

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

KARINE NICOLE SIQUEIRA

**ANÁLISE DO PAN-GENOMA DE CEPAS DE *Streptococcus agalactiae*
ISOLADAS DE TILÁPIA DO NILO (*Oreochromis niloticus*)**

DOIS VIZINHOS

2023

KARINE NICOLE SIQUEIRA

**ANÁLISE DO PAN-GENOMA DE CEPAS DE *Streptococcus agalactiae*
ISOLADAS DE TILÁPIA DO NILO (*Oreochromis niloticus*)**

**Pangenome analysis of *Streptococcus agalactiae* strains isolated from Nile
tilapia (*Oreochromis niloticus*)**

Trabalho de conclusão de curso de Especialização apresentado como requisito para obtenção do título de Especialização em Biologia Molecular – Habilitação Bioinformática da Universidade Tecnológica Federal do Paraná (UTFPR).

Orientadora: Prof. Dra. Betty Cristiane Kuhn

Coorientadores: Prof. Dra. Deborah Catharine de Assis Leite e Prof. Dr. Laurival Antônio Vilas-Boas

DOIS VIZINHOS

2023



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Esta licença permite download e compartilhamento do trabalho desde que sejam atribuídos créditos ao(s) autor(es), sem a possibilidade de alterá-lo ou utilizá-lo para fins comerciais. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

KARINE NICOLE SIQUEIRA

**ANÁLISE DO PAN-GENOMA DE CEPAS DE *Streptococcus agalactiae*
ISOLADAS DE TILÁPIA DO NILO (*Oreochromis niloticus*)**

Trabalho de Conclusão de Curso de Especialização
apresentado como requisito para obtenção do título de
Especialista em Biologia Molecular – Habilitação
Bioinformática da Universidade Tecnológica Federal
do Paraná (UTFPR).

Data de aprovação: 10/02/2023

Betty Cristiane Kuhn

Doutorado

Universidade Tecnológica Federal do Paraná (UTFPR)

Naiana Cristine Gabiatti

Doutorado

Universidade Tecnológica Federal do Paraná (UTFPR)

Same Costa Lima

Doutorado

Universidade Estadual de Londrina (UEL)

DOIS VIZINHOS

2023

RESUMO

A Estreptococose, causada pela bactéria *Streptococcus agalactiae*, é uma das principais causadoras de mortes e perdas econômicas em criações de tilápias do Nilo (*Oreochromis niloticus*) ao redor do mundo. Sendo assim, faz-se necessário estudos genômicos e de patogenicidade a fim de compreender melhor o processo de infecção e o potencial patogênico desse microrganismo. A análise de pan-genoma é uma ferramenta de estudo com grande potencial para investigar profundamente as semelhanças e diferenças dos genomas entre cepas de uma mesma espécie. Apesar de já terem sido publicados trabalhos acerca do pan-genoma de *S. agalactiae*, não há disponível na literatura pesquisas que foquem no pan-genoma de cepas isoladas de tilápia do Nilo. Desta forma, o objetivo desse trabalho é analisar o pan-genoma de 49 cepas de *S. agalactiae* isoladas de tilápia do Nilo. Para isso, os genomas foram analisados pelo Kbase Predictive Biology, que utilizaram principalmente, os algoritmos OrthoMCL e FastTree2 para a predição de genes ortólogos e análise filogenética das cepas, respectivamente. A comparação genômica mostrou que, de forma geral, os genomas das 49 cepas de *S. agalactiae* são muito conservados, mas apresentam diversas regiões compartilhadas e exclusivas, que garantem a diversidade genética. O estudo filogenético mostrou que as cepas foram agrupadas conforme o sorotipo capsular. Na análise de pan-genoma foram encontradas 5154 famílias gênicas, sendo 1277 famílias pertencentes ao genoma *core* e 3877 do genoma acessório. O genoma *core* apresentou, principalmente, genes relacionados a funções essenciais a células, como metabolismo de proteínas e parede celular, enquanto o genoma acessório exibiu tanto genes essenciais quanto regiões que poderiam oferecer vantagem às células. O estudo comparativo entre as cepas Maranhão, Recife e E8ang2 revelou a presença de importantes genes associados a virulência no genoma exclusivo da E8ang e no genoma compartilhado entre Maranhão e Recife. Com os resultados obtidos, foi permitido traçar um perfil geral acerca do pan-genoma do grupo estudado e caracterizar as regiões genômicas que compõe os genomas *core* e acessório, assim como entender o perfil filogenético das 49 cepas. Os dados gerados evidenciam a importância do estudo de cepas de *S. agalactiae* isoladas de tilápia do Nilo e mostram a necessidade de estudos que aprofundem ainda mais a caracterização genômica desses patógenos.

Palavras-chave: análise filogenética; estreptococose; genômica comparativa; Kbase; *Streptococcus* do grupo B.

ABSTRACT

Streptococcosis, caused by *Streptococcus agalactiae*, is one of the main causes of death and economic losses in Nile tilapia (*Oreochromis niloticus*) farms around the world. Therefore, it is necessary genomic and pathogenicity studies to have a better understanding of the infection process and the pathogenic potential of this microorganism. Pangenome analysis is a study tool with great potential to deeply analyze the similarities and differences of genomes among strains of the same species. Although studies on the pangenome of *S. agalactiae* have already been published, there are no studies available in the literature that focus on the pangenome of Nile tilapia isolated strains. Thus, this research aims to investigate the pangenome of 49 strains of *S. agalactiae* isolated from Nile tilapia. For that, the genomes were analyzed by KBase Predictive Biology software, which used mainly the OrthoMCL and FastTree2 algorithms for orthologues prediction and phylogenetic analysis of the strains, respectively. The genomic comparison showed that, in general, the genomes of the 49 strains of *S. agalactiae* are highly conserved, but exhibit several shared and exclusive regions, which guarantee genetic diversity. The phylogenetic study showed that the strains were grouped according to the capsular serotype. In the pan-genome analysis, 5154 gene families were found, with 1277 families belonging to the core genome and 3877 to the accessory genome. The core genome was composed mainly of genes related to essential cell functions, such as protein metabolism and cell wall, while the accessory genome showed both essential genes and regions that could offer advantages to cells. The comparative study among Maranhão, Recife, and E8ang2 strains revealed the presence of important virulence-associated genes in the E8ang exclusive genome and in the shared genome of Maranhão and Recife. With these results, it was possible to obtain a general profile of the studied group pan-genome and to characterize the genomic regions that make up the core and accessory genomes, as well as understand the phylogenetic profile of the 49 strains. These data show the importance of studying *S. agalactiae* strains isolated from Nile tilapia, besides showing the need for further studies that deepen the genomic characterization of these pathogens.

Keywords: phylogenetic analysis; streptococcosis; comparative genomics; Kbase; Group B Streptococcus.

LISTA DE ILUSTRAÇÕES

Quadro 1 - Código de acesso e origem de isolamento das 49 cepas de <i>S. agalactiae</i> analisadas nesta pesquisa.....	18
Figura 1 - Árvore filogenética das 49 cepas de <i>S. agalactiae</i> gerada pelo programa Insert Genome into Species Tree (KBase), usando os 5 parentes mais próximos como grupo externo	24
Figura 2 - Visualização circular dos 49 genomas de <i>S. agalactiae</i> gerada pelo programa Pangenome Circle Plot (Kbase)	26
Figura 3 - Análise filogenética baseada no pan-genoma, obtida através do programa Phylogenetic Pangenome Accumulation	29
Figura 4 - Papéis funcionais dos genes do genoma core obtidos através da anotação dos genomas pelo RAST e pelo programa View Function Profile for FeatureSet.....	31
Figura 5 - Papéis funcionais dos genes do genoma compartilhado obtidos através da anotação dos genomas pelo RAST e pelo programa View Function Profile for FeatureSet.....	33
Figura 6 - Papéis funcionais dos genes do genoma exclusivo obtidos através da anotação dos genomas pelo RAST e pelo programa View Function Profile for FeatureSet.....	34
Figura 7 - Diagrama de Venn mostrando os clusters compartilhados e exclusivos das cepas Maranhão, Recife e E8ang2, construído através da plataforma OrthoVenn2	35

LISTA DE TABELAS

Tabela 1 - Principais características dos 49 genomas de <i>S. agalactiae</i>	22
Tabela 2 - Tabela 2 - Dados referentes ao pan-genoma de 49 cepas de <i>S. agalactiae</i> geradas a partir do algoritmo OrthoMLC	26
Tabela 3 - Contagem de clusters e proteínas na comparação das cepas Maranhão, Recife e E8ang2 obtida através da plataforma OrthoVenn2	36

LISTA DE ABREVIATURAS E SIGLAS

CC	Complexo clonal
CG	Citosina Guanina
COG	Clusters of Orthologous Groups
CPS	Polissacarídeo capsular
CRISPR	<i>Clustered Regularly Interspaced Short Palindromic Repeats</i>
DNA	Ácido desoxirribonucleico
GBS	<i>Streptococcus</i> do grupo B
MCL	<i>Markov Cluster Algorithm</i>
MLST	Multilocus sequence typing
NCBI	<i>National Center for Biotechnology Information</i>
PATRIC	<i>Pathosystems Resource Integration Center</i>
RAST	<i>Rapid Annotation using Subsystem Technology</i>
RNA	Ácido ribonucleico
rRNA	Ácido ribonucleico ribossomal
ST	<i>Sequence type</i>
tRNA	Ácido ribonucleico transportador
UEL	Universidade Estadual de Londrina
UTFPR	Universidade Tecnológica Federal do Paraná

SUMÁRIO

1 INTRODUÇÃO	9
1.1 Objetivos	10
1.1.1 Objetivo Geral	10
1.1.2 Objetivos Específicos	10
2 REVISÃO DE LITERATURA	11
2.1 Tilápia do Nilo na aquacultura	11
2.2 <i>Streptococcus agalactiae</i>	12
2.2.1 Meios de controle da estreptococose em criações de tilápia	13
2.3 Caracterização genômica de <i>S. agalactiae</i>	14
2.4 Pan-genoma de bactérias	15
3. MATERIAL E MÉTODOS	18
3.1 Genomas sequenciados de <i>S. agalactiae</i>	18
3.2 Construção e análise do pan-genoma	19
3.3 Análise filogenética e caracterização do pan-genoma	20
4. RESULTADOS	22
4 DISCUSSÃO	37
5 CONSIDERAÇÕES FINAIS	46
REFERÊNCIAS	48

1 INTRODUÇÃO

Streptococcus agalactiae, também chamado de *Streptococcus* do grupo B, é um patógeno causador de altas taxas de mortalidade e perdas econômicas em pisciculturas ao redor do mundo, principalmente em criações de tilápias do Nilo, o peixe mais produzido no Brasil e terceiro mais produzido mundialmente (FAO, 2022; PEIXE BR, 2022; SUN et al., 2016). Essa bactéria já foi reportada como a causadora de surtos de Estreptococose em tilápias em diversos países, com sinais clínicos como natação errática, exoftalmia e escoliose e escurecimento da pele, e uma taxa de mortalidade que pode chegar a até 70% em menos de uma semana (PRADEEP et al., 2016; ZHANG, 2021).

A partir dos primeiros sequenciamentos de genomas de linhagens *S. agalactiae* no começo do milênio (GLASER et al., 2002; TETTELIN et al., 2002), foi possível evidenciar características genômicas únicas dessa espécie bem como as diferenças existentes tanto intra quanto inter-específicas dentro do gênero *Streptococcus* (BROCHET et al., 2006; LEFÉBURE; STANHOPE, 2007). Trabalhos prévios descreveram que o genoma de *S. agalactiae* possui diversos genes relacionados a virulência (BURNHAM; TYRRELL, 2003; SHABAYEK; SPELLERBERG, 2016), ilhas genômicas (HERBERT et al., 2005), contém genes específicos de acordo com a linhagem, além de várias características singulares em relação ao pan-genoma dessa espécie (GORI et al., 2020; TETTELIN et al., 2005).

A análise de pan-genoma é uma ferramenta da genômica comparativa utilizada para compreender o repertório genético de uma espécie ou conjunto de organismos. Essa análise permite definir o genoma “core”, ou seja, os genes presentes em todos os genomas estudados, o genoma compartilhado composto por genes presentes em vários organismos, mas não em todos, além de auxiliar a visualização de regiões genômicas exclusivas de cada cepa (HILLER; SÁ-LEÃO, 2018; KIM et al., 2020).

Estudos prévios de pan-genoma de *S. agalactiae* revelaram que o genoma core dessa espécie compreende cerca de 80% do genoma total e é composto principalmente de genes *housekeeping*, aqueles responsáveis por funções regulatórias e transporte de proteínas, por exemplo; em contrapartida, o genoma acessório, ou dispensável, abrange genes relacionados à virulência, resistência a antibióticos, elementos móveis e adaptação (TETTELIN et al., 2005; GORI et al.;

2020). Tettelin et al. (2005), ainda mostraram que o pan-genoma dessa espécie pode ser considerado aberto, pois novos genes continuarão a serem adicionados ao pan-genoma à medida que novas cepas forem sequenciadas e adicionadas à análise.

Apesar de toda informação disponível acerca da composição genética de *S. agalactiae*, até a presente data não há disponível na literatura estudos e análises de pan-genoma que foquem em cepas isoladas de tilápia do Nilo. Considerando o crescente número de genomas publicados em banco de dados públicos e a importância dessa espécie na piscicultura, evidenciar as particularidades, diferenças e semelhanças entre esse grupo poderá trazer novas perspectivas acerca do processo de infecção e patogenicidade dessas cepas, evolução do pan-genoma dessa bactéria, além de contribuir na produção de vacinas e terapias de controle da Estreptococose.

1.1 Objetivos

1.1.1 Objetivo Geral

Caracterizar os genomas *core*, compartilhado e exclusivo de 49 cepas de *S. agalactiae* isoladas de surtos de Estreptococose em tilápias do Nilo (*Oreochromis niloticus*).

1.1.2 Objetivos Específicos

- Descrever os papéis funcionais dos genes que compõe os genomas *core* e acessório do pan-genoma.
- Identificar a relação filogenética entre as cepas de *S. agalactiae* estudadas.
- Comparar o conteúdo genômico de 3 cepas de *S. agalactiae* (Maranhão, Recife e E8ang2) isoladas recentemente e ainda não publicadas.

2 REVISÃO DE LITERATURA

Na seguinte revisão de literatura, serão apresentados os principais pontos acerca da tilápia do Nilo, sua importância, produção e características, além de informações sobre a bactéria *S. agalactiae*, e como esse patógeno irá impactar o cultivo do peixe. Por fim, será apresentado um breve texto sobre pan-genoma de bactérias patogênicas.

2.1 Tilápia do Nilo na aquicultura

A pesca e o cultivo de organismos aquáticos (aquicultura) têm contribuído de forma significativa na segurança alimentar e nutrição ao redor do mundo. Em 2020, a produção mundial de animais aquáticos chegou a 178 milhões de toneladas, tendo a aquicultura contribuído para 88 milhões de toneladas dessa produção (FAO, 2022; OLIVEIRA, 2015). Em relação ao cultivo de peixes, as três maiores espécies que se destacaram na piscicultura foram as Carpas Capim (*Ctenopharyngodon idellus*), Carpas prateada (*Hypothalmichthys molitrix*) e as tilápias do Nilo (*Oreochromis niloticus*), sendo que as tilápias em 2020 apresentaram um cultivo de 4,4 milhões de toneladas, representando em torno de 2,2% da produção mundial (FAO, 2022).

A China é o país com a maior produção de tilápia, seguida da Indonésia, Egito, Brasil e Tailândia. O Brasil, se destaca como o quarto maior produtor, sendo a tilápia do Nilo o animal aquático mais produzido no país e representando cerca de 63,5% da produção nacional. Entre os estados brasileiros, o Paraná apresenta a maior produção de tilápia do Nilo, seguido de São Paulo, Minas Gerais, Santa Catarina e Mato Grosso do Sul. Além disso, esse peixe também é responsável por cerca de 88% das exportações, tendo os Estados Unidos como principal importador. (PEIXE BR, 2022).

A tilápia do Nilo é um peixe de água doce da família Cichlidae e representa a espécie de tilápia mais cultivada ao redor do mundo. Esse peixe possui diversas características que o fazem ser comercialmente atrativos, como rápido crescimento e fácil reprodução, alimentação diversa e sobrevivência e adaptação a diferentes condições de cultivo (AMAL; ZAMRI-SAAD, 2011; EL-SAYED, 2019). Além disso, é um peixe de carne branca com ótima textura e sabor, apresenta filés grandes, tem baixo teor de gordura e pode ser consumido de várias maneiras (SENAR, 2018a).

Na aquicultura, a criação de tilápias se dá principalmente em dois tipos de ambientes: semi-intensivo e intensivo. O cultivo semi-intensivo é caracterizado por

menores produções com baixas densidades de estocagem, maior controle da qualidade da água e uso de alimentação natural através da fertilização dos tanques em adição a alimentação fornecida (EL-SAYED, 2019; ROSSATO et al, 2021). Em contrapartida, no cultivo intensivo o principal objetivo é a maximização da produção, com altas densidades de estocagem e uso de alimentação comercial (EL-SAYED, 2019; EMBRAPA, 2009).

Principalmente nos sistemas intensivos, o uso de elevadas densidades de estocagem, diminuição da qualidade da água e manejo incorreto dos animais podem levar ao aumento do stress e provocar lesões nesses peixes deixando-os menos resistentes e mais susceptíveis a doenças (MARCUSO; SALVADOR; MARINHO-NETO, 2017; MAULU et al., 2021). A Estreptococose, infecção causada pela bactéria *Streptococcus agalactiae*, tem sido reportada em diversos países com climas tropicais e temperados, e tem se mostrado causadora de altas taxas de mortalidade e morbidade em criações de tilápia (KLESIOUS; SHOEMAKER; EVANS, 2008; ZHANG, 2021).

2.2 *Streptococcus agalactiae*

S. agalactiae é uma bactéria Gram-positiva, catalase e oxidase negativa, possui um formato esférico com cerca de 0,5 a 2µm de diâmetro e pode crescer em pares ou cadeias curtas quando cultivada; além disso, é anaeróbia facultativa, possui a capacidade de fermentar ácido láctico e pode ser α, β ou não-hemolítica (PHARES et al., 2008; ZHANG, 2021). Esse microrganismo também pode ser chamado de *Streptococcus* do grupo B, ou GBS, e recebeu esse nome quando foi diferenciado de outros *Streptococcus* pela primeira vez em 1933 por Rebecca Lancefield, ao ser isolado de vacas que possuíam mastite, uma inflamação que atinge as glândulas mamárias (LANCEFIELD, 1933). *S. agalactiae* ainda pode ser dividida em 10 sorotipos de acordo com seu polissacarídeo capsular (CPS), sendo eles os sorotipos Ia, Ib, II, III, IV, V, VI, VII, VIII e IX (DOGAN et al. 2005; SLOTVED et al. 2007).

Em humanos, essa bactéria coloniza de forma assintomática o trato genital e intestinal de ambos os sexos; os primeiros relatos de patogenicidade ocorreram a partir de 1935 com estudos que demonstraram infecções fatais em mulheres no puerpério (FRY, 1938; LANCEFIELD; HARE, 1935); além do pós parto, infecções invasivas por GBS também podem acometer gestantes, idosos e adultos com

comorbidades e/ou imunocomprometidos, e as principais manifestações clínicas são bacteremia e infecção de tecidos moles (RAABE; SHANE, 2019; SKOFF et al., 2009). Em neonatos, *S. agalactiae* está entre as principais causadoras de infecções bacterianas devido ao contato direto com o canal vaginal da mãe durante o parto, e pode se manifestar como pneumonia, sepse, infecção de tecidos moles e ossos e até mesmo meningite (RAABE; SHANE, 2019; YAGUPSKY; MENEGUS; POWELL, 1991).

A Estreptococose, atinge uma variedade de animais como, por exemplo, porcos, roedores, cães, bovinos e peixes, estes últimos com uma maior incidência (MAISANO et al., 2018; MORACH et al. 2018). Em tilápias, a infecção por *S. agalactiae* é um dos principais causadores de mortes e prejuízos econômicos em criações desse peixe por todo o mundo (MIAN et al., 2009; YE et al., 2011).

A transmissão em tilápias pode ocorrer entre contato direto de peixes saudáveis e doentes, por contato indireto através da água, além de haver evidências de transmissão vertical; a mortalidade pode chegar em até 70% em um período entre 3 e 7 dias, e ocorre principalmente em sistemas de criação intensivos e com temperatura e oxigenação fora dos níveis recomendados (PRADEEP et al., 2016). Os sinais clínicos de tilápias infectadas podem variar, porém entre os mais comuns estão a natação errática, exoftalmia e opacidade córnea uni ou bilateral, escurecimento da pele, distensão abdominal, escoliose, letargia, meningite e inflamação ao redor do coração e rim; também, com o agravamento da infecção, muitos peixes podem se recusar a comer, causando anorexia (PRADEEP et al., 2016; ZHANG, 2021).

2.2.1 Meios de controle da estreptococose em criações de tilápia

Os meios de prevenção e controle da Estreptococose nas tilápias incluem desde o manejo correto e biossegurança na produção, até o uso de antibióticos, alimentação suplementada, probióticos e vacinação (MAULU et al., 2021). Nos sistemas de criação, a água deve ser tratada e monitorada quanto a sua qualidade, com constante filtração, circulação, oxigenação, controle de patógenos e controle da temperatura; além disso, é importante isolar os peixes novos que serão introduzidos na produção para possível observação de doenças e sempre retirar dos tanques

aqueles que estão mortos ou com sinais patológicos, para que não haja transmissão horizontal (KLESIUS; SHOEMAKER; EVANS, 2008; ZAMRI-SAAD et al., 2014).

Em relação a estocagem, a quantidade de peixes deve estar de acordo com o tamanho do tanque utilizado, pois ao evitar o adensamento também se reduz a taxa de estresse e as feridas na pele causadas pelo atrito entre os animais, esses que são fatores relevantes no aumento da mortalidade por *S. agalactiae*, além de outros patógenos. Também, deve-se manter uma boa higienização dos tanques e equipamentos, e evitar transportes e manuseio dos peixes em excesso (XU; SHOEMAKER; KLESIUS, 2007; ZAMRI-SAAD et al., 2014).

A vacinação tem se mostrado altamente eficiente na prevenção de surtos da doença, induzindo uma grande resposta imune capaz de gerar uma resistência a longo prazo à infecção por *S. agalactiae* nas tilápias (CAIPANG; LUCANAS; LAY-YAG, 2014; KLESIUS; SHOEMAKER; EVANS, 2008). De modo geral, as vacinas através de injeção intra-visceral são as mais utilizadas, produzindo uma resposta mais rápida do sistema imunológico e uma alta produção de anticorpos; porém, essa técnica exige mão-de-obra intensiva, habilidade técnica, treinamento qualificado e equipamentos adequados, sendo ainda mais trabalhosa quando há uma grande quantidade de peixes (ABU-ELALA et al., 2019; SILVA et al., 2009). Já as vacinas por imersão, apresentam uma eficácia menor se comparadas com aquelas por injeção intravisceral, mas sua aplicação, por não haver contato direto entre os peixes e os manipuladores, reduz significativamente o stress dos animais, além de ser economicamente mais acessível e mais apropriada para vacinação em massa, principalmente de peixes jovens e pequenos (BØGWALD; DALMO, 2019).

2.3 Caracterização genômica de *S. agalactiae*

Os dois primeiros genomas completos sequenciados dessa bactéria foram isolados de humanos, onde NEM316, sorotipo III, continha um genoma de 2.211.485 pb com 2.118 genes codificadores de proteínas (GLASER et al., 2002), e a cepa 2603 V/R, sorotipo V, possuía 2.160.267 pb e 2.175 genes codificadores (TETTELIN et al., 2002). Ambos os genomas foram publicados em 2002, e com eles foi possível, pela primeira vez, ter uma visão da evolução e organização do material genético da *S. agalactiae* e compará-los com outros genomas já sequenciados, como de *Streptococcus pyogenes* e *Streptococcus pneumoniae*. (CHEN, 2019). Até a presente

data (novembro de 2022), é possível encontrar no banco de dados do National Center for Biotechnology Information (NCBI) 1.677 publicações de genomas montados e anotações, sendo que 143 são genomas completos, e destes, 46 são de cepas isoladas de tilápia do Nilo (NCBI, [2022]).

O tamanho e características do material genético da *S. agalactiae* podem variar de acordo com a cepa, mas de forma geral o genoma possui entre 1,8Mb e 2,2Mb, conteúdo CG variando de 35,3% a 35,9%, quantidade de rRNA de 18 a 21 e tRNA de 63 a 81, cerca de 15 ilhas genômicas e várias classes de elementos genéticos móveis (CHEN, 2019; TETTELIN et al., 2005).

Conhecer a estrutura, diversidade genética, evolução e especificidade das cepas da *S. agalactiae* é primordial para o desenvolvimento de vacinas efetivas e seguras contra as doenças que essa bactéria pode causar (KANNIKA et al., 2017). A determinação do sorotipo capsular e a *Multilocus sequence typing* (MLST) são métodos de referência utilizados para investigar a relação genética e epidemiológica desse microrganismo (GODOY et al., 2013).

O sorotipo capsular irá diferenciar as linhagens entre 10 sorotipos: Ia, Ib, II – IX, sendo que cada um possui uma estrutura diferente, além de serem antigenicamente distintos (BARONY et al., 2017; JOHRI et al., 2006); os sorotipos mais comuns isolados de peixes, por exemplo, são o Ia, Ib e III (EVANS et al., 2008), sendo que o Ib, segundo estudo de Barony et al. (2017), é o mais encontrado no Brasil. Já o MLST irá diferenciar as cepas em tipos de sequência (ST - do inglês *sequence type*) a partir de fragmentos de genes *housekeeping*, e através de agrupamentos de ST em complexos clonais (CC), as cepas poderão ser organizadas em linhagens genéticas (MAIDEN, 2006). Além disso, em decorrência do sequenciamento de nova geração (NGS) e avanços nas ferramentas de bioinformática, se tornou muito mais rápido e economicamente viável a comparação entre genomas de cepas de GBS, esse que consegue analisá-las e diferenciá-las a nível de nucleotídeos (GRAD; LIPSITCH, 2014).

2.4 Pan-genoma de bactérias

Os avanços nas técnicas de sequenciamento de nova geração, nas ferramentas de bioinformática e análises *in silico* fizeram com que a construção e compreensão dos genomas de microrganismos patogênicos se tornasse muito mais

viável e economicamente acessível, favorecendo estudos comparativos entre cepas da mesma espécie e até mesmo entre espécies diferentes (KIM et al., 2020). O pan-genoma, termo introduzido por Tettelin et al. (2005) em um estudo com *S. agalactiae*, pode ser definido com o conjunto global de genes e regiões genômicas compartilhadas por uma espécie ou grupo de interesse; esse repertório genético permite, a partir do compilado e análise de uma quantidade significativa de genomas, definir o genoma *core*, acessório e exclusivo de um conjunto de indivíduos (MEDINI et al., 2005).

O genoma *core* são famílias de genes comuns a todas as cepas estudadas, ou seja, são as regiões genômicas essenciais para a sobrevivência e crescimento da espécie; ele é formado, por exemplo, por genes relacionados ao metabolismo de carboidratos e aminoácidos, obtenção de energia, envelope celular e proteínas de transporte (TETTELIN et al., 2005). Já o genoma acessório ou compartilhado é composto por aqueles genes presentes em duas ou mais cepas, mas não em todas; ele pode ser composto por genes que conferem vantagens a cepa, como de resistência a antibióticos e colonização e adaptação a diferentes ambientes e hospedeiros (MEDINI et al., 2005). Por fim, tem-se os genes exclusivos, aqueles presentes em apenas uma única cepa (ROULI et al., 2015). Ambos os genomas acessório e exclusivo podem ser considerados dispensáveis por não apresentarem funções essenciais, mas são os grandes responsáveis pela diversidade entre organismos (MEDINI et al., 2005).

Um pan-genoma ainda pode ser classificado como aberto ou fechado, e essa natureza está relacionada com o estilo de vida da espécie. Um pan-genoma fechado é aquele onde assume-se que todo o repertório genético do grupo estudado já foi identificado e que, caso sejam adicionados mais genomas à análise, não serão descobertos novos genes (VERNIKOS et al., 2015) Já um pan-genoma aberto é aquele que seria necessário um número indeterminado de genomas para definir a quantidade exata de genes do grupo de interesse, ou seja, a medida que são sequenciados e adicionados genomas de novas cepas, novos genes são identificados (ROULI et al., 2015; TETTELIN et al. 2005). De acordo com estudos prévios, a grande maioria das bactérias patogênicas apresentam um pan-genoma aberto, e isso pode ser explicado pela plasticidade bacteriana, necessidade de genes de sobrevivência e adaptação a diferentes ambientes e hospedeiros e a capacidade de adquirir DNA

exógeno (HAMMOND et al., 2020; LAPIERRE; GOGARTEN, 2009; ROULI et al., 2015).

A análise de pan-genomas de microrganismos patogênicos, por exemplo, permite realizar um levantamento da diversidade genética da espécie, investigar a filogenia, desenvolvimento da patogenicidade e especialização nos diversos nichos, estudar o mobiloma (conjunto de elementos genéticos móveis) e resistoma (conjunto de genes de resistência a antibióticos), além de permitir traçar o perfil de virulência da bactéria estudada (CAPUTO; FOURNIER; RAOULT, 2019). Essa análise possibilita reclassificar os indivíduos dentro de uma espécie para além dos parâmetros tradicionais já utilizados, como sorotipo capsular e MLST, e pode revelar que organismos de sorotipos ou tipo de sequência (ST) diferentes podem ser mais semelhantes do que aqueles classificados no mesmo grupo, ou o contrário, onde indivíduos da mesma classificação podem ser distintos quando comparados os seus genomas completos (CAPUTO et al., 2015; KIM et al., 2020).

Outra grande aplicação das análises de pan-genoma está ligada às áreas médicas e biotecnológicas no desenvolvimento de estratégias terapêuticas e ferramentas e informações úteis para a manipulação de genes (DORON et al., 2018; KAVVAS et al., 2018). Com o grande aumento de patógenos resistentes a antibióticos, aumento de surtos, endemias e pandemias, as informações obtidas através de estudos de pan-genomas, principalmente sobre as regiões genômicas em comum entre as cepas estudadas de um determinado clado, podem ser ferramentas importantes para o desenvolvimento de vacinas e terapias de controle para essas bactérias patogênicas já caracterizadas bem como emergentes (KIM et al., 2020).

Em *S. agalactiae*, sabe-se que o genoma *core*, em sua grande maioria, é composto por genes *housekeeping*, aqueles que permitem a sobrevivência da célula, enquanto o genoma dispensável é representado por elementos móveis, ilhas genômicas, fatores de virulência e genes de adaptação que podem ter sido adquiridos por eventos de transferência horizontal. O pan-genoma de *S. agalactiae* é considerado aberto pois seu tamanho continuará aumentando à medida que novas cepas são sequenciadas, sendo assim necessário uma quantidade indeterminada de genomas para definir por completo o pan-genoma de *S. agalactiae* (GORI et al., 2020; KOIDE et al., 2022; TETTELIN et al., 2005).

3. MATERIAL E MÉTODOS

3.1 Genomas sequenciados de *S. agalactiae*

Para a análise de pan-genoma de cepas de *S. agalactiae* isoladas de tilápia do Nilo, foram utilizados 46 genomas completos disponíveis na base de dados National Center for Biotechnology Information (NCBI); como critério de inclusão, adicionou-se à análise apenas aqueles que possuíam nível de montagem completo e cujo hospedeiro de isolamento é a tilápia do Nilo. Além disso, foram incluídos nessa pesquisa os genomas de três cepas previamente isoladas, caracterizadas e sequenciadas pelo Laboratório de Micologia Veterinária da Universidade Estadual de Londrina (UEL), os quais ainda não foram publicados e seu nível de montagem está em contigs. Utilizou-se o BV-BRC v3.28.22 (Bacterial and Viral Bioinformatics Resource Center), uma ferramenta online de análise de genomas, para obtenção de alguns dados, como o conteúdo CG e o número de genes para tRNAs e rRNAs. No Quadro 1, estão listados os genomas utilizados e a origem de isolamento de cada um.

Quadro 1 - Código de acesso e origem de isolamento das 49 cepas de *S. agalactiae* analisadas nesta pesquisa

(continua)

Cepa	Número de acesso (GenBank - NCBI)	Origem de isolamento
WC1535	CP016501.2	Província de Hainan, China
TFJ0901	CP034315.1	Província de Fujian, China
HN016	CP011325.1	Província de Hainan, China
S73	CP030845.1	Piauí, Brasil
SA95	CP019812.1	Alagoas Brasil
SA97	CP019813.1	Pernambuco, Brasil
SA132	CP019815.1	Ceará, Brasil
SA85	CP019811.1	Alagoas, Brasil
SA218	CP019824.1	Espírito Santo, Brasil
SA102	CP019814.1	Pernambuco, Brasil
SA136	CP019816.1	Ceará, Brasil
SA75	CP019808.1	Ceará, Brasil
SA289	CP019828.1	Ceará, Brasil
SA256	CP019827.1	Ceará, Brasil
SA53	CP019802.1	Ceará, Brasil
SA245	CP019826.1	Minas Gerais, Brasil
SA73	CP019803.1	Ceará, Brasil
SA191	CP019819.1	Ceará, Brasil
SA374	CP019834.1	São Paulo, Brasil
SA375	CP019835.1	São Paulo, Brasil

Quadro 1 - Código de acesso e origem de isolamento das 49 cepas de *S. agalactiae* analisadas nesta pesquisa

(conclusão)

Cepa	Número de acesso (GenBank - NCBI)	Origem de isolamento
SA623	CP019836.1	Minas Gerais, Brasil
SA341	CP019831.1	Minas Gerais, Brasil
SA330	CP019829.1	São Paulo, Brasil
SA627	CP019837.1	Minas Gerais, Brasil
SA333	CP019830.1	Goiás, Brasil
SA346	CP019833.1	Minas Gerais, Brasil
SA343	CP019832.1	Minas Gerais, Brasil
SA220	CP019825.1	São Paulo, Brasil
SA212	CP019823.1	São Paulo, Brasil
SA20	CP003919.2	Paraná, Brasil
SA79	CP019809.1	Santa Catarina, Brasil
SA5	CP019805.1	Espírito Santo, Brasil
SA1	CP019804.1	Minas Gerais, Brasil
SA9	CP019806.1	Espírito Santo, Brasil
SA184	CP019818.1	Espírito Santo, Brasil
SA16	CP019807.1	São Paulo, Brasil
SA209	CP019822.1	São Paulo, Brasil
SA201	CP019821.1	Espírito Santo, Brasil
SA30	CP019800.1	São Paulo, Brasil
SA195	CP019820.1	NI
SA33	CP019801.1	São Paulo, Brasil
SA159	CP019817.1	Minas Gerais, Brasil
GX026	CP011328.1	Província de Guangxi, China
S25	CP015976.1	Paraná, Brasil
S13	CP018623.1	Paraná, Brasil
138P	CP007482.1	Idaho, EUA
Maranhão	Não Publicado	Maranhão, Brasil
Recife	Não Publicado	Pernambuco, Brasil
E8ang2	Não Publicado	Paraná, Brasil

NI: não informado

Fonte: autoria própria (2023)

3.2 Construção e análise do pan-genoma

A construção do pan-genoma das 49 cepas de *S. agalactiae* selecionadas se deu através do software Kbase Predictive Biology (<https://www.kbase.us/>), uma plataforma de acesso livre que reúne diversas ferramentas, programas e aplicativos para análise de dados biológicos. Os genomas foram adicionados à plataforma através de upload em formato FASTA ou diretamente do NCBI através do código de acesso. Genomas adicionados por upload foram anotados pelo programa Annotate

Microbial Assembly with RASTtk - v1.073 (BRETTIN T. et al, 2015), que fornece anotações rápidas usando a tecnologia de subsistemas. Todas as cepas foram agrupadas através do programa Batch Create GenomeSet v1.2.0. (ARKIN A. P. et al., 2018) e a anotação do restante dos genomas foi realizada pelo programa Annotate Multiple Microbial Genomes with RASTtk v1.073 (BRETTIN T. et al, 2015).

Para a predição de genes ortólogos e parálogos, o conjunto de genomas foi analisado através do programa Build Pangenome with OrthoMCL - v2.0 (LI; STOECKERT; ROOS, 2003) usando o algoritmo de agrupamento de Markov. Nessa ferramenta, de forma geral, comparou-se todas as proteínas dos genomas através do algoritmo BLASTp e calculou-se a porcentagem de correspondência entre elas; foram mantidas correspondências com $e\text{-value} < 1e\text{-}5$ e porcentagem de correspondência $\geq 50\%$. Ainda dentro do programa OrthoMCL, encontrou-se potenciais pares parálogos, ortólogos e co-ortólogos usando a ferramenta Orthomcl Pairs. Por fim, esses pares foram classificados em grupos de acordo com a função através do *Markov Cluster Algorithm* (MCL). Grupos de interesse foram analisados através da plataforma online OrthoVenn2 (<https://orthovenn2.bioinfotoolkits.net/>) para comparação e visualização de clusters de genes ortólogos (Xu et al. 2019). Utilizou-se o pan-genoma gerado para produzir uma visualização circular dos genomas através do programa Pangenome Circle Plot v1.6.0 (LI; STOECKERT; ROOS, 2003). Escolheu-se a cepa WC1535 (CP016501.2) para ser o “genoma base” para o alinhamento do restante dos genomas, visto que este é o maior genoma do conjunto analisado.

3.3 Análise filogenética e caracterização do pan-genoma

Na análise filogenética entre as cepas, o conjunto de genomas foi agrupado e analisado no programa Insert Set of Genomes Into SpeciesTree v2.2.0 (PRICE; DEHAL; ARKIN, 2010) que utiliza a ferramenta FastTree2 e permite a inferência da filogenia pelo método de Máxima Verossimilhança através do alinhamento e análise de 49 genes *housekeeping*, esses que incluíram genes envolvidos na estrutura de ribossomos, tradução e biogênese. Posteriormente, o pangenoma foi analisado no contexto filogenético no Phylogenetic Pangenome Accumulation v1.4.0 (LI; STOECKERT; ROOS, 2003).

Para a caracterização e análise dos genomas *core* e acessório dos *clusters* de interesse, utilizou-se o programa View Function Profile for FeatureSet v1.6.0. Este

programa utiliza da notação prévia feita através do RAST para agrupar os genes de acordo com seu papel funcional na base de dados de anotações de genomas The SEED (<https://pubseed.theseed.org/>). The SEED é uma plataforma de anotação que propicia a análise e agrupamento de genes em subsistemas e a criação de conjuntos de proteínas que compõem e possibilitam a tecnologia de anotação automatizada RAST (OVERBEEK et al., 2005).

4. RESULTADOS

O estudo de pan-genoma de cepas de *S. agalactiae*, isoladas de tilápia do Nilo, permitiu analisar 49 genomas de cepas provenientes do Brasil, China e Estados Unidos. O tamanho dos genomas variou entre 1.84 a 2.19 Mb, com conteúdo GC entre 35,2 a 35,8% e uma média de 1.920 genes. Em relação ao sorotipo capsular, mais de 85% dos genomas são do sorotipo Ib; além disso, vários genomas foram tipados pela técnica de MLST, e foram encontradas as linhagens ST-7, ST-927, ST-260, ST-552, ST-261 e ST-1525. A Tabela 1 apresenta as principais características dos genomas.

Tabela 1 - Principais características dos 49 genomas de *S. agalactiae* analisados

(continua)

Cepa	Sorotipo	ST*	Tamanho do genoma (Mb)	% CG	Genes	rRNA	tRNA
WC1535	Ia	7	2.19	35.8	2158	21	80
TFJ0901	Ia	NT	2.08	35.7	1981	21	80
HN016	Ia	NT	2.06	35.7	1954	21	80
S73	III	NT	2.06	35.4	1997	18	71
SA95	Ib	927	1.86	35.5	1790	18	65
SA97	Ib	927	1.86	35.5	2031	18	65
SA132	Ib	260	1.85	35.5	1949	18	65
SA85	Ib	927	1.85	35.5	1979	18	65
SA218	Ib	927	1.85	35.5	1925	18	65
SA102	Ib	927	1.85	35.5	2087	18	65
SA136	Ib	260	1.85	35.5	1970	18	65
SA75	Ib	260	1.85	35.5	1919	18	65
SA289	Ib	260	1.85	35.5	1786	18	65
SA256	Ib	260	1.85	35.5	1788	18	65
SA53	Ib	260	1.85	35.5	1784	18	65
SA245	Ib	260	1.85	35.5	1786	18	65
SA73	Ib	260	1.85	35.5	1991	18	65
SA191	Ib	260	1.85	35.5	2028	18	65
SA374	Ib	NT	1.84	35.5	1993	18	63
SA375	Ib	NT	1.84	35.5	1968	18	63

Tabela 1 - Principais características dos 49 genomas de *S. agalactiae* (conclusão)

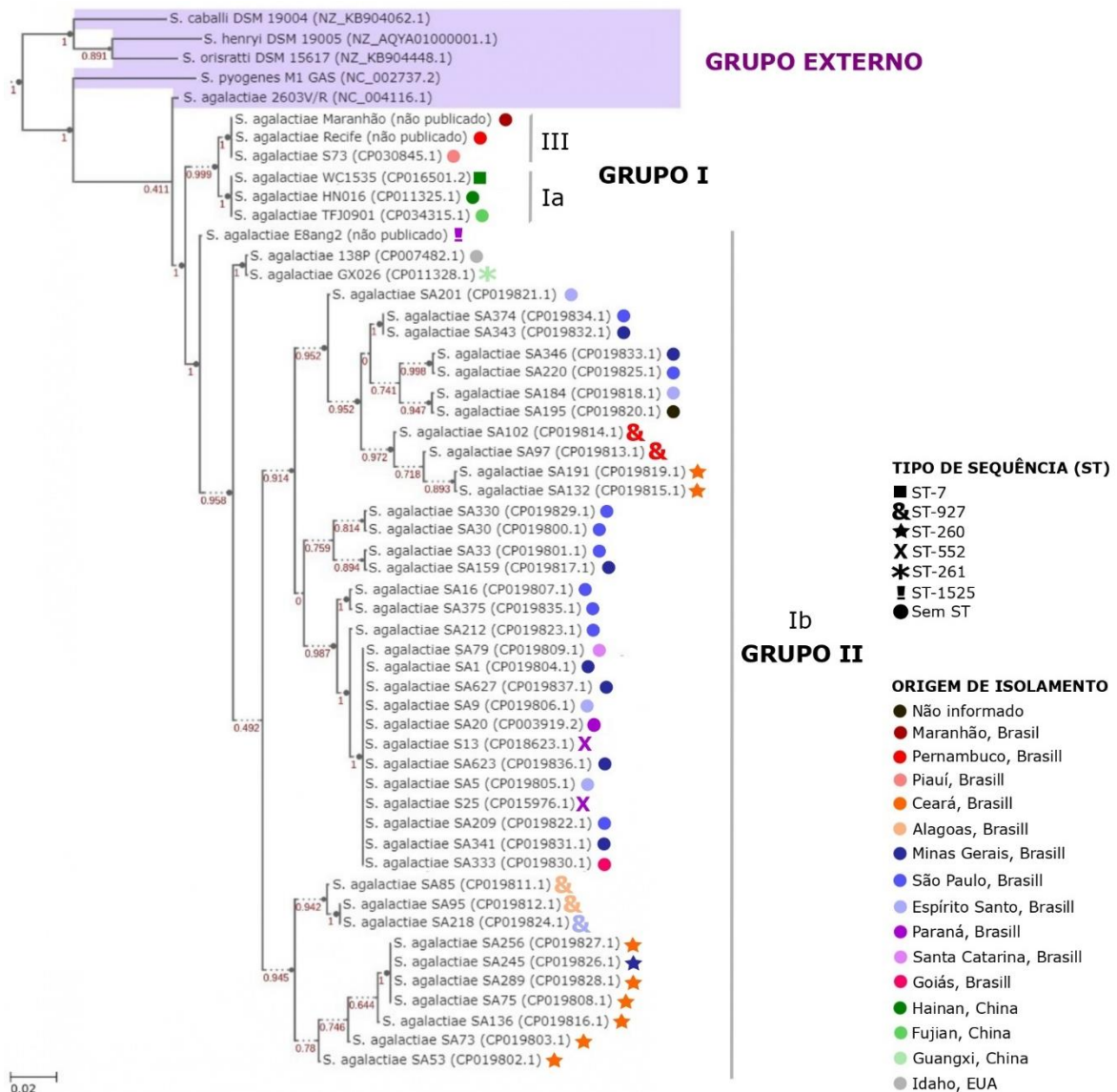
Cepa	Sorotipo	ST*	Tamanho do genoma (Mb)	% CG	Genes	rRNA	tRNA
SA623	lb	NT	1.84	35.5	1779	18	65
SA341	lb	NT	1.84	35.5	1956	18	63
SA330	lb	NT	1.84	35.5	1949	18	63
SA627	lb	NT	1.84	35.5	1780	18	65
SA333	lb	NT	1.84	35.5	1936	18	63
SA346	lb	NT	1.84	35.5	1947	18	63
SA343	lb	NT	1.84	35.5	1936	18	63
SA220	lb	NT	1.84	35.5	1930	18	63
SA212	lb	NT	1.84	35.5	1777	18	63
SA20	lb	NT	1.84	35.5	1923	18	63
SA79	lb	NT	1.84	35.5	1779	18	63
SA5	lb	NT	1.84	35.5	1777	18	63
SA1	lb	NT	1.84	35.5	1779	18	63
SA9	lb	NT	1.84	35.5	1777	18	63
SA184	lb	NT	1.84	35.5	2006	18	63
SA16	lb	NT	1.84	35.5	1931	18	63
SA209	lb	NT	1.84	35.5	1777	18	63
SA201	lb	NT	1.84	35.5	1775	18	63
SA30	lb	NT	1.84	35.5	1971	18	63
SA195	lb	NT	1.84	35.5	2024	18	63
SA33	lb	NT	1.84	35.5	1945	18	63
SA159	lb	NT	1.84	35.5	2097	18	63
GX026	lb	261	1.84	35.5	1778	18	83
S25	lb	552	1.84	35.5	1782	21	76
S13	lb	552	1.84	35.4	1785	15	59
138P	NI	NT	1.84	35.5	1905	16	72
Maranhão	III	NT	2.04	35.2	2092	5	52
Recife	III	NT	2.04	35.2	2099	5	51
E8ang2	lb	1525	2.08	35.3	2240	4	53

*ST: *sequence type*; NT: sem ST (do inglês, *non typeable*); NI: não informado

Fonte: autoria própria (2023)

Na análise de filogenia, o programa Set of Genomes Into SpeciesTree v2.2.0 utilizou 49 genes housekeeping para a construção da árvore filogenética (Figura 1); além disso, adicionou-se na análise os genomas de 5 cepas de bactérias pertencentes ao gênero *Streptococcus* como grupo externo.

Figura 1 - Árvore filogenética de cepas de *S. agalactiae* estimada pelo método de inferência de Máxima Verossimilhança por meio do programa Insert Genome into Species Tree (KBase)



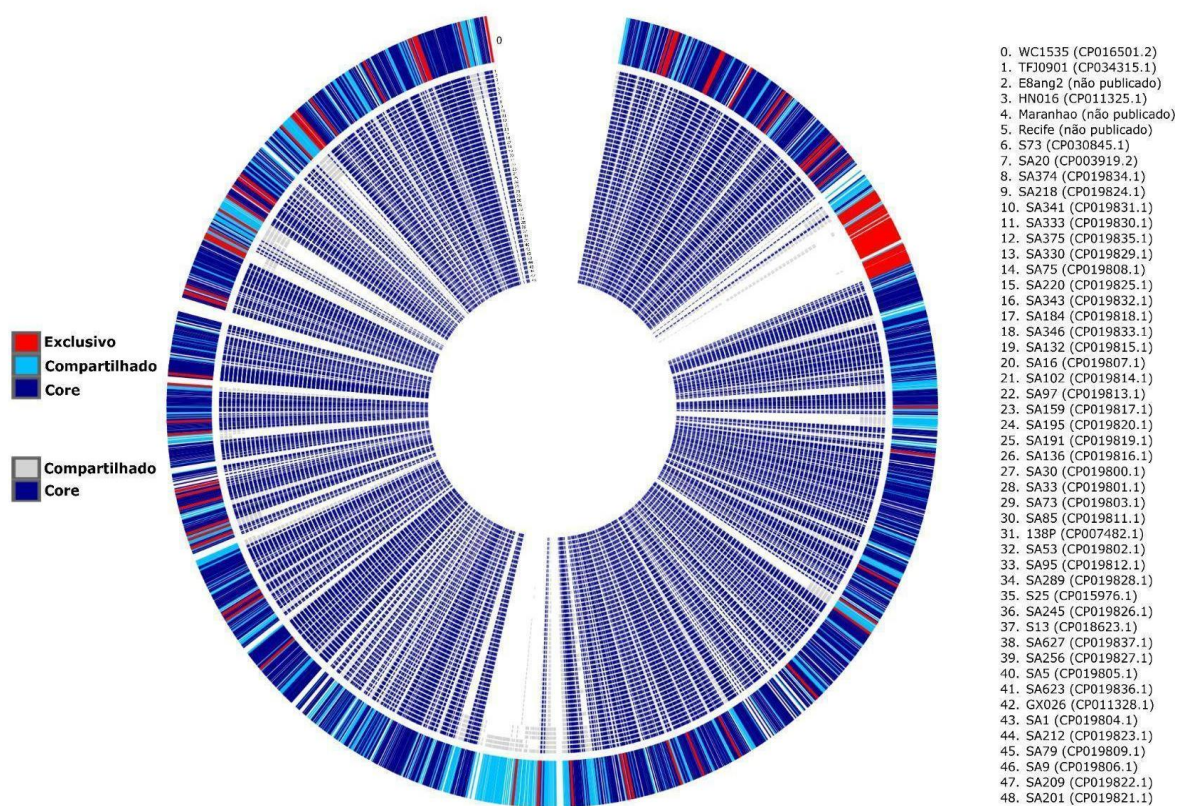
Fonte: Fonte: autoria própria (2023)

Na árvore, é possível observar que todos os genomas analisados nesta pesquisa foram agrupados junto ao genoma de *Streptococcus agalactiae* 2603V/R, uma cepa de referência desta espécie bacteriana. Também, analisou-se a formação de dois *clusters* referente ao sorotipo das bactérias: o Grupo I subdividiu-se em outros dois *clusters*, onde um agrupou todas as cepas do sorotipo III e o outro todas do sorotipo Ia, enquanto o segundo cluster (Grupo II) foi composto inteiramente por cepas de *S. agalactiae* do sorotipo Ib. Neste segundo agrupamento, a cepa *S. agalactiae* E8ang2, isolada no Paraná, foi a única a ficar separada do restante das cepas do sorotipo Ib, sugerindo-se uma menor similaridade quando comparada com o restante das bactérias.

Além disso, é possível observar uma tendência quanto a origem de isolamento e os cluster formados, mostrando que isolados de regiões brasileiras próximas estão mais intimamente relacionadas. Também, há uma tendência em relação ao ST; parte dos isolados do ST-260 foram agrupados em um mesmo clado, assim como parte das cepas do ST-927 e algumas cepas sem ST, formando clusters, como é possível observar na Figura 1. Todas as cepas de *S. agalactiae*, incluindo o genoma de referência 2603V/R analisados compartilham um ancestral comum com a espécie *S. pyogenes*, mostrando maior similaridade com essa espécie em relação as outras espécies do gênero *Streptococcus* incluídas nas análises.

No alinhamento e comparação dos 49 genomas de *S. agalactiae*, a visualização circular gerada pelo programa Pangenome Circle Plot (Figura 2), demonstrou uma alta taxa de conservação entre os genomas, bem como várias regiões ausentes e compartilhadas, o que mostra a diversidade genética entre as cepas de *S. agalactiae* isoladas de tilápia do Nilo.

Figura 2 - Visualização circular dos 49 genomas de *S. agalactiae* gerada pelo programa Pangenome Circle Plot (Kbase)



Fonte: autoria própria (2023)

Através do programa OrthoMCL, de forma geral, foram localizados 90.911 genes codificadores de proteínas nos 49 genomas, sendo que destes 88.179 são genes ortólogos entre as cepas, que estão distribuídos em cerca de 2.430 famílias ortólogas. Além disso, foram identificadas em torno de 2.726 famílias exclusivas de cepas, sendo que a média de genes exclusivos foi de 120 por genoma. A tabela 2 apresenta as principais estatísticas sobre o pan-genoma.

Tabela 2 - Tabela 2 - Dados referentes ao pan-genoma de 49 cepas de *S. agalactiae* geradas a partir do algoritmo OrthoMLC

(continua)

Genoma	Genes	Genes homólogos	Genes exclusivos	Famílias homólogas
WC1535	2158	1941	217	1886
TFJ0901	1981	1908	73	1864
HN016	1954	1891	63	1848

Tabela 2 - Dados referentes ao pan-genoma de 49 cepas de *S. agalactiae* geradas a partir do algoritmo OrthoMLC

(continua)

Genoma	Genes	Genes homólogos	Genes exclusivos	Famílias homólogas
S73	1997	1935	62	1911
SA95	1790	1605	185	1577
SA97	2031	1942	89	1864
SA132	1949	1882	67	1824
SA85	1979	1902	77	1840
SA218	1925	1870	55	1820
SA102	2087	1968	119	1893
SA136	1970	1892	78	1833
SA75	1919	1863	56	1814
SA289	1786	1603	183	1577
SA256	1788	1593	195	1567
SA53	1784	1603	181	1577
SA245	1786	1598	188	1573
SA73	1991	1896	95	1837
SA191	2028	1939	89	1870
SA374	1993	1909	84	1851
SA375	1968	1893	75	1838
SA623	1779	1591	188	1567
SA341	1956	1884	72	1830
SA330	1949	1889	60	1834
SA627	1780	1593	187	1569
SA333	1936	1875	61	1826
SA346	1947	1877	70	1829
SA343	1936	1872	64	1825
SA220	1930	1870	60	1822
SA212	1777	1581	196	1557
SA20	1923	1863	60	1817
SA79	1779	1581	198	1557
SA5	1777	1591	186	1567
SA1	1779	1583	196	1560

Tabela 2 - Dados referentes ao pan-genoma de 49 cepas de *S. agalactiae* geradas a partir do algoritmo OrthoMLC

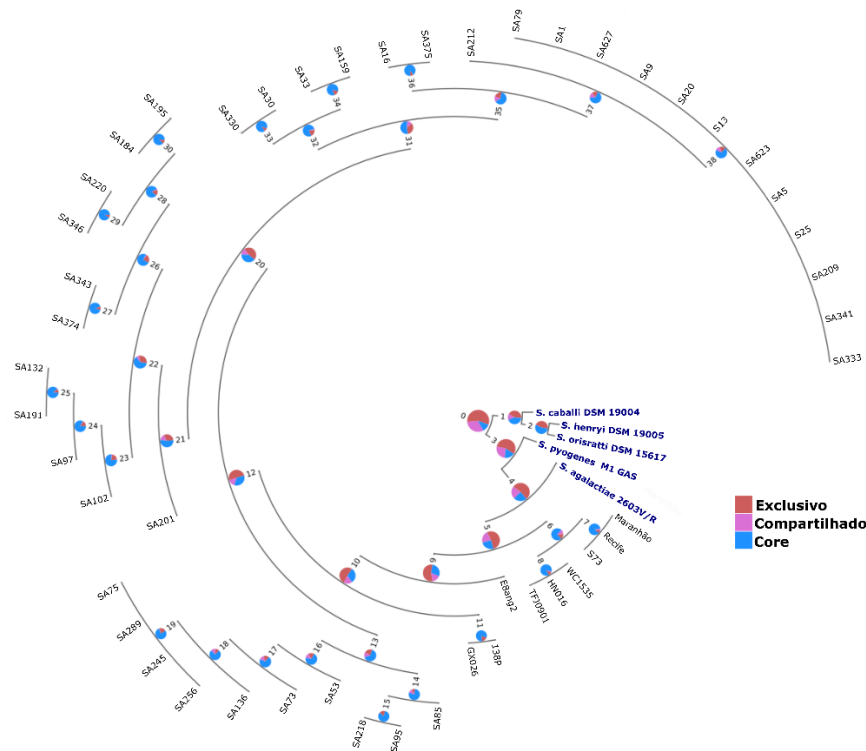
(conclusão)

Genoma	Genes	Genes homólogos	Genes exclusivos	Famílias homólogas
SA9	1777	1580	197	1556
SA184	2006	1928	78	1863
SA16	1931	1871	60	1822
SA209	1777	1579	198	1555
SA201	1775	1576	199	1553
SA30	1971	1891	80	1834
SA195	2024	1931	93	1863
SA33	1945	1874	71	1821
SA159	2097	1976	121	1896
GX026	1778	1573	205	1556
S25	1782	1588	194	1577
S13	1785	1597	188	1575
138P	1905	1797	108	1756
Maranhão	2092	2022	70	2001
Recife	2099	2020	79	1999
E8ang2	2240	2093	147	2031

Fonte: autoria própria (2023)

Através do programa Phylogenetic Pangenome Accumulation v1.4.0, analisou-se o pan-genoma em um contexto evolutivo e obteve-se a árvore filogenética, que revelou a formação de novos clusters, mas ainda mantendo o agrupamento de cepas do mesmo sorotipo, e a tendência acerca de origem de isolamento e ST (Figura 3). O cluster 5, que compreende todos os 49 genomas analisados nesta pesquisa, traz mais informações acerca do pan-genoma desse grupo: foi encontrado um total de 5154 famílias gênicas, sendo que, 2726 são exclusivas (52.9%), 1151 famílias compartilhadas (22.3%) e 1277 famílias pertencentes ao genoma *core* (24.8%), ou seja, presente em todos os genomas selecionados para este estudo.

Figura 3 - Análise filogenética baseada no pan-genoma, obtida através do programa Phylogenetic Pangenome Accumulation



NODE 0	TOTAL	7138	100%		NODE 13	TOTAL	2373	100%		NODE 26	TOTAL	2384	100%	
	singleton	4084	57.2%			singleton	517	21.8%			singleton	521	21.9%	
	partial	2215	31.0%			partial	338	14.2%			partial	73	3.0%	
	perfect core	839	11.8%			perfect core	1518	64.0%			perfect core	1790	75.1%	
NODE 1	TOTAL	3110	100%		NODE 14	TOTAL	2014	100%		NODE 27	TOTAL	2009	100%	
	singleton	1324	42.6%			singleton	208	10.3%			singleton	202	10.1%	
	partial	458	14.7%			partial	242	12.0%			partial	0	0.0%	
	perfect core	1328	42.7%			perfect core	1564	77.7%			perfect core	1807	89.9%	
NODE 2	TOTAL	2771	100%		NODE 15	TOTAL	1902	100%		NODE 28	TOTAL	2208	100%	
	singleton	1270	45.8%			singleton	333	17.5%			singleton	361	16.3%	
	partial	0	0.0%			partial	0	0.0%			partial	52	2.4%	
	perfect core	1501	54.2%			perfect core	1569	82.5%			perfect core	1795	81.3%	
NODE 3	TOTAL	5709	100%		NODE 16	TOTAL	2201	100%		NODE 29	TOTAL	1965	100%	
	singleton	3222	56.4%			singleton	374	17.0%			singleton	155	7.9%	
	partial	1507	26.4%			partial	295	13.4%			partial	0	0.0%	
	perfect core	980	17.2%			perfect core	1532	69.6%			perfect core	1810	92.1%	
NODE 4	TOTAL	5339	100%		NODE 17	TOTAL	2180	100%		NODE 30	TOTAL	2060	100%	
	singleton	2880	53.9%			singleton	374	17.0%			singleton	232	11.3%	
	partial	1201	22.5%			partial	290	13.3%			partial	0	0.0%	
	perfect core	1258	23.6%			perfect core	1537	70.5%			perfect core	1828	88.7%	
NODE 5	TOTAL	5154	100%		NODE 18	TOTAL	2052	100%		NODE 31	TOTAL	2800	100%	
	singleton	2726	52.9%			singleton	238	11.6%			singleton	929	33.2%	
	partial	1151	22.3%			partial	275	13.4%			partial	61	2.2%	
	perfect core	1277	24.8%			perfect core	1539	75.0%			perfect core	1489	53.2%	
NODE 6	TOTAL	2396	100%		NODE 19	TOTAL	1946	100%		NODE 32	TOTAL	2282	100%	
	singleton	345	14.4%			singleton	359	18.4%			singleton	443	19.4%	
	partial	302	12.6%			partial	46	2.4%			partial	61	2.7%	
	perfect core	1749	73.0%			perfect core	1541	79.2%			perfect core	1778	77.9%	
NODE 7	TOTAL	2171	100%		NODE 20	TOTAL	3703	100%		NODE 33	TOTAL	2005	100%	
	singleton	180	8.3%			singleton	1680	45.4%			singleton	209	10.4%	
	partial	95	4.4%			partial	551	14.8%			partial	0	0.0%	
	perfect core	1896	87.3%			perfect core	1472	39.8%			perfect core	1796	89.6%	
NODE 8	TOTAL	2075	100%		NODE 21	TOTAL	2855	100%		NODE 34	TOTAL	2108	100%	
	singleton	214	10.3%			singleton	897	31.4%			singleton	316	15.0%	
	partial	44	2.1%			partial	432	15.1%			partial	0	0.0%	
	perfect core	1817	87.6%			perfect core	1526	53.5%			perfect core	1792	85.0%	
NODE 9	TOTAL	4662	100%		NODE 22	TOTAL	2836	100%		NODE 35	TOTAL	2364	100%	
	singleton	2581	55.4%			singleton	879	31.0%			singleton	525	22.2%	
	partial	751	16.1%			partial	211	7.4%			partial	341	14.4%	
	perfect core	1330	28.5%			perfect core	1746	61.6%			perfect core	1498	63.4%	
NODE 10	TOTAL	4323	100%		NODE 23	TOTAL	2312	100%		NODE 36	TOTAL	1985	100%	
	singleton	2250	52.0%			singleton	440	19.0%			singleton	182	9.2%	
	partial	741	17.2%			partial	82	3.6%			partial	0	0.0%	
	perfect core	1332	30.8%			perfect core	1790	77.4%			perfect core	1803	90.8%	
NODE 11	TOTAL	1880	100%		NODE 24	TOTAL	2167	100%		NODE 37	TOTAL	2207	100%	
	singleton	331	17.6%			singleton	334	15.4%			singleton	377	17.1%	
	partial	0	0.0%			partial	40	1.9%			partial	328	14.8%	
	perfect core	1549	82.4%			perfect core	1793	82.7%			perfect core	1502	68.1%	
NODE 12	TOTAL	4177	100%		NODE 25	TOTAL	2040	100%		NODE 38	TOTAL	2191	100%	
	singleton	2125	50.9%			singleton	236	11.6%			singleton	361	16.5%	
	partial	626	15.0%			partial	0	0.0%			partial	323	14.7%	
	perfect core	1426	34.1%			perfect core	1804	88.4%			perfect core	1507	68.8%	

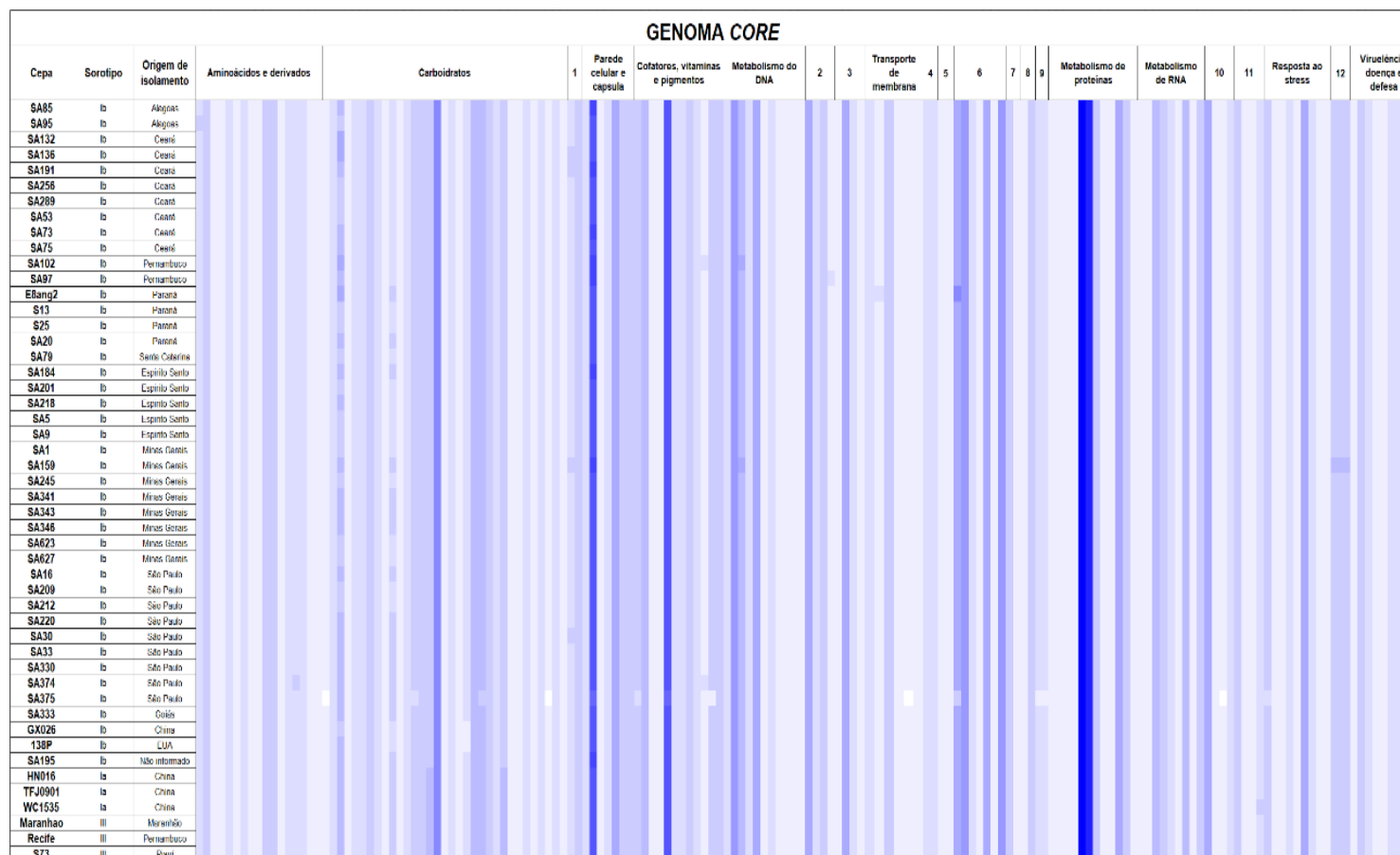
Fonte: Autoria própria (2023)

A anotação dos genomas por meio do RAST permitiu analisá-los quanto aos papéis funcionais dos genes, segundo a base de dados The SEED. Com isso, para o cluster 5 observado na Figura 3, através do programa View Function Profile For FeatureSet, foi possível gerar um *heatmap* a partir papéis funcionais das famílias que integram os três grupos que compõem o pan-genoma de *S. agalactiae*: genoma *core*, compartilhado e exclusivo.

Para melhor visualização, as cepas foram agrupadas conforme o sorotipo, e a origem de isolamento. Entre estes papéis funcionais, incluem-se, por exemplo, as categorias Carboidratos, Metabolismo do DNA, Regulação e sinalização celular, Resposta ao Stress, Virulência, doença e defesa, entre outros.

O *heatmap* do genoma *core* (Figura 4) mostrou homogeneidade em todas as categorias funcionais, visto que este grupo contém as famílias que são consenso entre os todas as cepas analisadas nesse estudo. As categorias que apresentaram maior número de genes foram (I) Carboidratos, nas funções de glicólise e gliconeogênese, (II) Parede celular e cápsula, na função de biossíntese de peptidoglicano, (III) Cofatores, vitaminas e pigmentos, na biossíntese de folato, e (IV) Metabolismo de proteínas, nos componentes envolvidos na subunidade menor e maior dos ribossomos.

Figura 4 - Papéis funcionais dos genes do genoma core obtidos através da anotação dos genomas pelo RAST e pelo programa View Function Profile for FeatureSet



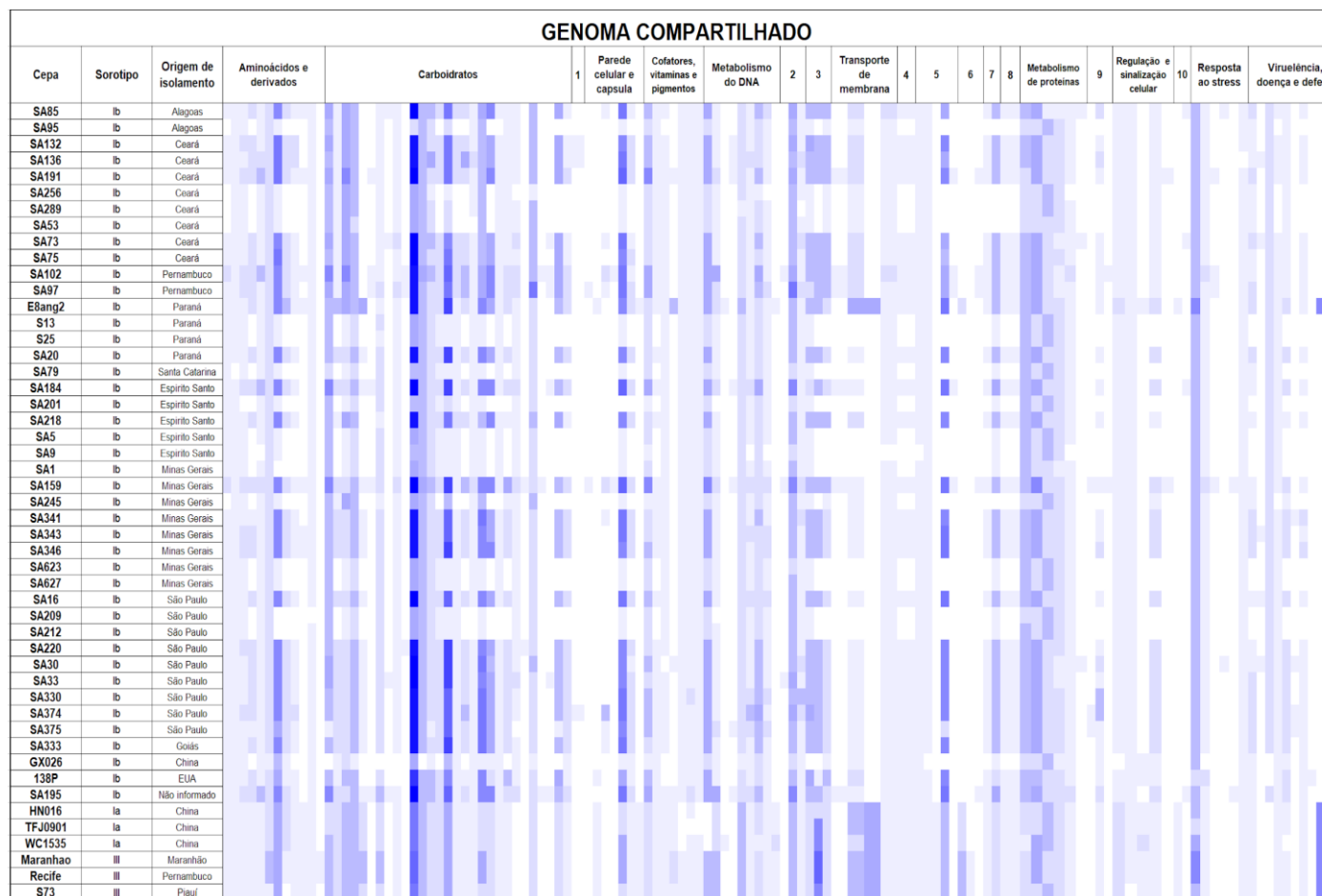
1. Divisão e ciclo celular 2. Ácidos graxos, lipídios e isoprenóides 3. Metabolismo e obtenção de ferro 4. Metabolismo de compostos aromáticos 5. Metabolismo de nitrogênio 6. Nucleosídeos e nucleotídeos 7. Fagos, profagos, elementos transponíveis e plasmídeos 8. Metabolismo do fósforo 9. Metabolismo do potássio 10. Regulação e sinalização celular 11. Respiração 12. Metabolismo do enxofre

Fonte: Autoria própria (2023)

Em contrapartida, no *heatmap* do genoma compartilhado (Figura 5), é possível observar a heterogeneidade característica desse grupo. Alguns papéis funcionais apresentaram, visualmente, diferenças significativas na quantidade de genes presentes nas cepas, como por exemplo na categoria (I) Aminoácidos e derivados, na função de biossíntese de metionina, categoria (II) Carboidratos, nas funções de fermentação ácida mista, metabolismo do glicogênio e utilização de lactose, galactose, maltose e maltodextrina, (III) Parede celular e cápsula, no metabolismo de ácido siálico, (IV) Cofatores, vitaminas e pigmentos, na biossíntese de biotina, (V) Metabolismo do DNA, em funções envolvidas com CRISPRs, (VI) Metabolismo e obtenção de ferro, em funções envolvidas no metabolismo do ferro e utilização de heme e hemina, (VII) Nucleosídeos e nucleotídeos, em conversões de purinas, e na categoria (VIII) Virulência, doença e defesa, em funções relacionadas a tolerância a vancomicina e zona recombinatória de *S. pneumoniae*.

Por fim, no *heatmap* do genoma exclusivo (Figura 6) é possível analisar os papéis funcionais dos genes exclusivos de cada cepa e, de forma muito clara, observa-se que todas as cepas possuem vários genes relacionados no metabolismo de proteínas, e em funções envolvidas na subunidade maior dos ribossomos. Também, percebe-se que os genes exclusivos estão distribuídos entre diversas categorias, sendo essas essenciais para a sobrevivência celular, como metabolismo de DNA, RNA e proteínas, parede celular e nucleotídeos, mas também em funções que podem conferir vantagens à essas bactérias, como virulência, doença e defesa, e resposta ao stress. Ainda nessa categoria, foi possível constatar que cerca de 56,39% das famílias são proteínas hipotéticas, desconhecidas ou regiões codificadoras ainda não anotadas, mostrando que parte significativa do genoma exclusivo ainda é pouco caracterizado.

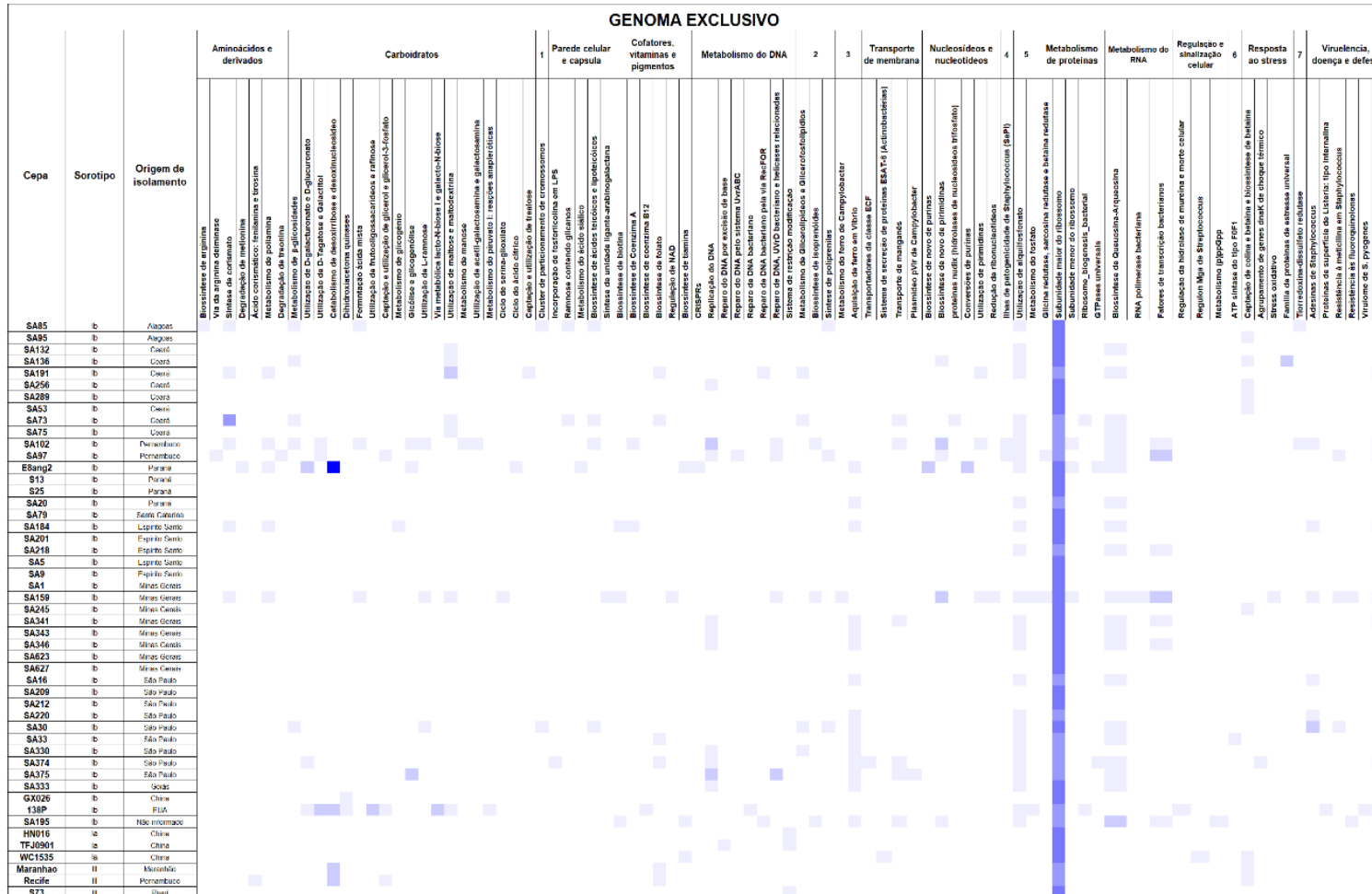
Figura 5 - Papéis funcionais dos genes do genoma compartilhado obtidos através da anotação dos genomas pelo RAST e pelo programa View Function Profile for FeatureSet



1. Divisão e ciclo celular 2. Ácidos graxos, lipídios e isoprenóides 3. Metabolismo e obtenção de ferro 4. Metabolismo de compostos aromáticos 5. Nucleosídeos e nucleotídeo 6. Fagos, profagos, elementos transponíveis e plasmídeos 7. Metabolismo do fósforo 8. Metabolismo do potássio 9. Metabolismo do RNA 10. Respiração

Fonte: Autoria própria (2023)

Figura 6 - Papéis funcionais dos genes do genoma exclusivo obtidos através da anotação dos genomas pelo RAST e pelo programa View Function Profile for FeatureSet

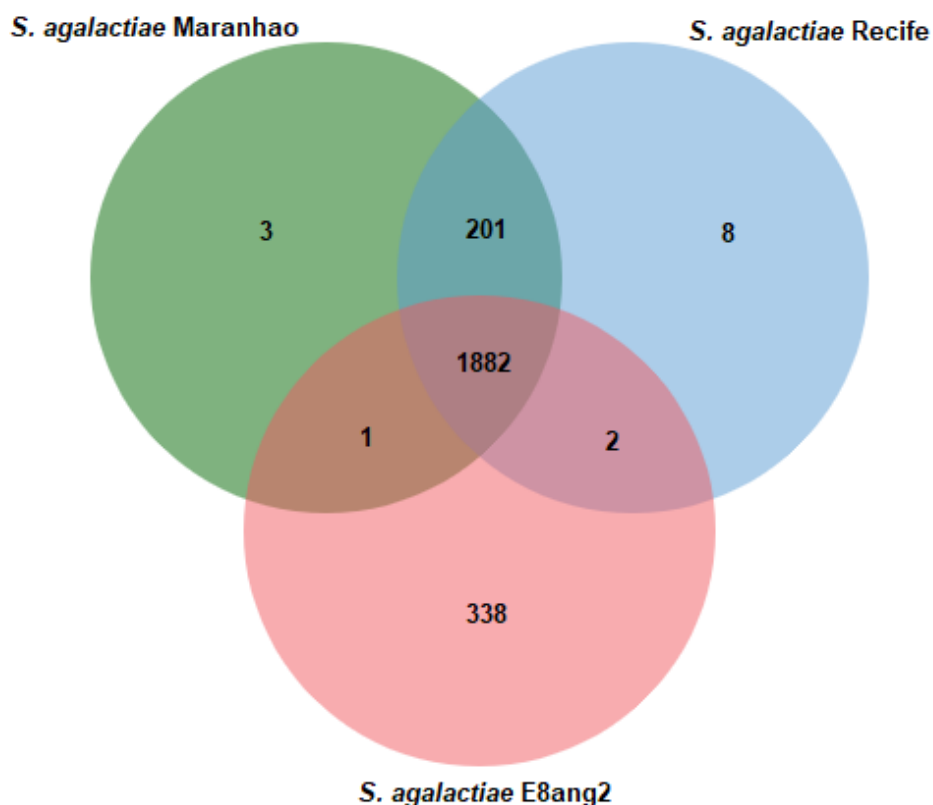


1. Divisão e ciclo celular 2. Ácidos graxos, lipídios e isoprenóides 3. Metabolismo e obtenção de ferro 4. Fagos, profagos, elementos transponíveis e plasmídeos 5. Metabolismo do fósforo 6. Respiração 7. Metabolismo do enxofre

Fonte: Autoria própria (2023)

As 3 cepas isoladas e sequenciadas pelo laboratório de Micologia Veterinária (UEL), Maranhão, Recife e E8ang2, foram selecionadas para comparação e visualização dos *clusters* ortólogos a fim de maior compreensão acerca dessas cepas recém isoladas e ainda não publicadas. Com a utilização da plataforma OrthoVenn2, obteve-se os dados apresentados na Figura 7 e Tabela 3. Os três genomas compartilham 1882 *clusters* ortólogos, sendo que destes, 1877 são clusters de gene de cópia única, ou seja, onde cada genoma possui apenas um gene incluso. Por outro lado, as cepas Maranhão, Recife e E8ang2 possuem 3, 8 e 350 proteínas em seu genoma exclusivo, respectivamente; todos os genes exclusivos de Maranhão e Recife, e 328 genes da E8ang não apresentaram ortólogos, e as 22 proteínas restantes foram agrupadas em 10 clusters dentro do próprio genoma da E8ang2. Além disso, Maranhão e Recife compartilham entre si 201 clusters ortólogos, o que mostra alta similaridade entre as duas cepas.

Figura 7 - Diagrama de Venn mostrando os *clusters* compartilhados e exclusivos das cepas Maranhão, Recife e E8ang2, construído através da plataforma OrthoVenn2



Fonte: Autoria própria (2023)

Tabela 3 - Contagem de *clusters* e proteínas na comparação das cepas Maranhão, Recife e E8ang2 obtida através da plataforma OrthoVenn2

Categoria do pan-genoma	Cepas inclusas	Contagem de <i>clusters</i>	Contagem de proteínas			
			Total	Maranhão	Recife	E8ang2
Core	Maranhão, Recife e E8ang2	1882	5653	1883	1883	1887
	Recife e E8ang2	2	4	-	2	2
Compartilhado	Maranhão e E8ang2	1	2	1	-	1
	Maranhão e Recife	201	411	205	206	-
	Maranhão	3	3	3	-	-
Exclusivo	Recife	8	8	-	8	-
	E8ang2	338	350	-	-	350
TOTAL		2435	6431	2092	2099	2240

Fonte: Autoria própria (2023)

Dos 328 genes exclusivos da cepa E8ang2, que não possuem ortólogos, 23,17% são regiões que codificam proteínas hipotéticas ou ainda não caracterizadas. Encontrou-se, também, vários genes relacionados a funções essenciais, como transportadores, proteínas de membrana e enzimas. Quanto a funções que podem conferir vantagem a cepa, foram localizados 3 genes em *tandem* que codificam as proteínas que compõe o pili, mais especificamente proteínas da ilha genômica PI-2b, além de dois genes para a proteína sortase; também foram encontrados vários genes relacionados a fagos, localizados todos em um mesmo *locus*, além de regiões referentes a elementos transponíveis e CRISPRs espalhados pelo genoma. Os clusters exclusivos da cepa E8ang2 também incluem o gene *PbsP*, relacionado a virulência, e duas regiões associadas a bacteriocinas.

Já nos clusters compartilhados entre Maranhão e Recife, também foram encontradas várias proteínas hipotéticas e genes *housekeeping*. Assim como a E8ang2, essas duas cepas também apresentaram genes para o pili, mas codificados em ilhas genômicas diferentes: PI-1 e PI-2a. Além disso, exibiram 3 genes em *tandem* relacionados ao pili do tipo IV, que codificam uma forma menor dessa estrutura (pseudo pili). Foram localizados diversos genes associados a elementos transponíveis e cerca de 15 relacionados a fagos. As duas cepas apresentaram 5 *clusters* referentes ao gene *bca*, proteína C- α , e outros 3 clusters do domínio Rib, que podem estar ligados a virulência e imunidade.

4 DISCUSSÃO

S. agalactiae se tornou uma bactéria de grande interesse clínico desde que foi diferenciada de outros *Streptococcus* pela primeira vez em 1933, isolada de mastite bovina (LANCEFELD, 1993). Os primeiros casos de infecções invasivas em humanos foram relatados em torno da década de 60 (EICKHOFF et al. 1964; LAZARUS; SELLERS; MARINE, 1965), mesma época em que também foram descritos casos dessa bactéria em peixes. Nesses animais, os primeiros relatos de infecções causadas por *S. agalactiae* foram em Carpas douradas (ROBINSON; MEYER, 1966), e após isso, essa bactéria foi isolada de várias outras espécies, como truta-arco-íris, yellowtail, goraz e tilápia (AMAL; ZAMRI-SAAD, 2011). Em tilápias, essa bactéria pode causar altas taxas de mortalidade e perdas econômicas em pisciculturas (MAULU, S. et al., 2021).

Atualmente, dos 1.677 relatórios de montagem e anotação de genomas de *S. agalactiae* encontrados no NCBI, cerca de 90 são de cepas isoladas de tilápia do Nilo, incluindo genomas completos, contigs, scaffolds e cromossomos. Esse grande volume de dados é resultado da diminuição dos custos e do tempo de execução dos sequenciamentos, que conseqüentemente, também permitiu um aumento das análises genômicas comparativas (GIANI et al., 2020; ROULI et al., 2015). Neste estudo, a comparação dos genomas de 49 cepas de *S. agalactiae* isoladas de tilápia do Nilo, sendo 46 genomas completos retirados do NCBI e 3 sequenciados recentemente pelo Laboratório de genética e taxonomia de bactéria da UEL, permitiu analisar as semelhanças e diferenças das cepas isoladas de uma mesma espécie hospedeira, além de ressaltar características comuns em bactérias desta espécie já descritas em outros trabalhos.

Dos genomas selecionados para essa pesquisa, 89,8% são de cepas isoladas do Brasil e 8,1% da China; ambos estão entre os países com maior produção de tilápia do Nilo, onde a China ocupa o primeiro lugar e o Brasil, quarto lugar. Em ambos os países, a tilápia possui destaque na aquicultura e importante contribuição na economia local, como o exemplo do Brasil que faturou US\$18,2 milhões em 2021 com as exportações desse peixe (PEIXE BR, 2022). Logo, relatos de surtos de estreptococose e isolamento de novas cepas de *S. agalactiae* tem sido amplamente descritos nos últimos anos pelos dois países, além de contribuições acerca do desenvolvimento de vacinas (LIU et al., 2016; PRETTO-GIORDANO et al., 2010),

caracterização da virulência e patogenicidade (MIAN et al., 2009; ZHANG et al., 2019), e classificação e genotipagem (GODOY et al., 2013; YE et al., 2011), por exemplo.

Em relação à sorotipagem, todas as cepas, com exceção de uma que não apresenta essa informação, são dos sorotipos Ia (6,12%), Ib (85,71%) ou III (6,12%). Estudos prévios, como de Godoy et al. (2013) e Evans et al. (2008) mostraram que os sorotipos Ia e Ib são os mais comuns isolados de peixes, corroborando com os resultados obtidos no presente estudo, onde foi possível observar entre os genomas uma elevada proporção (91,83%) de isolados com estes dois sorotipos. Apesar disso, o aumento de relatos nos últimos 10 anos mostra que o sorotipo III também tem grande importância para esses animais (KAYANSAMRUJ et al., 2019; LI et al., 2013, SYUHADA et al., 2020). Devido a isso, a caracterização do sorotipo capsular de isolados de peixes se torna imprescindível, pois pode auxiliar a compreender o padrão de surtos estreptocócicos, como por exemplo, quais os principais sorotipos causadores de doenças em tilápias (YE et al., 2011), e quais as principais características patogênicas de cada sorotipo (ZHANG et al., 2018).

A classificação das cepas conforme o sorotipo capsular foi observada na árvore filogenética (Figura 1), onde nela as cepas encontram-se divididas em dois grupos. Essa tendência à subdivisão da árvore em clados foi também observada em outros estudos, como o de Brochet et al. (2006) e de Gori et al. (2020), onde ambos analisaram cepas isoladas tanto de animais como de humanos. Além disso, no presente estudo, o Grupo I contemplou todas as cepas dos sorotipos Ia e III. As unidades de repetição de polissacarídeos desses dois sorotipos são idênticas em relação a composição e estrutura, sendo diferenciadas apenas pela ligação glicosídica entre as unidades (CIESLEWICZ et al., 2005), podendo explicar a semelhança entre as cepas.

Já em relação ao MLST, 19 genomas inclusos nessa pesquisa foram tipados por esse método, que é amplamente utilizado para classificação de cepas de *S. agalactiae* (JONES et al., 2003). Dos 30 genomas que não apresentaram ST, dois não foram submetidas ao MLST até o momento, quatro não apresentam essa informação na literatura, e 24 genomas apresentam uma deleção parcial no locus *glcK*, região genômica que faz parte do esquema de tipagem de *S. agalactiae* e, por isso, não foi possível a realização da tipagem por MLST; esse evento de deleção foi relatado por Barony et al. (2017) quando as mesmas 24 cepas foram estudadas no contexto evolucionário dentro do território brasileiro.

Na árvore filogenética (Figura 1) é possível observar um agrupamento por ST onde as cepas SA53, SA75, SA289, SA256, SA245, SA136 e SA73 (ST-260) fizeram parte de um mesmo clado, bem como cepas SA218, SA95 e SA85 (ST-927) compuseram outro clado. Além disso, 17 das 24 cepas sem ST do sorotipo Ib também foram agrupadas em um mesmo clado. Esse mesmo padrão de agrupamento foi visto no estudo de Barony et al. (2017) com as mesmas cepas, estudo no qual apesar de ter utilizado uma metodologia baseada em *whole-genome* MLST (analisa uma maior parcela do genoma), obteve resultados muito semelhantes com os obtidos no presente estudo.

Pode-se também, com a filogenia gerada, correlacionar a formação de clusters com a origem geográfica do isolado, principalmente para cepas brasileiras, o que pode ser devido a proporção de cepas brasileiras em relação as dos demais países. A tendência de agrupamentos foi principalmente nas grandes regiões brasileiras: Nordeste, Sudeste e Sul. Por outro lado, observou-se também que cepas com origens geográficas distintas se agruparam em um mesmo clado, como a cepa SA245 (MG) compartilhando um mesmo ancestral que as cepas SA256, SA289, SA75, SA136, SA73 e SA53 (CE); e a cepa SA218 (ES) que é mais semelhante as cepas SA85 e SA95 (AL) do que outras isoladas na região Sudeste. Como mostrado por Barony et al. (2017), esses achados podem ser explicados pelo deslocamento de peixes pelos estados brasileiros e também pela forma que a estreptococose pode ser transmitida, no contato direto entre peixes doentes e peixes saudáveis (PRADEEP et al., 2016).

Com a visualização circular da sobreposição dos grupos de genes dos 49 genomas alinhados em relação ao genoma de base (WC1535) (Figura 2), foi possível observar uma alta taxa de conservação entre as regiões genômicas. Estudos como de Tettlin et al. (2005) e Brochet et al. (2006) já mostraram que essa espécie possui um genoma base ou fundamental (do inglês *chromosomal backbone*) altamente conservado, intercalado com ilhas genômicas com maiores variações, que são responsáveis por conferir a diversidade genética entre as cepas.

A partir do pan-genoma gerado, observou-se que o genoma *core* das cepas analisadas neste estudo possui principalmente genes com papéis fundamentais para a sobrevivência e funcionamento celular de um procarionto. Além destes, foram amplamente encontrados genes relacionados a glicólise, gliconeogênese, biossíntese de peptidoglicano e as subunidades dos ribossomos.

Além de papel funcional relacionado com obtenção de energia pela célula (TORTORA et al., 2017), o processo de gliconeogênese mostrou-se estar envolvido na virulência de várias bactérias patogênicas, como *Mycobacterium tuberculosis* (PUCKETT et al., 2014) e *Francisella tularensis* (BRISSAC et al., 2015), e inclusive em *S. agalactiae*, para o qual Liu et al. (2019) mostraram que genes relacionados a via do fosfato e piruvato estavam diferencialmente expressos em cepas virulentas e não virulentas de GBS isoladas de tilápias, interferindo no crescimento celular e, conseqüentemente, na virulência das bactérias.

Os peptidoglicanos são componentes essenciais da parede celular da maioria das bactérias, em *S. agalactiae* essa estrutura está covalentemente ligada ao antígeno de superfície B, que define essa bactéria como um Streptococcus do grupo B, e aos polissacarídeos capsulares, importante fator de virulência dessa espécie (CALIOT et al., 2012; DENG et al., 2000). A biossíntese do peptidoglicano é complexa, e envolve cerca de 20 processos que ocorrem no citoplasma, interior e exterior da membrana plasmática, desde a formação dos açúcares e peptídeos que formam sua estrutura até a montagem na superfície celular, além de envolverem uma grande diversidade de enzimas (BARRETEAU et al., 2008); isso pode explicar tanto a quantidade de genes envolvidos com a biossíntese dessa estrutura, como também a elevada quantidade de genes envolvidos na estrutura dos ribossomos encontrados no genoma *core*, isso porque em procariotos, além dos RNAs ribossomais, os ribossomos são compostos por cerca de 52 proteínas diferentes (ZIMMERMANN et al., 2003).

Por outro lado, ao analisar os papéis funcionais dos principais genes encontrados no genoma compartilhado, pode-se observar que estes estão mais relacionados a características de vantagem do que funções essenciais para o funcionamento celular, como já visto em outros estudos de pan-genoma do gênero *Streptococcus* (TETTELIN et al., 2005; SINHA et al., 2021), o que faz com esses genes sejam também chamados de genoma dispensável. Temos como exemplo regiões genômicas relacionados a CRISPRs que se mostraram ausentes nas cepas do sorotipo Ia, enquanto estavam presentes em diferentes quantidades no restante das cepas estudadas. Ma et al. (2018) e Dong et al. (2021) mostraram que genes associados a CRISPRs, ao reprimirem reguladores transcricionais, podem aumentar a capacidade de invasão, adesão, hemólise, citotoxicidade e sobrevivência intracelular de *S. agalactiae*, além de aumentar a permeabilidade à barreira

hematoencefálica em animais infectados, sendo importantes regiões genômicas envolvidas com a virulência dessa espécie.

Diversos estudos demonstraram que cepas de *S. agalactiae* isoladas de diferentes tipos de hospedeiros tendem a apresentar sensibilidade ao antibiótico vancomicina (ALI et al., 2010; MIRANDA et al., 2018; PINTO et al., 2013), corroborando com nossos dados, uma vez que para 85,7% das cepas analisadas nesse estudo tais genes de resistência estiveram ausentes. Por outro lado, todas as cepas dos sorotipos Ia, III e a E8ang2 (Ib), apresentaram genes relacionados a tolerância à vancomicina. Este achado pode sugerir que a resistência a esta droga pode se tornar um problema epidemiológico, visto que a Vancomicina é um importante antibiótico usado para controle de doenças causadas por *S. agalactiae*, além de que evidencia a propriedade bacteriana de transferência horizontal de genes que podem conferir vantagens a célula (RICHARDS et al., 2011).

Quando observado os dados estatísticos gerados pela análise do pan-genoma do grupo estudado, observou-se que estes possuem 2.726 famílias exclusivas. Uma família exclusiva é aquela onde apenas um genoma possui genes incluídos e, cada família, pode possuir um ou mais genes. A quantidade média de genes exclusivos de cepa foi de 120, onde o menor número foi encontrado na cepa SA218, com 55 genes, e o maior na cepa WC1535, que contém 217 genes exclusivos. Esses números estão muito acima daqueles reportados por outros estudos de pan-genoma de *S. agalactiae*, como de Tettelin et al. (2005), mas pode ser explicado pela diferente metodologia empregada para analisar os genomas e até mesmo a quantidade de genomas adicionados à análise. Também, vale destacar que o estudo de Tettelin et al. (2005) foi o pioneiro desse modelo de análise de genômica comparativa, além de ter sido publicado há quase 20 anos com programas e algoritmos que funcionavam para a pequena quantidade de dados disponíveis no período.

Nesse primeiro estudo de 2005 foram analisados 8 genomas de *S. agalactiae* para investigação do pan-genoma e, através de modelos matemáticos baseados em seus achados, concluíram que cerca de 33 novos genes seriam identificados e adicionados à análise à medida que uma nova cepa dessa bactéria fosse sequenciada, ou seja, que cada cepa teria em torno de 33 genes exclusivos. Esse resultado trouxe pela primeira vez, para essa espécie, o conceito de pan-genoma aberto. Em contrapartida, na presente pesquisa a média de genes exclusivos foi de 120 para cada novo genoma. O algoritmo OrthoMCL (LI et al., 2003), utilizado para

identificação das famílias ortólogas e cálculos do pan-genoma, comparou os genomas através de uma estratégia baseada no BLASTp, onde todas as proteínas são comparadas contra todas (do inglês *all-against-all*) para identificação de ortólogos e parálogos recentes, que em seguida são segregados através do algoritmo de *Markov Cluster*. É importante ressaltar que essa comparação de proteínas, mais especificamente a semelhança entre elas, deriva de um *p-value* que se baseia na semelhança e comprimento das sequências alinhadas, então, se duas sequências com a mesma função forem alinhadas, mas uma delas apresentar uma deleção em uma região, por exemplo, elas podem ser consideradas diferentes uma da outra por não atingirem os valores limites que as consideram ortólogas.

É por essa razão que os resultados se mostram muito diferentes quando comparados com outros estudos de pan-genoma. Por exemplo, o gene que codifica a proteína L24p da subunidade maior do ribossomo aparece não somente no genoma *core*, mas como um gene exclusivo em 6 outras cepas que, devido a diferenças nas sequências, não atingiram o *p-value* e por isso foram considerados parte do genoma exclusivo. Essa mesma situação aconteceu com outras proteínas do ribossomo, como a L33p, L34p e L36p, que apareceram tanto no genoma *core* quanto no exclusivo e, neste caso, explicam a elevada quantidade de genes relacionados a ribossomos no genoma específico de cepa (Figura 6).

Essa limitação no algoritmo OrthoMCL também impactou no tamanho do genoma *core*, que seria em torno de 1277 famílias, e em sua análise, visto que várias famílias encontradas no genoma acessório também fazem parte do *core*, o que necessitaria de análise individual de todos os genes inclusos nos genomas compartilhado e exclusivo. Além disso, o programa traz o tamanho desse genoma em famílias, que podem conter um ou mais genes, o que torna difícil a comparação com outros trabalhos, como o de Tettelin et al. (2005), que mostrou que o genoma *core* dessa espécie possui em torno de 1.800 genes, e de Gori et al. (2020) que apresenta o valor de 1.374 genes como representantes desse genoma compartilhado por todos os membros da análise.

O algoritmo OrthoMCL, ainda que tenha mostrado essa limitação no agrupamento de proteínas, é uma ferramenta que tem sido amplamente utilizada desde o seu desenvolvimento no início do milênio até os dias atuais, a qual já foi citada por mais de 4.050 trabalhos até a presente data (fevereiro de 2023) segundo a base

de dados Web of Science Clarivate Analytics™ (<https://clarivate.com/webofsciencegroup/solutions/web-of-science/>).

É fundamental destacar que cerca de 56% das famílias consideradas exclusivas compreendem genes que codificam proteínas hipotéticas, regiões codificadoras ainda não anotadas ou com pouco conhecimento acerca da função, mostrando que parte do pan-genoma dessa espécie ainda é desconhecido ou pouco estudado, e pode conter características importantes. Tais dados estão em conformidade com os encontrados na literatura: Sinha et al. (2021), investigando o pan-genoma de *Streptococcus intermedius*, também verificaram que 55,6% dos genes do genoma exclusivo não estavam descritos na base de dados *Clusters of Orthologous Groups* (COG) e conseqüentemente tinham função desconhecida, assim como Tettelin et al. (2005), que concluíram que cerca de um terço dos genes do genoma acessório das cepas *S. agalactiae* de seu estudo codificavam proteínas de função hipotética.

Além disso, apesar das cepas terem sido agrupadas de acordo com o sorotipo nas análises filogenéticas, tanto na inferência por Máxima Verossimilhança (Figura 1) quanto na árvore baseada no pan-genoma (Figura 3), não foi possível observar tendências quanto a esta classificação na distribuição dos papéis funcionais das famílias do genoma acessório, mostrando que o sorotipo capsular não reflete na diversidade genética dessa espécie, resultado que também foi demonstrado por Tettelin et al. (2005).

A análise comparativa das 3 cepas isoladas pelo laboratório de Micologia Veterinária (UEL) revelou importantes características genômicas que diferenciam essas 3 bactérias. A presença de genes que codificam o pili foi observada em todos os genomas, com a diferença de que Maranhão e Recife possuem regiões que codificam o pili do tipo PI-1 e PI-2a, além de genes do pseudo pili, enquanto E8ang codifica apenas o pili do tipo PI-2b. Genes dessa estrutura podem estar localizados dentro de duas diferentes ilhas genômicas, PI-1 e PI-2, sendo que a segunda ilha pode ter duas variantes: PI-2a e PI-2b; segundo Xia et al. (2015), o tipo PI-2a é vastamente encontrado em cepas de diversos sorotipos, enquanto o PI-2b é restrito a poucos isolados clínicos dessa espécie. De forma geral, o pili é um importante fator de virulência relacionado ao aumento da adesão, translocação e formação de biofilme (SHABAYEK; SPELLERBERG, 2018), mas vários estudos já mostraram características distintas acerca de cada variante dessa estrutura. Por exemplo, Jiang

et al. (2012) mostrou que o PI-1 tem papel importante na evasão do sistema imunológico inato do hospedeiro, mas que possui pouca contribuição na adesão à tecidos. Já o PI-2a se mostrou amplamente envolvido na aderência e formação de biofilme (RINAUDO et al., 2010), enquanto reportou-se que o PI-2b aumentou a sobrevivência das células em macrófagos (CHATTOPADHYAY et al., 2011) e promoveu maior taxa de invasão nas células do hospedeiro (LAZZARIN et al., 2017). Visto que as cepas analisadas possuem diferentes tipos de pili, essa característica pode fazer com que Maranhão e Recife tenham mecanismos de invasão distintos quando comparadas a E8ang2, interferindo na patogenicidade das 3 cepas.

Ainda em relação a patogenicidade, diferentes genes relacionados a virulência foram encontrados nos genes exclusivos da E8ang e nos genes compartilhados de Maranhão e Recife. O gene *PbsP*, que codifica a proteína de ligação ao plasminogênio, está incluso nos genes exclusivos da E8ang2; essa proteína de adesão tem papel importante na patogenicidade de *S. agalactiae* e pode aumentar a capacidade de invasão e colonização de células do hospedeiro ao ligar-se ao plasminogênio (BUSCETTA et al., 2016). Já a bacteriocina, também encontrada nessa cepa, é um tipo de peptídeo antimicrobiano que, ao reprimir o crescimento de outras espécies bacterianas, pode favorecer a infecção e crescimento do patógeno que a produziu (VOGEL; SPELLERBERG, 2021); esse tipo de toxina não é comumente encontrado em cepas de *S. agalactiae*, além de haver poucos relatos na literatura de caracterização de bacteriocinas encontradas nessa espécie (WESCOMBE; TAGG, 2003; VIDAL AMARAL et al., 2022).

Entre os fatores de virulência compartilhados pela Maranhão e Recife, pode-se citar o gene *bca* e o domínio Rib, ambos codificando produtos que fazem parte do grupo de proteínas do tipo-alpha (do inglês *alpha-like proteins*), amplamente encontradas em cepas de *S. agalactiae*. Proteínas tipo-alpha ficam localizadas na superfície da célula e possuem domínios com alta similaridade, o que faz que com sejam grandes candidatas no desenvolvimento de vacinas (GABRIELSEN et al., 2017). As proteínas do tipo-alpha podem estar relacionadas com o aumento da virulência, já que funcionam como adesinas, e por isso podem facilitar a invasão e colonização no hospedeiro (BARON et al., 2007).

A diferente composição dos genes exclusivos da cepa E8ang2 e daqueles compartilhados entre Maranhão e Recife podem conferir vantagens distintas à essas linhagens, e conseqüentemente, podem interferir na patogenicidade, visto que em

estudos prévios [dados não publicados] as 3 cepas mostraram sinais clínicos e taxas de mortalidade muito diferentes quando foram inoculadas em tilápias do Nilo.

5 CONSIDERAÇÕES FINAIS

O estudo do pan-genoma de 49 cepas de *S. agalactiae* isoladas de tilápia do Nilo permitiu fazer uma análise comparativa preliminar e geral das principais características em relação a esse grupo de bactérias. Foi possível confirmar aspectos já consolidados sobre o pan-genoma dessa espécie, como a composição do genoma *core* por proteínas essenciais às células e a presença de diversos genes exclusivos que fazem com que o pan-genoma de *S. agalactiae* seja considerado aberto. Além disso, observou-se que, apesar de não haver tendência quanto ao sorotipo capsular na distribuição dos genes no genoma acessório, a relação filogenética acaba agrupando cepas do mesmo sorotipo; também se verificou uma tendência quanto a origem de isolamento e o tipo de sequência nos clados da árvore filogenética.

A análise também permitiu investigar com maior profundidade as semelhanças e diferenças de 3 cepas recentemente isoladas, Maranhão, Recife e E8ang2, e indicou várias regiões genômicas que podem conferir vantagens e torná-las cepas virulentas e com maior capacidade de infecção, seja pela presença de genes de virulência, ou pela maior quantidade de genes relacionados a funções essenciais.

Além disso, limitações dentro das ferramentas escolhidas mostram a necessidade de análises adicionais, principalmente que investiguem mais a fundo os genomas compartilhado e exclusivo, para que assim, mais informações sejam levantadas acerca desse grupo de microrganismos. A plataforma usada para as análises (Kbase), possui interface de fácil acesso e utilização, permitindo explorar o conteúdo genômico de todas as 49 cepas de uma forma ampla e superficial. Contudo, necessita-se de análises mais aprofundadas por meio de outros programas que permitam investigações mais completas.

De qualquer forma, destaca-se a importância de estudar cepas de *S. agalactiae* isoladas de tilápia do Nilo devido ao grande impacto que a doença causada por essa espécie bacteriana provoca em pisciculturas em todo o mundo. Os dados obtidos serão importantes para futuros estudos, principalmente para uma melhor compreensão dos mecanismos de patogenicidade, que poderão impactar diretamente na implementação de melhores estratégias de manejo da Estreptococose, e conseqüentemente, poderão reduzir o impacto econômico e ambiental nos sistemas de criação. Além disso, a vasta quantidade de genomas disponíveis em banco de dados públicos torna possível a comparação de grandes conjuntos de dados,

principalmente informações genômicas que podem conter informações essenciais para entender o processo de infecção desses patógenos.

REFERÊNCIAS

- ABU-ELALA, N. M.; SAMIR, A.; WASFY, M.; ELSAYED, M. Efficacy of injectable and immersion polyvalent vaccine against streptococcal infections in broodstock and offspring of Nile Tilapia (*Oreochromis niloticus*). **Fish & shellfish immunology**, v. 88, p. 293-300, 2019.
- ALI, A.; HASSAN, D.; SALEHA, A. A.; SITI, K. B.; MILUD, A. *Streptococcus agalactiae* the etiological agent of mass mortality in farmed red tilapia (*Oreochromis* sp.). **Journal of Animal and Veterinary Advances**, v. 9, n. 20, p. 2640-2646, 2010.
- AMAL, M. N. A.; ZAMRI-SAAD, M. Streptococcosis in tilapia (*Oreochromis niloticus*): a review. **Pertanika Journal of Tropical Agricultural Science**, v. 34, n. 2, p. 195-206, 2011.
- ARKIN, A. P.; COTTINGHAM, R. W.; HENRY, C. S.; HARRIS, N. L.; STEVENS, R. L.; MASLOV, S.; ... YU, D. KBase: the United States department of energy systems biology knowledgebase. **Nature biotechnology**, v. 36, n. 7, p. 566-569, 2018.
- BARATO, P.; MARTINS, E. R.; VASQUEZ, G. M.; RAMIREZ, M.; MELO-CRISTINO, J.; MARTÍNEZ, N.; IREGUI, C. Capsule impairs efficient adherence of *Streptococcus agalactiae* to intestinal epithelium in tilapias *Oreochromis* sp. **Microbial pathogenesis**, v. 100, p. 30-36, 2016.
- BARON, M. J.; FILMAN, D. J.; PROPHETE, G. A.; HOGLE, J. M.; MADOFF, L. C. Identification of a glycosaminoglycan binding region of the alpha C protein that mediates entry of group B streptococci into host cells. **Journal of biological chemistry**, v. 282, n. 14, p. 10526-10536, 2007.
- BARONY, G. M.; TAVARES, G. C.; PEREIRA, F. L.; CARVALHO, A. F.; DORELLA, F. A.; LEAL, C. A.; FIGUEIREDO, H. C. Large-scale genomic analyses reveal the population structure and evolutionary trends of *Streptococcus agalactiae* strains in Brazilian fish farms. **Scientific reports**, v. 7, n. 1, p. 1-10, 2017.
- BARRETEAU, H.; KOVAČ, A.; BONIFACE, A.; SOVA, M.; GOBEC, S.; BLANOT, D. Cytoplasmic steps of peptidoglycan biosynthesis. **FEMS microbiology reviews**, v. 32, n. 2, p. 168-207, 2008.
- BØGWALD, J.; DALMO, R. A. Review on immersion vaccines for fish: An update 2019. **Microorganisms**, v. 7, n. 12, p. 627, 2019.
- BRETTIN, T.; DAVIS, J. J.; DISZ, T.; EDWARDS, R. A.; GERDES, S.; OLSEN, G. J.; ... XIA, F. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. **Scientific reports**, v. 5, n. 1, p. 1-6, 2015.
- BRISSAC, T.; ZIVERI, J.; RAMOND, E.; TROS, F.; KOCK, S.; DUPUIS, M.; ... CHARBIT, A. Gluconeogenesis, an essential metabolic pathway for pathogenic *Francisella*. **Molecular microbiology**, v. 98, n. 3, p. 518-534, 2015.

- BROCHET, M.; COUVÉ, E.; ZOUINE, M.; VALLAEYS, T.; RUSNIOK, C.; LAMY, M. C.; ... GLASER, P. Genomic diversity and evolution within the species *Streptococcus agalactiae*. **Microbes and Infection**, v. 8, n. 5, p. 1227-1243, 2006.
- BURNHAM, C. D.; TYRRELL, G. J. Virulence factors of group B streptococci. **Reviews in Medical Microbiology**, v. 14, n. 4, p. 109-118, 2003.
- BUSCETTA, M.; FIRON, A.; PIETROCOLA, G.; BIONDO, C.; MANCUSO, G.; MIDIRI, A.; BENINATI, C. PbsP, a cell wall-anchored protein that binds plasminogen to promote hematogenous dissemination of group B *Streptococcus*. **Molecular Microbiology**, v. 101, n. 1, p. 27-41, 2016.
- CAIPANG, C. M. A.; LUCANAS, J. B.; LAY-YAG, C. M. Updates on the vaccination against bacterial diseases in tilapia, *Oreochromis spp.* and Asian seabass, *Lates calcarifer*. **Aquaculture, Aquarium, Conservation & Legislation**, v. 7, n. 3, p. 184-193, 2014.
- CALIOT, E.; DRAMSI, S.; CHAPOT-CHARTIER, M. P.; COURTIN, P.; KULAKAUSKAS, S.; PECHOUX, C.; ... MISTOU, M. Y. Role of the Group B antigen of *Streptococcus agalactiae*: a peptidoglycan-anchored polysaccharide involved in cell wall biogenesis. **PLoS pathogens**, v. 8, n. 6, p. e1002756, 2012.
- CAPUTO, A.; FOURNIER, P. E.; RAOULT, D. Genome and pan-genome analysis to classify emerging bacteria. **Biology direct**, v. 14, n. 1, p. 1-9, 2019.
- CAPUTO, A.; MERHEJ, V.; GEORGIADES, K.; FOURNIER, P. E.; CROCE, O.; ROBERT, C.; RAOULT, D. Pan-genomic analysis to redefine species and subspecies based on quantum discontinuous variation: the *Klebsiella* paradigm. **Biology Direct**, v. 10, n. 1, p. 1-12, 2015.
- CHANG, Y. C.; OLSON, J.; BEASLEY, F. C.; TUNG, C.; ZHANG, J.; CROCKER, P. R.; ... NIZET, V. Group B *Streptococcus* engages an inhibitory Siglec through sialic acid mimicry to blunt innate immune and inflammatory responses in vivo. **PLoS Pathogens**, v. 10, n. 1, p. e1003846, 2014.
- CHATTOPADHYAY, D.; CAREY, A. J.; CALIOT, E.; WEBB, R. I.; LAYTON, J. R.; WANG, Y.; ... ULETT, G. C. Phylogenetic lineage and pilus protein Spb1/SAN1518 affect opsonin-independent phagocytosis and intracellular survival of Group B *Streptococcus*. **Microbes and infection**, v. 13, n. 4, p. 369-382, 2011.
- CHEN, S. L. Genomic insights into the distribution and evolution of group B *Streptococcus*. **Frontiers in microbiology**, v. 10, p. 1447, 2019.
- CIESLEWICZ, M. J.; CHAFFIN, D.; GLUSMAN, G.; KASPER, D.; MADAN, A.; RODRIGUES, S.; ... RUBENS, C. E. Structural and genetic diversity of group B *Streptococcus* capsular polysaccharides. **Infection and immunity**, v. 73, n. 5, p. 3096-3103, 2005.
- DENG, L.; KASPER, D. L.; KRICK, T. P.; WESSELS, M. R. Characterization of the linkage between the Type III capsular polysaccharide and the bacterial cell wall of group B *Streptococcus*. **Journal of Biological Chemistry**, v. 275, n. 11, p. 7497-7504, 2000.
- DOGAN, B.; SCHUKKEN, Y. H.; SANTISTEBAN, C.; BOOR, K. J. Distribution of serotypes and antimicrobial resistance genes among *Streptococcus agalactiae*

isolates from bovine and human hosts. **Journal of Clinical Microbiology**, v. 43, n. 12, p. 5899-5906, 2005.

DONG, Y.; MA, K.; CAO, Q.; HUANG, H.; NIE, M.; LIU, G.; ... LIU, Y. CRISPR-dependent endogenous gene regulation is required for virulence in piscine *Streptococcus agalactiae*. **Emerging microbes & infections**, v. 10, n. 1, p. 2113-2124, 2021.

DORON, S.; MELAMED, S.; OFIR, G.; LEAVITT, A.; LOPATINA, A.; KEREN, M.; ... SOREK, R. Systematic discovery of antiphage defense systems in the microbial pangenome. **Science**, v. 359, n. 6379, p. eaar4120, 2018.

EICKHOFF, T C.; KLEIN, J. O.; DALY, A. K.; INGALL, D.; FINLAND, M. Neonatal sepsis and other infections due to group B beta-hemolytic streptococci. **New England Journal of Medicine**, v. 271, n. 24, p. 1221-1228, 1964.

EL-SAYED, A. M. **Tilapia culture**. 2nd ed. San Diego: Academic Press (Elsevier Science Publishing Co Inc.), 2019.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. **Piscicultura em tanques-rede / Embrapa Amazônia Oriental**. 120 p.: il. – (Coleção Criar, 6). Brasília, DF: Embrapa Informação Tecnológica, 2009.

EVANS, J. J.; BOHNSACK, J. F.; KLESIUS, P. H.; WHITING, A. A.; GARCIA, J. C.; SHOEMAKER, C. A.; TAKAHASHI, S. Phylogenetic relationships among *Streptococcus agalactiae* isolated from piscine, dolphin, bovine and human sources: a dolphin and piscine lineage associated with a fish epidemic in Kuwait are also associated with human neonatal infections in Japan. **Journal of medical microbiology**, v. 57, n. 11, p. 1369-1376, 2008.

FAO (Food and Agriculture Organization the United Nations). **The State of World Fisheries and Aquaculture 2022**. Towards Blue Transformation. Rome, 2022. Disponível em: <https://www.fao.org/3/cc0461en/online/cc0461en.html>

FRY, R. M. Fatal Infections by Haemolytic Streptococcus Group B. **Lancet**, p. 199-201, 1938.

GABRIELSEN, C.; MÆLAND, J. A.; LYG, R. V.; RADTKE, A.; AFSET, J. E. Molecular characteristics of *Streptococcus agalactiae* strains deficient in alpha-like protein encoding genes. **Journal of Medical Microbiology**, v. 66, n. 1, p. 26-33, 2017.

GIANI, A. M.; GALLO, G. R.; GIANFRANCESCHI, L.; FORMENTI, G. Long walk to genomics: History and current approaches to genome sequencing and assembly. **Computational and Structural Biotechnology Journal**, v. 18, p. 9-19, 2020.

GLASER, P.; RUSNIOK, C.; BUCHRIESER C.; CHEVALIER, F.; FRANGEUL, L.; MSADEK, T.; ZOUINE, M.; COUVÉ, E.; LALIOUI, L.; POYART, C.; TRIEU-CUOT, P.; KUNST F. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. **Molecular microbiology**, v. 45, n. 6, p. 1499-1513, 2002.

GODOY, D. T.; CARVALHO-CASTRO, G. A.; LEAL, C. A. G.; PEREIRA, U. P.; LEITE, R. C.; FIGUEIREDO, H. C. P. Genetic diversity and new genotyping scheme for fish

pathogenic *Streptococcus agalactiae*. **Letters in applied microbiology**, v. 57, n. 6, p. 476-483, 2013.

GORI, A.; HARRISON, O. B.; MLIA, E.; NISHIHARA, Y.; CHAN, J. M.; MSEFULA, J.; ... HEYDERMAN, R. S. Pan-GWAS of *Streptococcus agalactiae* highlights lineage-specific genes associated with virulence and niche adaptation. **MBio**, v. 11, n. 3, p. e00728-20, 2020.

GRAD, Y. H.; LIPSITCH, M. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. **Genome biology**, v. 15, n. 11, p. 1-14, 2014.

HAMMOND, J. A.; GORDON, E. A.; SOCARRAS, K. M.; CHANG MELL, J.; EHRLICH, G. D. Beyond the pan-genome: current perspectives on the functional and practical outcomes of the distributed genome hypothesis. **Biochemical Society Transactions**, v. 48, n. 6, p. 2437-2455, 2020.

HERBERT, M. A.; BEVERIDGE, C. J.; MCCORMICK, D.; ATEN, E.; JONES, N.; SNYDER, L. A.; SAUNDERS, N. J. Genetic islands of *Streptococcus agalactiae* strains NEM316 and 2603VR and their presence in other Group B streptococcal strains. **BMC microbiology**, v. 5, n. 1, p. 1-13, 2005.

HILLER, N. L.; SÁ-LEÃO, R. Puzzling over the pneumococcal pangenome. **Frontiers in microbiology**, p. 2580, 2018.

JIANG, S.; PARK, S. E.; YADAV, P.; PAOLETTI, L. C.; WESSELS, M. R. Regulation and function of pilus island 1 in group B *Streptococcus*. **Journal of bacteriology**, v. 194, n. 10, p. 2479-2490, 2012.

JOHRI, A. K.; PAOLETTI, L. C.; GLASER, P.; DUA, M.; SHARMA, P. K.; GRANDI, G.; RAPPUOLI, R. Group B *Streptococcus*: global incidence and vaccine development. **Nature Reviews Microbiology**, v. 4, n. 12, p. 932-942, 2006.

JONES, N.; BOHNSACK, J. F., TAKAHASHI, S., OLIVER, K. A., CHAN, M. S., KUNST, F., ... & SPRATT, B. G. Multilocus sequence typing system for group B *Streptococcus*. **Journal of clinical microbiology**, v. 41, n. 6, p. 2530-2536, 2003.

KANNIKA, K.; PISUTTHARACHAI, D.; SRISAPOOME, P.; WONGTAVATCHAI, J.; KONDO, H.; HIRONO, I.; UNAJAK, S.; AREECHON N. Molecular serotyping, virulence gene profiling and pathogenicity of *Streptococcus agalactiae* isolated from tilapia farms in Thailand by multiplex PCR. **Journal of applied microbiology**, v. 122, n. 6, p. 1497-1507, 2017.

KAVVAS, E. S. CATOIU, E.; MIH, N.; YURKOVICH, J. T.; SEIF, Y.; DILLON, N.; ... PALSSON, B. O. Machine learning and structural analysis of *Mycobacterium tuberculosis* pan-genome identifies genetic signatures of antibiotic resistance. **Nature communications**, v. 9, n. 1, p. 1-9, 2018.

KAYANSAMRUJAJ, P.; SOONTARA, C.; UNAJAK, S.; DONG, H. T.; RODKHUM, C.; KONDO, H.; ... AREECHON, N. Comparative genomics inferred two distinct populations of piscine pathogenic *Streptococcus agalactiae*, serotype Ia ST7 and

serotype III ST283, in Thailand and Vietnam. **Genomics**, v. 111, n. 6, p. 1657-1667, 2019.

KIM, Y.; GU, C.; KIM, H. U.; LEE, S. Y. Current status of pan-genome analysis for pathogenic bacteria. **Current opinion in biotechnology**, v. 63, p. 54-62, 2020.

KLESIOUS, P. H.; SHOEMAKER, C. A.; EVANS, J. J. *Streptococcus*: a worldwide fish health problem. In: **Proceedings of the 8th International Symposium on Tilapia in Aquaculture**. Ag. Press Unit Abbassa, Egypt, p. 83-107, 2008.

KOIDE, S.; NAGANO, Y.; TAKIZAWA, S.; SAKAGUCHI, K.; SOGA, E.; HAYASHI, W.; ... NAGANO, N. Genomic traits associated with virulence and antimicrobial resistance of invasive group B *Streptococcus* isolates with reduced penicillin susceptibility from elderly adults. **Microbiology Spectrum**, v. 10, n. 3, p. e00568-22, 2022.

LANCEFIELD, R. C. A serological differentiation of human and other groups of hemolytic streptococci. **The Journal of experimental medicine**, v. 57, n. 4, p. 571-595, 1933.

LANCEFIELD, R. C.; HARE, R. The serological differentiation of pathogenic and non-pathogenic strains of hemolytic streptococci from parturient women. **The Journal of experimental medicine**, v. 61, n. 3, p. 335, 1935.

LAPIERRE, P.I; GOGARTEN, J. P. Estimating the size of the bacterial pan-genome. **Trends in genetics**, v. 25, n. 3, p. 107-110, 2009.

LAZARUS, J. M.; SELLERS, D. P.; MARINE, W. M. Meningitis due to the group B beta-hemolytic *Streptococcus*. **New England Journal of Medicine**, v. 272, n. 3, p. 146-147, 1965.

LAZZARIN, M.; MU, R.; FABBRINI, M.; GHEZZO, C.; RINAUDO, C. D.; DORAN, K. S.; MARGARIT, I. Contribution of pilus type 2b to invasive disease caused by a *Streptococcus agalactiae* ST-17 strain. **BMC microbiology**, v. 17, n. 1, p. 1-8, 2017.

LEFÉBURE, T.; STANHOPE, M. J. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. **Genome biology**, v. 8, n. 5, p. 1-17, 2007.

LI, L; STOECKERT, C. J.; ROOS, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. **Genome research**, v. 13, n. 9, p. 2178-2189, 2003.

LI, L.; WANG, R.; LIANG, W.; GAN, X.; HUANG, T.; HUANG, Y.; ... LUO, H. Rare serotype occurrence and PFGE genotypic diversity of *Streptococcus agalactiae* isolated from tilapia in China. **Veterinary microbiology**, v. 167, n. 3-4, p. 719-724, 2013.

LIU, G.; ZHU, J.; CHEN, K.; GAO, T.; YAO, H.; LIU, Y.; ... LU, C. Development of *Streptococcus agalactiae* vaccines for tilapia. **Diseases of aquatic organisms**, v. 122, n. 2, p. 163-170, 2016.

LIU, Y.; LI, L.; HUANG, T.; WANG, R.; LIANG, W.; YANG, Q.; ... CHEN, M. Comparative multi-omics systems analysis reveal the glycolysis/gluconeogenesis

signal pathway play an important role in virulence attenuation in fish-derived GBS YM001. **Plos one**, v. 14, n. 8, p. e0221634, 2019.

MAIDEN, M. C. J. Multilocus sequence typing of bacteria. **Annual Review of Microbiology**, v. 60, p. 561-588, 2006.

MAISANO, A. M.; ROMANÒ, A.; MARASCHI, F.; MORANDI, S.; ORLANDELLI, C.; SPELTA, C.; VEZZOLI, F.; BERTOCCHI, L.; LUINI, M. Epidemiological, molecular and antimicrobial-resistance investigations in dairy farms infected with *Streptococcus agalactiae*. **Large Animal Review**, v. 24, n. 1, p. 11-18, 2018.

MA, K.; CAO, Q.; LUO, S.; WANG, Z.; LIU, G.; LU, C.; LIU, Y. Cas9 enhances bacterial virulence by repressing the regR transcriptional regulator in *Streptococcus agalactiae*. **Infection and immunity**, v. 86, n. 3, p. e00552-17, 2018.

MARCUSSO, P. F.; SALVADOR, R.; MARINHO-NETO, F. A. Infecção por *Streptococcus agalactiae* em tilápias do Nilo (*Oreochromis niloticus*). **Revista de Ciências Agroveterinárias**, v. 16, n. 2, p. 165-169, 2017.

MAULU, S.; HASIMUNA, O. J.; MPHANDE, J.; MUNANG'ANDU, H. M. Prevention and Control of Streptococcosis in Tilapia Culture: A Systematic Review. **Journal of Aquatic Animal Health**, 2021.

MEDINI, D.; DONATI, C.; TETTELIN, H.; MASIGNANI, V.; RAPPUOLI, R. The microbial pan-genome. **Current opinion in genetics & development**, v. 15, n. 6, p. 589-594, 2005.

MIAN, G. F.; GODOY, D. T.; LEAL, C. A. G.; YUHARA, T. Y.; COSTA, G. M.; FIGUEIREDO, H. C. P. Aspects of the natural history and virulence of *S. agalactiae* infection in Nile tilapia. **Veterinary microbiology**, v. 136, n. 1-2, p. 180-183, 2009.

MIRANDA, P. S. D.; LANNES-COSTA, P. S.; PIMENTEL, B. A. S.; SILVA, L. G.; FERREIRA-CARVALHO, B. T.; MENEZES, G. C.; ... NAGAO, P. E. Biofilm formation on different pH conditions by *Streptococcus agalactiae* isolated from bovine mastitic milk. **Letters in Applied Microbiology**, v. 67, n. 3, p. 235-243, 2018.

MORACH, M.; STEPHAN, R.; SCHMITT, S.; EWERS, C.; ZSCHÖCK, M.; REYESVELEZ, J.; ... JOHLER, S. Population structure and virulence gene profiles of *Streptococcus agalactiae* collected from different hosts worldwide. **European Journal of Clinical Microbiology & Infectious Diseases**, v. 37, n. 3, p. 527-536, 2018.

OLIVEIRA, R. C. O panorama da aquicultura no Brasil: a prática com foco na sustentabilidade. **Revista INTERTOX de toxicologia, risco ambiental e sociedade**, v. 2, n. 1, p. 71-89, 2015.

OVERBEEK, R.; BEGLEY, T.; BUTLER, R. M.; CHOUDHURI, J. V.; CHUANG, H. Y.; COHOON, M.; ... VONSTEIN, V. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. **Nucleic acids research**, v. 33, n. 17, p. 5691-5702, 2005.

PEIXE BR. **Anuário Peixe BR da Piscicultura 2022**. Associação Brasileira de Piscicultura. Brasil, 2022. Disponível em: <https://www.peixebr.com.br/anuario2022/>

PHARES, C. R.; LYNFIELD, R.; FARLEY, M. M.; MOHLE-BOETANI, J.; HARRISON, L. H.; PETIT, S.; ... SCHRAG, S. J. Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005. **Jama**, v. 299, n. 17, p. 2056-2065, 2008.

PINTO, T.; COSTA, N. S.; SOUZA, A. R. V.; SILVA, L. G. D.; CORRÊA, A. B. D. A.; FERNANDES, F. G.; ... BENCHETRIT, L. C. Distribution of serotypes and evaluation of antimicrobial susceptibility among human and bovine *Streptococcus agalactiae* strains isolated in Brazil between 1980 and 2006. **Brazilian Journal of Infectious Diseases**, v. 17, p. 131-136, 2013.

PRADEEP, P. J.; SUEBSING, R.; SIRTHAMMAJAK, S.; KAMPEERA, J.; JITRAKORN, S.; SAKSMERPROME, V.; ... WITHYACHUMANARNKUL, B. Evidence of vertical transmission and tissue tropism of Streptococcosis from naturally infected red tilapia (*Oreochromis* spp.). **Aquaculture Reports**, v. 3, p. 58-66, 2016.

PRETTO-GIORDANO, L. G.; MULLER, E. E.; KLESIOUS, P.; DA SILVA, V. G. Efficacy of an experimentally inactivated *Streptococcus agalactiae* vaccine in Nile tilapia (*Oreochromis niloticus*) reared in Brazil. **Aquaculture Research**, v. 41, n. 10, p. 1539-1544, 2010.

PRICE, M. N.; DEHAL, P. S.; ARKIN, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. **PloS one**, v. 5, n. 3, p. e9490, 2010.

PUCKETT, S.; TRUJILLO, C.; EOH, H.; MARRERO, J.; SPENCER, J.; JACKSON, M.; ... EHRT, S. Inactivation of fructose-1, 6-bisphosphate aldolase prevents optimal catabolism of glycolytic and gluconeogenic carbon substrates in *Mycobacterium tuberculosis*. **PLoS pathogens**, v. 10, n. 5, p. e1004144, 2014.

RAABE, V. N.; SHANE, A. L. Group B streptococcus (*Streptococcus agalactiae*). **Microbiology spectrum**, v. 7, n. 2, p. 7.2. 17, 2019.

RICHARDS, V. P.; LANG, P.; BITAR, P. D. P.; LEFÉBURE, T.; SCHUKKEN, Y. H.; ZADOKS, R. N.; STANHOPE, M. J. Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*. **Infection, Genetics and Evolution**, v. 11, n. 6, p. 1263-1275, 2011.

RINAUDO, C. D.; ROSINI, R.; GALEOTTI, C. L.; BERTI, F.; NECCHI, F.; REGUZZI, V.; MAIONE, D. Specific involvement of pilus type 2a in biofilm formation in group B *Streptococcus*. **PloS one**, v. 5, n. 2, p. e9216, 2010.

ROBINSON, J. A.; MEYER, F. P. Streptococcal fish pathogen. **Journal of Bacteriology**, v. 92, n. 2, p. 512-512, 1966.

ROSSATO, S.; NOVACK, M. M. E.; CEREZER, E.; FRONZA, R. T. L.; BISOGNIN, J.; SOUTO, M. S.; DE BRUM, P. F.; DA FONTOURA, M. P. Comparação entre sistemas de cultivo: tanques-rede x tanques escavados. **Brazilian Journal of Development**, v. 7, n. 12, p. 110465-110481, 2021.

ROULI, L.; MERHEJ, V.; FOURNIER, P. E.; RAOULT, D. The bacterial pangenome as a new tool for analyzing pathogenic bacteria. **New microbes and new infections**, v. 7, p. 72-85, 2015.

SENAR (Serviço Nacional de Aprendizagem Rural). **Piscicultura: criação de tilápias em tanques-rede**. 108 p.; il. – (Coleção SENAR, 208). Brasília: SENAR, 2018a.

SHABAYEK, S; SPELLERBERG, B. Group B streptococcal colonization, molecular characteristics, and epidemiology. **Frontiers in microbiology**, v. 9, p. 437, 2018.

SILVA, B. C.; MARTINS, M. L.; JATOBÁ, A.; BUGLIONE NETO, C. C.; VIEIRA, F. N.; PEREIRA, G. V.; ... MOURIÑO, J. L. P. Hematological and immunological responses of Nile tilapia after polyvalent vaccine administration by different routes. **Pesquisa Veterinária Brasileira**, v. 29, p. 874-880, 2009.

SINHA, D.; SUN, X.; KHARE, M.; DRANCOURT, M.; RAOULT, D.; FOURNIER, P. E. Pangenome analysis and virulence profiling of *Streptococcus intermedius*. **BMC Genomics**, v. 22, n. 1, p. 1-17, 2021.

SKOFF, T. H.; FARLEY, M. M.; PETIT, S.; CRAIG, A. S.; SCHAFFNER, W.; GERSHMAN, K.; ... SCHRAG, S. J. Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990–2007. **Clinical Infectious Diseases**, v. 49, n. 1, p. 85-92, 2009.

SLOTVED, H. C.; KONG, F.; LAMBERTSEN, L.; SAUER, S.; GILBERT, G. L. Serotype IX, a proposed new *Streptococcus agalactiae* serotype. **Journal of clinical microbiology**, v. 45, n. 9, p. 2929-2936, 2007.

SUN, J.; FANG, W.; KE, B.; HE, D.; LIANG, Y.; NING, D.; ... DENG, X. Inapparent *Streptococcus agalactiae* infection in adult/commercial tilapia. **Scientific reports**, v. 6, n. 1, p. 1-11, 2016.

SYUHADA, R.; ZAMRI-SAAD, M.; INA-SALWANY, M. Y.; MUSTAFA, M.; NASRUDDIN, N. N.; DESA, M. N. M.; ... AMAL, M. N. A. Molecular characterization and pathogenicity of *Streptococcus agalactiae* serotypes Ia ST7 and III ST283 isolated from cultured red hybrid tilapia in Malaysia. **Aquaculture**, v. 515, p. 734543, 2020.

TETTELIN, H.; MASIGNANI, V.; CIESLEWICZ, M. J.; EISEN, J. A.; PETERSON, S.; WESSELS, M. R.; ... FRASER, C. M. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. **Proceedings of the National Academy of Sciences**, v. 99, n. 19, p. 12391-12396, 2002.

TETTELIN, H.; MASIGNANI, V.; CIESLEWICZ, M. J.; DONATI, C.; MEDINI, D.; WARD, N. L.; ... FRASER, C. M. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. **Proceedings of the National Academy of Sciences**, v. 102, n. 39, p. 13950-13955, 2005.

TORTORA, G. J.; CASE, C. L.; FUNKE, B. R. **Microbiologia**. 12. ed. – Porto Alegre: Artmed, 2017.

VERNIKOS, G.; MEDINI, D.; RILEY, D. R.; TETTELIN, H. Ten years of pan-genome analyses. **Current Opinion in Microbiology**, v. 23, p. 148–154, 2015.

VIDAL AMARAL, J. R.; Ramos, R. T. J.; Araújo, F. A.; Kato, R. B.; Aburjaile, F. F.; Soares, S. C.; ... Rosado, A. S. Bacteriocin Producing *Streptococcus agalactiae*

Strains Isolated from Bovine Mastitis in Brazil. **Microorganisms**, v. 10, n. 3, p. 588, 2022.

VOGEL, V.; SPELLERBERG, B. Bacteriocin production by beta-hemolytic streptococci. **Pathogens**, v. 10, n. 7, p. 867, 2021.

WESCOMBE, P. A.; TAGG, J. R. Purification and characterization of streptin, a type A1 lantibiotic produced by *Streptococcus pyogenes*. **Applied and environmental microbiology**, v. 69, n. 5, p. 2737-2747, 2003.

XIA, F. D.; MALLET, A.; CALIOT, E.; GAO, C.; TRIEU-CUOT, P.; DRAMSI, S. Capsular polysaccharide of Group B *Streptococcus* mediates biofilm formation in the presence of human plasma. **Microbes and infection**, v. 17, n. 1, p. 71-76, 2015.

XU, D. H.; SHOEMAKER, C. A.; KLESIOUS, P. H. Evaluation of the link between gyrodactylosis and streptococcosis of Nile tilapia, *Oreochromis niloticus* (L.). **Journal of Fish Diseases**, v. 30, n. 4, p. 233-238, 2007.

XU, L.; DONG, Z.; FANG, L.; LUO, Y.; WEI, Z.; GUO, H.; ... WANG, Y. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. **Nucleic acids research**, v. 47, n. W1, p. W52-W58, 2019.

YAGUPSKY, P.; MENEGUS, M. A.; POWELL, K. R. The changing spectrum of group B streptococcal disease in infants: an eleven-year experience in a tertiary care hospital. **The Pediatric infectious disease journal**, v. 10, n. 11, p. 801-808, 1991.

YE, X.; LI, J.; LU, M.; DENG, G.; JIANG, X.; TIAN, Y.; ... JIAN, Q. Identification and molecular typing of *Streptococcus agalactiae* isolated from pond-cultured tilapia in China. **Fisheries Science**, v. 77, n. 4, p. 623-632, 2011.

ZAMRI-SAAD, M.; AMAL, M. N. A.; SITI-ZAHRAH, A.; ZULKAFI, A. R. Control and Prevention of Streptococcosis in Cultured Tilapia in Malaysia: A Review. **Pertanika Journal of Tropical Agricultural Science**, v. 37, n. 4, 2014.

ZHANG, D.; KE, X.; LIU, Z.; CAO, J.; SU, Y.; LU, M.; ... QIN, F. Capsular polysaccharide of *Streptococcus agalactiae* is an essential virulence factor for infection in Nile tilapia (*Oreochromis niloticus* Linn.). **Journal of fish diseases**, v. 42, n. 2, p. 293-302, 2019.

ZHANG, Z.; LAN, J.; LI, Y.; HU, M.; YU, A.; ZHANG, J.; WEI, S. THE pathogenic and antimicrobial characteristics of an emerging *Streptococcus agalactiae* serotype IX in Tilapia. **Microbial pathogenesis**, v. 122, p. 39-45, 2018.

ZHANG, Z. Research Advances on Tilapia Streptococcosis. **Pathogens**, v. 10, n. 5, p. 558, 2021.

ZIMMERMANN, Robert A. The double life of ribosomal proteins. **Cell**, v. 115, n. 2, p. 130-132, 2003.