FEDERAL UNIVERSITY OF MINAS GERAIS

VETERINARY SCHOOL

GRADUATE PROGRAM IN VETERINARY SCIENCE

GENOMIC DIVERSITY OF THE BRAZILIAN STRAINS OF *Streptococcus agalactiae* ISOLATED FROM OUTBREAKS ON FISH FARMS

Gustavo Morais Barony

Belo Horizonte, MG

Veterinary School, UFMG

2017

GUSTAVO MORAIS BARONY

GENOMIC DIVERSITY OF THE BRAZILIAN STRAINS OF *Streptococcus agalactiae* ISOLATED FROM OUTBREAKS ON FISH FARMS

Dissertation submitted to the Department of Preventive Veterinary Medicine of the Graduate Program in Veterinary Science of the Veterinary School of the Federal University of Minas Gerais in partial fulfillment of the requirements for the degree of Master in Animal Science, under orientation of professor Henrique César Pereira Figueiredo.

Belo Horizonte, MG

Veterinary School, UFMG

2017

Barony, Gustavo Morais, 1991-
B266g Genomic diversity of the brazilian strains of *Streptococcus agalactiae* isolated from outbreaks on fish farms / Gustavo Morais Barony. – 2017.
65 p. : il.

Orientador: Henrique César Pereira Figueiredo
Dissertação (mestrado) – Universidade Federal de Minas Gerais. Departamento de Medicina Veterinária Preventiva
Inclui bibliografia

1. Tilápia (Peixe) – Doenças – Teses. 2. Estreptococo – Teses. 3. Genômica – Teses. I. Figueiredo, Henrique César Pereira. II. Universidade Federal de Minas Gerais. Departamento de Medicina Veterinária Preventiva. III. Título.

CDD – 639.31

UFMG UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

GUSTAVO MORAIS BARONY

Dissertação submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIA ANIMAL, como requisito para obtenção do grau de MESTRE em CIÊNCIA ANIMAL, área de concentração MEDICINA VETERINÁRIA PREVENTIVA.

Aprovada em 07 de Fevereiro de 2017, pela banca constituída pelos membros:

Prof. Henrique César Pereira Figueiredo
Presidente - Orientador

Prof. Carlos Augusto Gomes Leal
Escola de Veterinária - UFMG

Prof. Vasco Ariston de Carvalho Azevedo
Instituto de Ciências Biológicas - UFMG

Prof. Rodrigo Otávio Silveira Silva
Escola de Veterinária - UFMG

Universidade Federal de Minas Gerais
Escola de Veterinária
Av. Antônio Carlos, 6627 -
Caixa Postal 567 – CEP 30123-970
Belo Horizonte – Minas Gerais
Telefone: (31) 3409-2057/2059(fax)
www.vet.ufmg.br
E-mail cap@vet.ufmg.br

Escola de Veterinária
UFMG

ACKNOWLEDGEMENTS

I would like to thank:

God, for His provision;

My parents José and Virginia, for their lessons;

My sister Fernanda, for her patience and companionship;

Daniel, for his support;

My friends, for all the moments together;

My co-workers, for the uncountable contributions, especially Felipe and Guilherme;

Professor Siomar Soares, for the assistance;

All my professors of the Federal University of Minas Gerais, for teaching me how to learn;

Professor Henrique Figueiredo, for helping me to grow;

All the employees of the Veterinary School, for their helpfulness;

The federal employees of the Postgraduate Secretary, especially Luzete, for her patience and joy;

All the federal employees and staff who take care of the University, for contributing to everybody's welfare;

The Federal University of Minas Gerais, for my nine years of permanence;

The Coordination for the Improvement of Higher Education Personnel (CAPES), for granting me a Master degree scholarship;

SUMMARY

ABBREVIATION LIST

BIGSdb: Bacterial Isolate Genome Sequence Database
BHI: Brain Heart Infusion
BP: Base Pairs
BURST: Based Upon Related Sequences Types
CC: Clonal Complex
CDS: Coding Sequence
CNS: Central Nervous System
ESS: Effective Sample Size
FAO: Food and Agriculture Organization
FNO: *Francisella noatunensis* subsp. *orientalis*
GBS: Group B Streptococcus
GCS: Group C Streptococcus
GTR: General Time-Reversible
HPD: Highest Probability Density
IBGE: Brazilian Institute of Geography and Statistics
MALDI-TOF: Matrix Assisted Laser Desorption/Ionization – Time of Flight
MB: Megabases (1,000,000 of bases)
MCMC - Markov Chain Monte Carlo
MLST: Multilocus Sequence Typing
MPA: Brazilian Ministry of Fish and Aquaculture
NGS: Next Generation Sequencing
NT: Nucleotide
PCR: Polymerase Chain Reaction
qPCR: Quantitative Polymerase Chain Reaction
ST: Sequence-Type
tMRCA: Time to Most Recent Common Ancestor
TS: Tryptic Soy Medium
TSA: Tryptic Soy Agar
TSB: Tryptic Soy Broth
wgMLST: Whole Genome Multilocus Sequence Typing
WGS: Whole Genome Sequencing

FIGURE LIST

TABLE LIST

RESUMO

A tilapia do Nilo é o peixe mais produzido no Brasil e a expansão de seu cultivo é afetada pelo patógeno *Streptococcus agalactiae*. Os objetivos dessa dissertação foram realizar uma revisão de literatura sobre estreptococose por *S. agalactiae* e sua epidemiologia, e analisar a população brasileira desse micro-organismo, estabelecer uma ferramenta para rastrear surtos de estreptococose e distinguir amostras geneticamente próximas. Um total de 39 amostras foram obtidas de surtos e seus genomas foram sequenciados e anotados para análises comparativas de tipagem multilocus (MLST), similaridade genômica, MLST de genoma inteiro (wgMLST) e uma análise evolutiva com inferência Bayesiana da espécie. Os isolados brasileiros de *S. agalactiae* apresentaram dois STs, dentre os quais um novo descrito pela primeira vez nesse trabalho, e também uma linhagem não tipável. O wgMLST diferenciou cada isolado como um clone único e estabeleceu correlações temporais e geográficas entre as amostras, sendo que os STs 260 e 927 prevaleceram no Nordeste e as amostras não tipáveis, na região Centro-Sul. A análise evolutiva Bayesiana mostrou que a população brasileira de *S. agalactiae* tem uma emergência muito recente, de cerca de 585 anos. A população brasileira de *S. agalactiae* se mostrou genomicamente heterogênea e distribuida em diferentes regiões do país conforme seu genótipo, e o wgMLST pode rastrear cada evento de surto individualmente.

Palavras-chave: MLST; estreptococose; genômica; genômica comparativa; tMRCA; tilápia; filogenética; wgMLST.

ABSTRACT

Nile tilapia is the most produced fish in Brazil and the evolution of its cultivation is affected by the pathogen *Streptococcus agalactiae*. The aims of this thesis were to review the literature about streptococcosis by *S. agalactiae* and its epidemiology and to analyze genetic structure of the Brazilian isolates, to establish a method to track outbreaks of streptococcosis and to distinguish closely related strains. A total of 39 strains were obtained from outbreaks and their whole genomes were sequenced and annotated for comparative analysis of multilocus sequence typing (MLST), genomic similarity, whole genome MLST (wgMLST) and a Bayesian evolutionary analysis of the species. The Brazilian strains of *S. agalactiae* presented two STs, among which a new one just described, and also a non-typeable lineage. The wgMLST could differentiate each strain in a single clone and establish temporal and geographical correlations among strains, wherein STs 260 and 927 predominated in Northeast, and non-typeable strains, on Central-South region. The Bayesian evolutionary analysis revealed that the Brazilian population of *S. agalactiae* has a remarkably recent emergence, about 585 years ago. Brazilian strains of S. agalactiae showed to be heterogeneous in its genome sequence, distributed in different regions of the country according to the genotype, and wgMLST analysis could track each outbreak event individually.

Keywords: MLST; streptococcosis; genomics; comparative genomics; tMRCA; tilapia; phylogenetics; wgMLST.

1. INTRODUCTION

Streptococcus agalactiae, or Group B Streptococcus (GBS), is a major pathogen to several fish species worldwide, especially Nile tilapia (*Oreochromis niloticus*). Tilapia is the second most produced fish in the World and the most produced fish in Brazil, and its chain is severely hampered by infectious diseases, as streptococcosis by *S. agalactiae*. Besides its impact on aquaculture, this pathogen may also be the aetiological agent of diseases in newborn infants and immunocompromised humans, and may cause sepsis, meningitis and pneumonia, or even other illnesses in many animal species. In fish, streptococcosis by GBS evolves mainly to an acute disease characterized by septicaemia and meningoencephalitis, associated with clinical signals of exophthalmia, erratic swimming, ascites, hyperaemia on the fin basis, and high mortality rates. In some cases, it can remain as a chronic infection which may lead to decrease of growth and performance loss, which can become economically infeasible.

The epidemiology of this disease is specially studied by multilocus sequence typing (MLST), which consists in typing seven housekeeping genes and the combination of seven alleles originates a sequence-type (ST). The STs may be evolutionarily related if they share enough alleles, and it can be grouped in a clonal complex (CC) if a ST shares six alleles with the others, being this ST the group's founder. Through MLST it has been discovered STs and CCs specially involved in human meningitis, or cow mastitis, or even fish meningoencephalitis, but in many cases, strains with several host, geographical and temporal origins share a ST or a CC, what made it necessary more powerful methods for epidemiological studies.

With the advent and growth of genomics, many *S. agalactiae* whole genome sequences (WGSs) are being published, and it is increasing the number of genomic-based epidemiological studies of diseases by this micro-organism. Through comparative genomics it is possible to ensure bigger and lower proximity between strains in a population, and it can lead to more assertive evolutionary studies, or clinical ones, as patterns of antibiotic resistance, spatial circulation of strains, pool of virulence factors and vaccine targets. A big step on genetic typing was the whole genome MLST (wgMLST), which instead of typing seven conserved alleles, it types every allele in a genome and, so, allows a bigger discrimination power for separating closely-related strains.

2. OBJECTIVE

The aim of this dissertation was to study the genetic structure of the Brazilian piscine isolates of *S. agalactiae* through comparative genomics in addition to track each outbreak and distinguish closely related strains.

3. CHAPTER ONE – Literature review

3.1. BRAZILIAN AQUACULTURE

Aquaculture is the food production chain which most quickly evolves worldwide, with an average annual rate of expansion of 8% in the last three decades (THE…, 2014) . In the year of 2012, the world's production reached 90.4 million Tons, moving a financial amount of 144,4 billion (THE…, 2014). Tilapia species, specially Nile tilapia (*Oreochromis niloticus*) is the most widespread fish on the global aquaculture, occupying the second place on the ranking of the most produced fish worldwide with an average production of 3.7 million Tons in 2014 and losing uniquely to carps (THE…, 2014; CULTURED..., 2017). Nowadays, Tilapia production is in expansion in Asia, South America and Africa, with a growing volume of supply entering domestic markets in the major producing countries (THE…, 2016). Brazil took the 14th place on the ranking of the 25 top aquaculture producers in 2014, with a total production of 562.5 thousand Tons (THE…, 2016). Tilapia production led the Brazilian aquaculture in 2015, with a total of production of 219.33 thousand Tons, with a growth of 9.7% from the previous year (PRODUÇÃO…, 2015), and other major produced fish were carps and native fish species as tambaqui (*Colossoma macropomum*) and catfish species and its hybrids (BOLETIM…, 2013; PRODUÇÃO…, 2015). Furthermore, Brazil has been growing in shrimp farming, with a production of 69.9 thousand Tons in 2015 (PRODUÇÃO…, 2015), which represented an annual growth of 7.4% compared to 2014. Brazil has a huge potential to aquaculture development, an extent shore of about 8,400 km plus 5,500,000 ha of freshwater in reservoirs. In addition, Brazil has a favorable weather and abundant manpower, added to the increase of national consumption of fish (Crepaldi, 2006).

Notwithstanding, the occurrence of outbreaks of infectious diseases has been shown as the main hindrance to the expansion of the tilapia farming in Brazil (Figueiredo, 2012). Bacteria are the main aetiological agents of diseases on farmed tilapia in Brazil, among which outbreaks by *Streptococcus agalactiae* (GBS, Group B Streptococcus), *Streptococcus dysgalactiae* (GCS, Group C Streptococcus), *Aeromonas hydrophila*, *Flavobacterium columnare* and *Francisella noatunensis subp. orientalis* (FNO) highlight as the most frequent ones (Figueiredo *et al.*, 2005; Mian *et al.*, 2009; Carvalho-Castro *et al.*, 2010; Netto *et al.*, 2011; Sebastião *et al.*, 2011; Figueiredo, 2012; Leal *et al.*, 2014; Assis *et al.*, 2016). Although GBS is currently the most associated bacterial pathogen with tilapia in Brazil, FNO francisellosis and even coinfection with GBS is an emerging health problem on this field (Assis *et al.*, 2016).

3.2. *Streptococcus agalactiae* BIOLOGY AND STREPTOCOCCOSIS

S. agalactiae is a Gram-positive coccus which has already been described as an etiological agent of disease in many animals, as human, bovine and fish, and, less often, in dolphins, camels, horses, cats, dogs, frogs, among others (Edelstein and Pegram, 1974; Dow *et al.*, 1987; Elliott *et al.*, 1990; Vandamme *et al.*, 1997; Evans *et al.*, 2008). Besides having a pathogenicity potential and a broad host range, *S. agalactiae* is a member of the human

intestinal and vaginal microbiota, with carriage percentage varying between 10 and 36% in both women and men (Sorensen *et al.*, 2010). This micro-organism is a facultative anaerobe, catalase and oxidase negative, grows in temperatures between 25-37°C in common media as Brain Heart Infusion (BHI) and Tryptic Soy Medium (TS), can present complete, partial or absent hemolysis (β, α or γ, respectively) in blood agar (Evans, 2002; Evans *et al.*, 2008; Soto *et al.*, 2015). Furthermore, *S. agalactiae* is a commonly capsulated bacterium, and there are currently 10 described serotypes, Ia, Ib and II-IX (Yao *et al.*, 2013). Piscine strains of *S. agalactiae* may present total or absent hemolysis and are mostly serotypes Ia, Ib and III, but while Ia and III occur predominantly in Asia and are commonly hemolytic, Ib is commonly related in Americas and Europe and does not cause hemolysis (Delannoy *et al*. 2016). Nevertheless human and bovine isolates are mostly β-hemolytic (Elliott *et al.*, 1990; Evans *et al.*, 2002; Evans *et al.*, 2008; Delannoy *et al.*, 2016). The mostly reported serotypes of human GBS strains are Ia, II, III and V, as in bovines, serotypes Ia and III are most reported (Chen *et al.*, 2015). The diagnosis of GBS can be performed biochemically, through both ready-to-use kits as API® 20 Strep (BioMérieux, France) or manually, as described on Bergey's Manual of Bacteriology (Holt, 1994). Also, there are molecular diagnosis tests through specific PCR or qPCR (Bergh *et al.*, 2004; Mata *et al.*, 2004; Su *et al.*, 2016). Some of the non-hemolytic piscine streptococci isolates were, in the past, erroneously named *Streptococcus difficile*, which was primarily described in Israel as an agent of septicemia in Nile tilapia, but genetic similarity studies showed it to be, actually, *S. agalactiae* (Berridge *et al.*, 2001).

3.2.1. DISEASES BY GBS IN HUMAN

GBS is responsible for causing diseases in both newborn infants and immunocompromised adults. In those adults, this bacterium may cause invasive diseases as meningitis and pneumonia or systemic infection (Baker, 1995; Dermer *et al.*, 2004; Edwards and Baker, 2005; Rajagopal, 2009). However, the newborns are the main affected population by this pathogen. The vaginal colonization is the main risk factor for the transmission to the fetus, and this can occur in the pre-parturition, when the bacterium can cause inflammation in the placenta membranes with a risk of disruption and fetal amniotic aspiration, or during the childbirth, by direct contact of the child to the vagina mucus (Dermer *et al.*, 2004; Johri *et al.*, 2006; Almeida *et al.*, 2015). The infected newborn may have two distinct progressions, an early- or a late-onset disease (EOD and LOD, respectively). The EOD occurs in the first week of life, and correspond to 80% of the cases (Schuchat, 1998), and the main occurrences are respiratory insufficiency, pneumonia, bacteremia and septic shock (Rajagopal, 2009). On the other hand, the LOD occurs after the first week and before the third month of life, after the systemic infection and colonization of other body sites, mainly central nervous system, causing meningitis (Schuchat, 1998; Rajagopal, 2009; Almeida *et al.*, 2015). The children who survive from LOD often acquire neurologic sequelae, which may be severe (Schuchat, 1998).

3.2.2. BOVINE STREPTOCOCCOSIS BY GBS

Streptococcus agalactiae is a parasite of the bovine mammary gland (Mcdonald, 1977), and it is highly contagious (Keefe, 1997). It is responsible to cause mastitis, mostly chronic and

subclinical, with some economic impact on the milk production in many countries. Hotspots of GBS mastitis are, currently, Europe and South Africa (Mweu *et al.*, 2012; Petzer *et al.*, 2013; Jorgensen *et al.*, 2016). In cows, GBS can survive for long periods on the mammary gland, and its transmission occurs by direct contact between animals, by fomites, or orofecally (Keefe, 1997; Petzer *et al.*, 2013). The occurrence of mastitis is often linked to a bad sanitary management, as in the milking process and tits and equipment sanitation, or a failed biosecurity system (Hillerton and Berry, 2005). The subclinical infection of GBS in the mammary gland has a negative impact on the milk quality, that is reflected by the raise of the somatic cell count (Lopes Júnior *et al.*, 2012). Sometimes, there can be a clinical infection, in which there can be clinical signals in the milk, as a yellowish coloration, pus, blood and nodules, or alterations in the udder, as edema, redness and the response to the touch, or even systemic signals in cows as fever, anorexia, and reduction in the milk production (Lago *et al.*, 2011; 2011).

3.2.3. *S. agalactiae* INFECTIONS IN FISH

Streptococcosis by GBS in fish is a worldwide issue. Inside five continents, this disease has been already related in many countries as China, Vietnam, Thailand, Kuwait, Iran, Israel, Egypt, Australia, Malaysia, Indonesia, Philippines, Ireland, USA, Honduras, Costa Rica, Colombia, Peru and Brazil (Eldar, 1995 ; Evans, 2002; Evans *et al.*, 2008; Hernandez *et al.*, 2009; Mian *et al.*, 2009; Ye *et al.*, 2011; Amal *et al.*, 2012; Bowater *et al.*, 2012; Ruane *et al.*, 2013; Ortega Asencios *et al.*, 2016).Besides affecting tilapia farming, streptococcosis has been related in many fresh and saltwater fish species, as the golden shiner (*Notemigonus crysoleucas*), silver pomfret (*Pampus artenteus*), golden pompano (*Trachinotus blochii*), hybrid striped bass (*Morone saxatilis x M. chrysops*), Gilt-head bream (*Sparus aurata*), mullet (*Liza klunzingeri*), and rainbow trout (*Onchorhynchus mykiss*) (Robinson and Meyer, 1966; Plumb *et al.*, 1974; Eldar *et al.*, 1995; Evans, 2002; Duremdez *et al.*, 2004; Mian *et al.*, 2009; Pourgholam, 2011; Soto *et al.*, 2015). This disease often occurs as outbreaks with high morbidity and mortality, sometimes above 90% (Mian *et al*., 2009; Pereira *et al*., 2010, Ye *et al.*, 2011), and some of main risk factors are the increase of water temperature over than 27°C, high stocking density and intensive husbandry (Evans *et al*., 2002; Figueiredo *et al*., 2006; Tavares *et al*., 2015). The most frequent clinical signals in fish with streptococcosis are ascites, anorexia, lethargy, exophthalmia, hyperaemia in the basis of the fins, and the neurological signal of erratic swimming, because of the ability of GBS to overpass the blood-brain barrier and infect the central nervous system (CNS), (Eldar *et al.*, 1995; Evans *et al.*, 2002; Mian *et al.*, 2009).

3.3. MULTILOCUS SEQUENCE TYPING OF *S. agalactiae*

S. agalactiae strains have been typed by different methods for epidemiological studies of streptococcosis, whereas it occurs globally and with a broad host range. The capsular polysaccharide was the first target for typing, what brought us to the current ten described serotypes (Ia, Ib, II-X) (Yao *et al*., 2013). The current genetic pillar of GBS epidemiology is the multilocus sequence typing, which consists in sequencing and typing seven housekeeping genes (*adhP* – alcohol dehydrogenase, *pheS* – phenylalanine tRNA synthetase, *atr* – amino acid

transporter, *glnA* – glutamine synthetase, *sdhA* – serine dehydratase, *glcK* – glucokinase and *tkt* – transketolase), combining them into sequence-types (STs) (Jones *et al*., 2003). Furthermore, through the algorithm BURST (Based Upon Related Sequences) on the software eBURST (Feil *et al*., 2004), the STs can be grouped if they share many alleles into clonal complexes (CCs), which share an evolutionary story.

MLST catches mutational events in the seven housekeeping genes, and it uses alleles as a basis of comparison, instead of nucleotide sequences. So, the number of polymorphic changes in each gene sequence is not taken in consideration in the establishment of a new allele, but it is taken as a single genetic event (Maiden *et al.*, 2013). Clonal complex is an informal definition, named when there is a central ST and its related ones, which must share enough alleles between themselves, and the identification of a CC depends on its frequency and longevity on the analyzed population (Maiden, 2006). MLST counts with the principle of a not necessarily clonal structure of bacterial population, in which genetic mutations can perpetuate to descents not only by binary fission after a mutational event, but by lateral genetic transfer (Maiden, 2006; Belén *et al*., 2009). The MLST database houses the deposited alleles sequences, makes it publicly available to everyone and allows a rapid genetic comparison study.

GBS strains from fish were already typed as ST-7, ST-103, ST-257, ST-258, ST-259, ST-260, ST-261, ST-552 and ST-553, combined with the different capsular serotypes, mainly Ia, Ib or III (Evans *et al*., 2008; Godoy *et al.*, 2013). The CC552 is the only one arising exclusively from piscine isolates, and comprises ST-552 (founder), ST-553, ST-260, ST-257, ST-259 and is closely related to the STs 261 (Evans *et al*. 2008) and ST-246 (Brochet *et al*. 2006), both constituting a CC with an unknown founder (Godoy *et al.*, 2013). Although, other STs and CCs harbour strains from fish origin, as ST-7, from CC7 (Evans *et al.*, 2008), and ST-103, a singleton (Godoy *et al.*, 2013), but these lineages also encompass strains from other origins as humans, bovines, cats and guinea pig (Brochet *et al*., 2006; Bohnsack *et al*., 2008; Evans *et al.*, 2008; Springman *et al*., 2008; Godoy *et al*., 2013). The eBURST analysis revealed that around 96% of the piscine Brazilian strains of GBS belonged to CC552 (Godoy *et al*., 2013), together with strains from Honduras and USA (Evans *et al*., 2008), Costa Rica and Colombia (Delannoy *et al*. 2013). CC552 and CC7 are known to be the major lineages of fish-isolated GBS worldwide (Evans *et al.*, 2008; Delannoy, 2013), but while CC7 occurrence is very limited to Asia, CC552 is being reported worldwide in a larger number of fish species (Delannoy *et al.*, 2016).

MLST genotyping is in constant evolution, since new alleles come arising. Even though this technique is still the most employed for *S. agalactiae* epidemiological studies, it lacks resolution to study strains that share a ST, which may be epidemiologically distant from different hosts, as humans and bovines (Sorensen *et al.*, 2010), or may have arisen from a single outbreak in a farm (Godoy *et al.*, 2013), and it happens due to the little variation which is expected for the constitutive genes.

3.4. EPIDEMIOLOGY OF *S. agalactiae* ON THE POST-GENOMIC ERA

With the advent of the new sequencing methods, known as NGS (*next generation sequencing*), the post-genomic era of microbiology advanced (Maiden *et al.*, 2013). By sequencing platforms as 454 GS FLX system (Roche), Illumina GA IIx (Illumina), HeliScope (Helicos), SOLiD 5500 XL system (ABI), PacBio RS system (Pacific Biosciences) and Personal Genome Machine - Ion Torrent (Life Technologies) the amount of genomic data and information raised quickly (Kato, 2009), and bioinformatics is a major pillar for the interpretation of these data (Croucher *et al.*, 2013). So, with the advent and growth of the genomics, the whole genome sequencing (WGS) is allowing to bigger resolution in epidemiological studies of streptococcosis (Brochet *et al.*, 2006; Liu *et al.*, 2013; Puymege *et al.*, 2015; Delannoy *et al.*, 2016).

The post genomic era brought new tools on the microbiological field, and *S. agalactiae* genomes have been used on reverse vaccinology (Maione *et al.*, 2005; Pereira *et al.*, 2013) and as support to the development of new diagnostic methods (Su *et al.*, 2016). On the other hand, complete genomes are allowing the advance of epidemiological and evolutionary studies of *Streptococcus agalactiae* and streptococcosis. There are, currently (January 14th, 2017), about 30 genomes of fish-origin GBS on NCBI database, and less than the half of them comprises complete sequences. However, there is a total of 891 *S. agalactiae* genomes on NCBI, among which only 35 are complete. Comparative genomics have been used to study GBS. From comparative studies, it is possible to search for shared and unique genes between strains, insertion and deletion (indel) events or mutations, and predict the mRNA transcripts and proteins (Wang *et al.*, 2015).

Population genomic studies can split the genome on the core genome which is the part shared by every strain, the dispensable genome, which comprises the parts present in some strains, but not all, and the new genes, which are strain-specific. The joining of all the parts gives birth to the pan-genome, which is the full repertoire of genes for that species in that time (Tettelin *et al.*, 2005; Wang *et al.*, 2015). A previous study by Tettelin *et al.* (2005) studied eight human GBS genomes, one complete and seven drafts, and showed an opened pan-genome, which means a remarkable addition of new genes when a new strain is added to the population, 33 in this case. It was seen that, the core genome comprised housekeeping functions, as molecules transporting and binding proteins, the dispensable genome comprised mostly genes with unknown function and, both dispensable and strain-specific genome harbored a big amount of horizontally acquired genes, which are mainly organized into genomic islands. A genomic island is an entire segment acquired from other organism by horizontal gene transfer (HGT) and inserted in the genome (Pallen and Wren, 2007; Soares *et al.*, 2015) It was suggested by that *Streptococcus agalactiae* differences of virulence, environmental and host predilections are due to this genomic differences between the strains, which may be taken individually from other bacteria by HGT (Tettelin *et al.*, 2005). A pan-genome study on a population of 138 strains of bovine origin and from CC61 from Portugal (Almeida *et al.*, 2016), and it was seen that, inside the accessory genome, a variant of the lactose operon, highly similar to *S. uberis* and *S. dysgalactiae* subsp. *dysgalactiae*, was present and conserved in some strains of this population, and it is known that it contributes to the pathogenesis of mastitis (Richards *et al.*, 2011).

Streptococcus agalactiae genomes show a pool of variable genes in spite of a stable "backbone", and this variation in metabolic features is believed to be linked to the vast environment and host interactions (Brochet *et al.*, 2006).

A study by Liu *et al.* (2013) with 15 GBS strains from human, bovine and fish origin from China, North America, Latin America, and Europe was performed to study the phylogenetics of this organism and host-associated traits. The core-genome of the strains represented 54.7% of all the genes, and each strain had a considerably difference in the accessory genome. It supports the idea of the not-essential genome being linked to the different host and environmental associations by Brochet *et al.* (2006). It was seen by Liu *et al.* (2013) that the Latin-American CC552 piscine strains SA20-06 (ST-553) and STIR-CD-17 (ST-260) share more specific genes with each other than the other strains, even Chinese fish strains from ST-7, and had a smaller genome (about 1.8 mb) than the other isolates (>2 mb). It supports the previous evidences of a reductive evolution of the fish and frog isolates from STs 261, 260 and 552 by Rosinski-Chupin (2013), in which pseudogenization is a highlight.

The single nucleotide polymorphisms (SNP)-based phylogeny is largely applied in evolutionary studies of *S. agalactiae*. Liu *et al.* (2013) made a SNP-phylogenetic analysis of all orthologous clusters found in the alignment of 15 *S. agalactiae* genomes, nine from fish divided into CC7, STs-261, 552 and 261, three from human, divided into CC7, CC19 and CC23, one bovine from CC67, and two from frog, one ST-261 and one ST-260. By this phylogenetic analysis, the authors showed that the two fish GBS isolates from ST-7, which predominate in Asia, are very close to two human strains from STs-7 and ST-6, both CC7, and besides that, all these ST-7 strains were genomically similar, both in genome size (around 2 mb) and in gene content. Delannoy *et al.* (2016) studied 16 GBS isolates, two CC552 from fish, two CC67 from bovine, one CC17 from human, two CC23, one from human and one from seal, two CC19 from human, one CC1 from human, and five CC7, three from fish and two from human origins, all epidemiologically independent events. Their phylogenetic analysis of the SNPs from the core-genome supported the analysis by Liu *et al.* (2013), as the CC7 strains were very close, but very distant from CC552 strains from fish. Furthermore, the genomic comparison could separate eight loci which are present on all the fish-GBS genomes, but only one was shared by CCs 7 and 552, being called fish-associated while seven were CC552 exclusives, named fish-specific.

Next generation sequencing is spreading in clinical environments as diagnosis laboratories and hospitals. Through NGS, it is possible to sequence multiple isolates of pathogens in a single run, from the same species or from different ones, with metagenomics (Deurenberg *et al.*, 2017). Currently, some molecular epidemiological databases are under adaptation for NGS data, as MLST for the Enterobacteria from genera *Salmonella, Shigella, Escherichia, Yersinia* and *Moraxella*, for which a specific database was created, the Enterobase (available at http://enterobase.warwick.ac.uk/).

Recently, a new genetic tool is emerging as an extended MLST. Through NGS data, it is possible to choose as many alleles as wanted to type on each isolate, even the whole genome (wgMLST), so the discrimination power becomes remarkably bigger if compared to the classic

seven alleles typing MLST (Jolley and Maiden, 2010). In this context, the BIGSdb (Bacterial Isolate Genome Sequence Database) tool (Jolley and Maiden, 2010) was developed to handle with NGS data and it allows the execution of the extended MLST and it allows to discern strains from situations of difficult epidemiological discrimination, such as isolates from a single host or a single epidemiological unit (Maiden *et al.*, 2013). The wgMLST types each of all the alleles of a provided genome sequence. So, differently of MLST and SNP analysis, the events of HGT, which comprises one of the mechanisms of bacterial evolution and diversification, is taken in consideration (Maiden *et al.*, 2013; Figueiredo *et al.*, 2015) to study a population evolution. A recent work from Figueiredo *et al.* (2015) with the Brazilian emerging pathogen *Weissella ceti* sequenced three genomes and it performed a wgMLST-phylogenetic analysis with another American strain of this bacterium and other species of the same genus. The study showed that, besides the big genomic similarity in the genetic content on the Brazilian strains, there were inter-farms variation, each isolate was a single clone, and showed a close relationship between a strain from the first outbreaks in Brazil with a contemporary American isolate. It is believed that the integration of the principles of the classical microbial typing methods and genomics is the new advance for microbial molecular epidemiology.

3.5. REFERENCES

ALMEIDA, A.; ALVES-BARROCO, C.; SAUVAGE, E. *et al*. Persistence of a dominant bovine lineage of group B Streptococcus reveals genomic signatures of host adaptation. *Environ Microbiol*, v.18, p.4216-4229, 2016.

ALMEIDA, A.; VILLAIN, A.; JOUBREL, C. *et al*. Whole-Genome Comparison Uncovers Genomic Mutations between Group B Streptococci Sampled from Infected Newborns and Their Mothers. *J Bacteriol*, v.197, p.3354-3366, 2015.

AMAL, M.N.; ZAMRI-SAAD, M.; IFTIKHAR, A.R. *et al.* An outbreak of Streptococcus agalactiae infection in cage-cultured golden pompano, Trachinotus blochii (Lacepede), in Malaysia. *J Fish Dis*, v.35, p.849-852, 2012.

ASENCIOS, Y.O.; SÁNCHEZ, F.B.; MENDIZÁBAL, H.B. *et al*. First report of Streptococcus agalactiae isolated from Oreochromis niloticus in Piura, Peru: Molecular identification and histopathological lesions. *Aquaculture Reports*, v.4, p.74-79, 2016.

ASSIS, G.B.; TAVARES, G.C.; PEREIRA, F.L. *et al.* Natural coinfection by Streptococcus agalactiae and Francisella noatunensis subsp. orientalis in farmed Nile tilapia (Oreochromis niloticus L.). *J Fish Dis*, 2016.

BAKER, C. J.; EDWARDS, M.S. Group B streptococcal infections. In: Remington, J.S.; Klein, J.O. *Infectious Disease of the Fetus and New-born Infant*. USA, TheW.B. Saunders Co, 1995. p.980-1054.

BELÉN, A.; PAVÓN, I.; MAIDEN, M.C.J. Multilocus Sequence Typing. *Methods in molecular biology* (Clifton, N.J.), v.551, p.129-140, 2009.

BERGH, K.; STOELHAUG, A.; LOESETH, K. *et al*. Detection of group B streptococci (GBS) in vaginal swabs using real-time PCR with TaqMan probe hybridization. *Indian J Med Res*, v.119 Suppl, p.221-223, 2004.

BERRIDGE, B.R.; BERCOVIER, H.; FRELIER, P.F. Streptococcus agalactiae and Streptococcus difficile 16S-23S intergenic rDNA: genetic homogeneity and species-specific PCR. *Vet Microbiol*, v.78, p.165-173, 2001.

BOHNSACK, J.F.; WHITING, A.; GOTTSCHALK, M. *et al.* Population structure of invasive and colonizing strains of *Streptococcus agalactiae* from neonates of six U.S. Academic Centers from 1995 to 1999. *J Clin Microbiol*, v.46, p.1285-1291, 2008.
BOWATER, R.O.; FORBES-FAULKNER, J.; ANDERSON, I.G. *et al*. Natural outbreak of Streptococcus agalactiae (GBS) infection in wild giant Queensland grouper, Epinephelus lanceolatus (Bloch), and other wild fish in northern Queensland, Australia. *J Fish Dis*, v.35, p.173-186, 2012.
BOLETIM ESTATÍSTICO DA PESCA E AQUICULTURA 2011. Brasília: Ministério da Pesca e Aquicultura, 2013. Available in: <http://www.icmbio.gov.br/cepsul/images/stories/biblioteca/download/estatistica/est_2011_bol__bra.pdf>. Accessed in: 01/20/2017.
BROCHET, M.; COUVE, E.; ZOUINE, M. *et al*. Genomic diversity and evolution within the species Streptococcus agalactiae. *Microbes Infect*, v.8, p.1227-1243, 2006.
CARVALHO-CASTRO, G.A.; LOPES, C.O.; LEAL, C.A. *et al*. Detection of type III secretion system genes in Aeromonas hydrophila and their relationship with virulence in Nile tilapia. *Vet Microbiol*, v.144, p.371-376, 2010.
CHEN, M.; WANG, R.; LUO, F.G. *et al.* Streptococcus agalactiae isolates of serotypes Ia, III and V from human and cow are able to infect tilapia. *Vet Microbiol*, v.180, p.129-135, 2015.
CREPALDI, D.V., FARIA, P.M.C., TEIXEIRA, E.A. *et al.* A situação da Aquacultura no Brasil e no Mundo. *Rev. Bras. Reprod. Anim.*, v.30, p.81-85, 2006.
CROUCHER, N.J.; HARRIS, S.R.; GRAD, Y.H. *et al*. Bacterial genomes in epidemiology--present and future. P*hilos Trans R Soc Lond B Biol Sci*, v.368, p.20120202, 2013.
DELANNOY, C.M.; ZADOKS, R.N.; CRUMLISH, M. *et al*. Genomic comparison of virulent and non-virulent Streptococcus agalactiae in fish. *J Fish Dis*, v.39, p.13-29, 2016.
DELANNOY, C.M.J., CRUMLISH, M., FONTAINE, M. C. *et al*. Human Streptococcus agalactiae strains in aquatic mammals and fish. *Bmc Microbiology*, v.13, 2013.
DERMER, P.; LEE, C.; EGGERT, J. *et al*. A history of neonatal group B streptococcus with its related morbidity and mortality rates in the United States. J Pediatr Nurs, v.19, p.357-363, 2004.
DEURENBERG, R.J., BATHOORN, E., CHLEBOWICZ, M.N. *et al*. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol*, v.243, p.16-24, 2017.
DOW, S.W.; JONES, R.L.; THOMAS, T.N. *et al*. Group B streptococcal infection in two cats. *J Am Vet Med Assoc*, v.190, p.71-72, 1987.
DUREMDEZ, R.; AL-MARZOUK, A.; QASEM, J.A. *et al*. Isolation of Streptococcus agalactiae from cultured silver pomfret, Pampus argenteus (Euphrasen), in Kuwait. *J Fish Dis*, v.27, p.307-310, 2004.
EDELSTEIN, R.M.; PEGRAM, R.G. Contagious skin necrosis of Somali camels associated with Streptococcus agalactiae. *Trop Anim Health Prod*, v.6, p.255-256, 1974.
EDWARDS, M.S.; BAKER, C.J. Group B streptococcal infections in elderly adults. *Clin Infect Dis*, v.41, p.839-847, 2005.
ELDAR, A., BEJERANO, Y., LIVOFF, A. *et al*. Experimental streptococcal meningo-encephalitis in cultured fish. *Vet Microbiol*, v.43, p.33-40, 1995
ELLIOTT, J.A.; FACKLAM, R.R.; RICHTER, C.B. Whole-cell protein patterns of nonhemolytic group B, type Ib, streptococci isolated from humans, mice, cattle, frogs, and fish. *J Clin Microbiol*, v.28, p.628-630, 1990.
EVANS, J.J.; BOHNSACK, J.F.; KLESIUS, P.H. *et al*. 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 is also associated with human neonatal infections in Japan. *J Med Microbiol*, v.57, p.1369-1376, 2008.

EVANS, J.J.; KLESIUS, P.H.; GILBERT, P.M. *et al*. Characterization of β-haemolytic Group B Streptococcus agalactiae in cultured seabream, Sparus auratus L., and wild mullet, Liza klunzingeri (Day), in Kuwait. *J Fish Dis*, v.25, p.505-513, 2002.
CULTURED Aquatic Species Information Programme - Oreochromis niloticus (Linnaeus, 1758). Available in: http://www.fao.org/fishery/culturedspecies/Oreochromis_niloticus/en#tcNA00EA. Accessed in: 01/10/2017.
THE State of World Fisheries and Aquaculture 2014. Rome: FAO, 2014. Available in: <http://www.fao.org/3/a-i3720e.pdf>. Accessed in: 01/10/2017.
THE State of World Fisheries and Aquaculture 2016. Rome: FAO, 2016. Available in: <http://www.fao.org/3/a-i5555e.pdf>. Accessed in: 01/10/2017.
FEIL, E.J.; LI, B.C.; AANENSEN, D.M. *et al*. eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. *J Bacteriol*, v.186, p.1518-1530, 2004.
FIGUEIREDO, H.C.P.; CARNEIRO, D.O.; FARIA, F.C. *et al*. *Streptococcus agalactiae* associado à meningoencefalite e infecção sistêmica em tilápia-do-nilo (*Oreochromis niloticus*) no Brasil. *Arq Bras Med Vet Zootec*, v.58, p.678-680, 2006.
FIGUEIREDO, H.C.; KLESIUS, P.H.; ARIAS, C.R. *et al*. Isolation and characterization of strains of Flavobacterium columnare from Brazil. *J Fish Dis*, v.28, p.199-204, 2005.
FIGUEIREDO, H.C.P.; LEAL, C.A.G. Infecções por *Streptococcus spp*. em peixes. In: SILVA-SOUZA, A.T., LIZAMA, M.A.P, TAKEMOTO, R.M. P*atologia e Sanidade de Organismos Aquáticos*. 1.ed. Maringá: ABRAPOA, 2012. p. 275-314.
FIGUEIREDO, H.C.P.; SOARES, S.C.; PEREIRA, F.L. *et al*. Comparative genome analysis of Weissella ceti, an emerging pathogen of farm-raised rainbow trout. *BMC Genomics*, v.16, 2015.
GODOY, D.T.; CARVALHO-CASTRO, G.A.; LEAL, C.A. *et al*. Genetic diversity and new genotyping scheme for fish pathogenic Streptococcus agalactiae. *Lett Appl Microbiol*, v.57, p.476-483, 2013.
HERNANDEZ, E.; FIGUEROA, J.; IREGUI, C. Streptococcosis on a red tilapia, Oreochromis sp., farm: a case study. *J Fish Dis*, v.32, p.247-252, 2009.
HILLERTON, J.E.; BERRY, E.A. Treating mastitis in the cow--a tradition or an archaism. *J Appl Microbiol*, v.98, p.1250-1255, 2005.
HOLT, J.G.; SNEATH, P.H.; KRIEG, N.R. *et al*. *Bergey's Manual of Determinative Bacteriology*. 9. ed. Baltimore: Lippincott Williams & Wilkins, 1994. 787p.
PRODUÇÃO da Pecuária Municipal 2015. Rio de Janeiro: IBGE, 2015. Available in: <http://www.ibge.gov.br/home/estatistica/economia/ppm/2015/default.shtm>. Accessed in: 12/12/2016.
JOHRI, A.K.; PAOLETTI, L.C.; GLASER, P. *et al*. Group B Streptococcus: global incidence and vaccine development. *Nat Rev Microbiol*, v.4, p.932-942, 2006.
JOLLEY, K.A.; MAIDEN, M.C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, v.11, p.595, 2010.
JONES, N.; BOHNSACK, J.F.; TAKAHASHI, S. *et al*. Multilocus Sequence Typing System for Group B Streptococcus. *J Clin Microbiol*, v.41, p.2530-2536, 2003.
JORGENSEN, H.J.; NORDSTOGA, A.B.; SVILAND, S. *et al*. Streptococcus agalactiae in the environment of bovine dairy herds--rewriting the textbooks? *Vet Microbiol*, v.184, p.64-72, 2016.
KATO, K. Impact of the next generation DNA sequencers. *International J Clin Exp Med*, v.2, p.193-202, 2009.
KEEFE, G.P. Streptococcus agalactiae mastitis: a review. *Can Vet J*, v.38, p.429-437, 1997.

LAGO, A.; GODDEN, S.M.; BEY, R. *et al*. The selective treatment of clinical mastitis based on on-farm culture results: I. Effects on antibiotic use, milk withholding time, and short-term clinical and bacteriological outcomes. *J Dairy Sci*, v.94, p.4441-4456, 2011.
LEAL, C.A.; TAVARES, G.C.; FIGUEIREDO, H.C. Outbreaks and genetic diversity of Francisella noatunensis subsp orientalis isolated from farm-raised Nile tilapia (Oreochromis niloticus) in Brazil. *Genet Mol Res*, v.13, p.5704-5712, 2014.
LIU, G.; ZHANG, W.; LU, C. Comparative genomics analysis of Streptococcus agalactiae reveals that isolates from cultured tilapia in China are closely related to the human strain A909. *BMC Genomics*, v.14, p.775, 2013.
LOPES JÚNIOR, J.E.F.; LANGE, C.C.; BRITO, M.A.V.P. *et al*. Relationship between total bacteria counts and somatic cell counts from mammary quarters infected by mastitis pathogens. *Ciênc Rural*, v.42, p.691-696, 2012.
MAIDEN, M.C. Multilocus sequence typing of bacteria. *Annu Rev Microbiol*, v.60, p.561-588, 2006.
MAIDEN, M.C.; JANSEN VAN RENSBURG, M.J.; BRAY, J.E. *et al*. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol*, v.11, p.728-736, 2013.
MAIONE, D.; MARGARIT, I.; RINAUDO, C.D. *et al*. Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science*, v.309, p.148-150, 2005.
MATA, A.I.; GIBELLO, A.; CASAMAYOR, A. *et al*. Multiplex PCR assay for detection of bacterial pathogens associated with warm-water Streptococcosis in fish. *Appl Environ Microbiol*, v.70, p.3183-3187, 2004.
MCDONALD, J.S. Streptococcal and staphylococcal mastitis. *J Am Vet Med Assoc*, v.170, p.1157-1159, 1977.
MIAN, G.F.; GODOY, D.T.; LEAL, C.A. *et al.* Aspects of the natural history and virulence of S. agalactiae infection in Nile tilapia. *Vet Microbiol*, v.136, p.180-183, 2009.
MWEU, M.M.; TOFT, N.; KATHOLM, J. *et al*. Evaluation of two herd-level diagnostic tests for Streptococcus agalactiae using a latent class approach. *Vet Microbiol*, v.159, p.181-186, 2012.
NETTO, L.N.; LEAL, C.A.; FIGUEIREDO, H.C. Streptococcus dysgalactiae as an agent of septicaemia in Nile tilapia, Oreochromis niloticus (L.). *J Fish Dis*, v.34, p.251-254, 2011.
PALLEN, M.J.; WREN, B.W. Bacterial pathogenomics. *Nature*, v.449, p.835-842, 2007.
PEREIRA, U.P.; SOARES, S.C.; BLOM, J. *et al*. In silico prediction of conserved vaccine targets in Streptococcus agalactiae strains isolated from fish, cattle, and human samples. *Genet Mol Res*, v.12, p.2902-2912, 2013.
PETZER, I.-M.; KARZIS, J.; LESOSKY, M. *et al*. Host adapted intramammary infections in pregnant heifers which were co-housed and reared on fresh milk as calves. *BMC Vet Res*, v.9, p.49, 2013.
PLUMB, J.A.; SCHACHTE, J.H.; GAINES, J.L. *et al*. Streptococcus sp. from Marine Fishes Along the Alabama and Northwest Florida Coast of the Gulf of Mexico. *Trans Am Fish Soc*, v.103, p.358-361, 1974.
POURGHOLAM, R., LALUEI, F., SAEEDI, A.A. *et al.* Distribution and Molecular identification of some causative agents of streptococcosis isolated from farmed rainbow trout (*Oncorhynchus mykiss*, Walbaum) in Iran. *Iranian J Fish Sci*, v.10, p.109-122, 2011.
PUYMEGE, A.; BERTIN, S.; GUEDON, G. *et al*. Analysis of Streptococcus agalactiae pan-genome for prevalence, diversity and functionality of integrative and conjugative or mobilizable elements integrated in the tRNA(Lys CTT) gene. *Mol Genet Genomics*, v.290, p.1727-1740, 2015.
RAJAGOPAL, L. Understanding the regulation of Group B Streptococcal virulence factors. *Future Microbiol*, v.4, p.201-221, 2009.

RICHARDS, V.P.; LANG, P.; BITAR, P.D. *et al*. Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted Streptococcus agalactiae. *Infect Genet Evol*, v.11, p.1263-1275, 2011.
ROBINSON, J.A.; MEYER, F.P. Streptococcal fish pathogen. *J Bacteriol*, v.92, p.512, 1966.
ROSINSKI-CHUPIN, I., SAUVAGE, E., MAIREY, B. *et al*. Reductive evolution in Streptococcus agalactiae and the emergence of a host adapted lineage. *BMC Genomics*, v.14, 2013.
RUANE, N.M.; COLLINS, E.M.; GEARY, M. *et al*. Isolation of Streptococcus agalactiae and an aquatic birnavirus from doctor fish Garra rufa L. *Irish Vet J*, v.66, p.16-16, 2013.
SCHUCHAT, A. Epidemiology of group B streptococcal disease in the United States: shifting paradigms. *Clin Microbiol Rev*, v.11, p.497-513, 1998.
SEBASTIÃO, F.A.; NOMURA, D.; SAKABE, R. *et al*. Hematology and Productive Performance of Nile Tilapia (Oreochromis niloticus) Naturally Infected With Flavobacterium columnare. *Braz J Microbiol*, v.42, p.282-289, 2011.
SOARES, S.C.; GEYIK, H.; RAMOS, R.T. *et al*. GIPSy: Genomic island prediction software. *J Biotechnol*, 2015.
SORENSEN, U.B.; POULSEN, K.; GHEZZO, C. *et al*. Emergence and global dissemination of host-specific Streptococcus agalactiae clones. *MBio*, v.1, 2010.
SOTO, E.; WANG, R.; WILES, J. *et al*. Characterization of Isolates of Streptococcus agalactiae from Diseased Farmed and Wild Marine Fish from the U.S. Gulf Coast, Latin America, and Thailand. *J Aquat Anim Health*, v.27, p.123-134, 2015.
SPRINGMAN, A.C.; LACHER, D.W.; WU, G. *et al*. Selection, recombination, and virulence gene diversity among group B streptococcal genotypes. *J Bacteriol*, v.191, p.5419-5427, 2009.
SU, Y.L.; FENG, J.; LI, Y.W. *et al*. Development of a quantitative PCR assay for monitoring Streptococcus agalactiae colonization and tissue tropism in experimentally infected tilapia. *J Fish Dis*, v.39, p.229-238, 2016.
TETTELIN, H.; MASIGNANI, V.; CIESLEWICZ, M.J. *et al*. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc Natl Acad Sci*, v.102, p.13950-13955, 2005.
VANDAMME, P.; DEVRIESE, L.A.; POT, B. *et al*. Streptococcus difficile is a nonhemolytic group B, type Ib Streptococcus. *Int J Syst Bacteriol*, v.47, p.81-85, 1997.
WANG, R.; LI, L.; HUANG, Y. *et al*. Comparative genome analysis identifies two large deletions in the genome of highly-passaged attenuated Streptococcus agalactiae strain YM001 compared to the parental pathogenic strain HN016. *BMC Genomics*, v.16, p.897, 2015.
YAO, K.; POULSEN, K.; MAIONE, D. *et al*. Capsular gene typing of Streptococcus agalactiae compared to serotyping by latex agglutination. *J Clin Microbiol*, v.51, p.503-507, 2013.
YE, X.; LI, J.; LU, M. *et al*. Identification and molecular typing of Streptococcus agalactiae isolated from pond-cultured tilapia in China. *Fish Sci*, v.77, p.623-632, 2011.

4. CHAPTER TWO – Large-scale genomic analyses reveal the population structure and evolutionary trends of *Streptococcus agalactiae* strains in Brazilian fish farms

Original manuscript submitted to the Journal "Emerging Microbes & Infections". The formatting is according to the journal requirements.

Large-scale genomic analyses reveal the population structure and evolutionary trends of *Streptococcus agalactiae* strains in Brazilian fish farms

G. M. Barony[1*], G. C. Tavares[1*], F. L. Pereira[1*], A. F. Carvalho[1*], F. A. Dorella[1*], C. A. G. Leal[1*], H. C. P. Figueiredo[1*]

[1] National Reference Laboratory for Aquatic Animal Diseases (AQUACEN) of Ministry of Agriculture, Livestock and Food Supply, Federal University of Minas Gerais, Belo Horizonte, Brazil.

Correspondence: HCP Figueiredo, E-mail: figueiredoh@ufmg,br

[*]These authors contributed equally to this work

Abstract

Streptococcus agalactiae is a major pathogen and a hindrance on tilapia farming worldwide. The aims of this work were to analyse the Brazilian population of *S. agalactiae,* to establish a method to track strains associated with outbreaks of streptococcosis and to distinguish closely related strains. A total of 39 strains were obtained from outbreaks and their whole genomes were sequenced and annotated for comparative analysis of MLST, genomic similarity and wgMLST. The Brazilian strains presented two STs, among which a new one just described, and also a non-typeable lineage. The wgMLST could differentiate each strain in a single clone and to establish

temporal and geographical correlations among strains. The Bayesian evolutionary analysis revealed that the Brazilian population of *S. agalactiae* has a remarkably recent emergence, about 585 years ago. Brazilian population of *S. agalactiae* showed to be heterogeneous in its genome sequence, distributed in different regions of the country according to the genotype, and wgMLST analysis was able to track each outbreak event individually.

Keywords: MLST; streptococcosis; comparative genomics; tMRCA; tilapia; phylogenetics; wgMLST.

Introduction

Streptococcus agalactiae (Lancefield's group B Streptococcus, GBS) is a gram-positive coccus which causes septicaemia and meningoencephalitis in many species of marine and freshwater fish worldwide (Eldar, 1995 ; Evans, 2002; Mian *et al.*, 2009; Godoy *et al.*, 2013; Soto *et al.*, 2015). Also, this bacterium may cause septicaemia and meningitis in human new-borns (Bohnsack *et al.*, 2008) and has already been reported in other animals as guinea pig, camels, cats, dolphins, horses and frogs (Johri *et al.*, 2006). This disease is a major obstacle to expansion of Brazilian's aquaculture, because it has a high prevalence in Nile tilapia (*Oreochromis niloticus*) farming, the most produced fish in Brazil (Mian *et al.*, 2009). GBS streptococcosis in tilapia farming occurs mainly in

temperatures above 27° C and provokes high economic impact due to high mortalities and fast-evolving scenario (Mian *et al.*, 2009; Soto *et al.*, 2015).

Several genotyping methods have been used to study the population structure of *S. agalactiae* infecting humans and animals, including the possibility of cross-species transmission (Ismail and Anthony, 2012; Chen *et al.*, 2015). The main epidemiological tool applied in studies of GBS diseases is multilocus sequence typing (MLST) (Maiden *et al.*, 1998; Maiden, 2006), which can discriminate strains in lineages (sequence-type, ST) and combine them based on genetic proximity in clonal complexes (CC). This method has proved itself efficient to understand evolutionary stories between lineages and has so far discriminated fish isolates worldwide in several STs or CCs, as CC552 (Godoy *et al.*, 2013; Delannoy *et al.*, 2016), ST-261 and ST-246 (Evans *et al.*, 2008; Godoy *et al.*, 2013; Chen *et al.*, 2015). However, it does not have full capacity to discriminate many strains from different host species, geographical origin and dates, grouping them into the same sequence-type or complex clonal. Examples of this lack of resolution are several strains from human, bovine, feline and rodent sources belonging to the ST-103 (Brochet *et al.*, 2006; Bohnsack *et al.*, 2008; Springman *et al.*, 2009; Godoy *et al.*, 2013) and ST-7 which groups isolates from fish, human and dolphin (Evans *et al.*, 2008). A previous study revealed that Brazilian fish GBS were almost exclusively from CC552, being ST-552, ST-553 or ST-260 (Godoy *et al.*, 2013), and this CC was also seen predominant in Central America (Delannoy, 2013).

Notwithstanding, with the current genotyping methods it is usually not possible to obtain the epidemiological links between isolates and outbreak events of GBS streptococcosis, even using MLST, which uses only seven genes to discriminate strains. It has been previously thought and proved that by using a bigger set of genes or even all the genome, it would be possible to discriminate closely related strains of a given species even in single clones (Maiden *et al.*, 2013; Cody *et al.*, 2014), through a MLST analysis with hundreds of loci or the whole genome (wgMLST). So, wgMLST is emerging as a tool for epidemiological and evolutionary studies and has a promising future in molecular epidemiology of microbial pathogens. This work aimed to evaluate the population structure of *Streptococcus agalactiae* pathogenic to fish in Brazil and to establish a farm-to-farm approach that allow the epidemiological track of different bacterial clones through genomic-based methods.

Materials and Methods

Outbreaks and strains Isolation

A total of 39 strains of *S. agalactiae* isolated from different outbreaks of streptococcosis in Nile tilapia and Amazon hybrid catfish (*Pseudoplatystoma fasciatum* x *Leiarius marmoratus*) farms, between the years of 2003 and 2015 were evaluated (figure 1), representing all outbreaks assisted by the National Reference Laboratory for

Aquatic Animal Diseases of Brazilian Ministry of Agriculture, Livestock and Food Supply (AQUACEN). The clinical signals presented by fish were erratic swimming, exophthalmia and corneal opacity. The animals were sent to Aquacen well-conditioned on ice and were submitted to necropsy, and swabs of brain and kidney were spread on Blood Agar (Tryptic Soy Agar plus 5% of sheep blood) and grew at 28°C for 48h. Grown colonies were confirmed as *S. agalactiae* by MALDI-TOF (Bruker, Germany) mass spectrometry and were stored at -80°C in brain-heart infusion broth with 15% of glycerol until posterior use. The city, state, region of origin and year of isolation of each strain is described on supplementary table 1.

DNA extraction and genomic sequencing

For DNA obtainment, the strains were thawed and grown in blood agar at 28°C for 48h. Colonies were collected and their DNA was extracted using Maxwell® 16 Tissue DNA Purification Kit with Maxwell® 16 MDx Research Instrument (both from Promega, USA), following manufacturer's instructions. Genomic DNAs were quantified using Qubit 2.0 Fluorometer (Life Technologies, Thermo Scientific, USA). The strain SA20 (Pereira *et al.*, 2013) was previously re-sequenced (Pereira *et al.*, 2016) on MiSeq sequencer (Illumina, USA) with Nextera™ DNA Library Prep Kit. The other strains were sequenced on Ion Torrent Personal Genome Machine™ (PGM) (Life Technologies, USA), with Ion PGM Sequencing 200 Kit, all following manufacturer's recommendations.

Genomes' assembly and annotation

The quality of all the sequenced raw data was analysed by FastQC 0.11.1 (Andrews, 2010) and an in-house script was used to obtain reads with at least 20 on PHRED quality score. After this triage, the genomes were assembled *ab initio* on Newbler 2.0 (Roche, USA) and scaffolds were generated on CONTIGuator (Galardini, 2011), all using as reference the strain SA20, which was reassembled with optical mapping (Onmus-Leone, 2013). The gaps filling was performed manually on CLC Genomics Workbench 7.0 (Qiagen, USA), mapping the genomes to the reads and gradually extending the flanks of the gaps. Annotation was performed manually for SA20 (Pereira *et al.*, 2013; Pereira *et al.*, 2016) and the other strains were submitted to annotation transfer on the software prokka (Seemann, 2014) using nested basis on this order: CDSs from SA20, RefSeq database only with *Streptococcus agalactiae* proteins, and finally RefSeq's all proteins in databases, in this order of priority, being either annotation or manual curation done on Artemis (Rutherford, 2000) and CLC Genomics Workbench 7.0.

***In silico* capsular serotyping**

Based on the work of Sheppard *et al.* (2016), an in-house script was developed (available at https://github.com/aquacen/serotype_Sagalactiae) to access the capsular serotype of *S. agalactiae* isolates through the whole genome sequence file. The 39 strains of the present work had the serotype obtained.

MLST genotyping and eBURST analysis

All the seven MLST genes *adhP* (alcohol dehydrogenase), *pheS* (phenylalanyl-tRNA ligase subunit alpha), *atr* (amino acid ABC transporter), *glnA* (glutamine synthetase), *sdhA* (L-serine dehydratase subunit alpha), *glcK* (glucokinase) and *tkt* (transketolase) were selected from the 39 GBS whole genomes and all loci and STs were obtained by an in-house script (available at https://www.github.com/aquacen/mlst_Sagalactiae), which searches the described alleles deposited on pubMLST on the genomes through Blastn aligment. Moreover, one strain from each ST previously related on fish was picked from pubMLST and genbank databases, all data were submitted to the eBURST algorithm (Feil *et al.*, 2004) and clonal complexes were defined using Single Locus Variant bias.

Genomic similarity percentage analysis

The whole genomes of interest were used on Gegenees 2.0 (Agren *et al.*, 2012) in addition to two piscine strains which had complete genome sequences, *S. agalactiae* 138spar (ST-261) and GD201008-001 (ST-7) (GenBank access numbers CP007565 and CP003810, respectively), to obtain the percentage of similarity matrix, and then build a heatplot map of similarity. The used parameters by the software were fragmentation length of 200 nucleotides and threshold of 50%.

Prediction of polymorphic sites and phylogenomic analysis

The complete genome sequences of the 39 presented strains plus 138spar, GD201008-001, and the draft genome sequence of ST-491 *S. agalactiae* strain STIR-CD-14 (GenBank access number ANEJ01) were submitted to BIGSdb (Bacterial Isolate Genome Sequence Database) and all the obtained loci were compared one by one between isolates using the gene-by-gene approach (Maiden *et al.*, 2013). The distance matrix with the relative genomic divergence between all isolates was picked up and used on SplitsTree 4.0 (Huson and Bryant, 2006) to construct a phylogenomic tree with NeighborNet.

Substitution rates and Bayesian analysis

All the 39 strains of this work together with the *S. agalactiae* genomes available at Genbank that fulfilled the parameters of at least scaffolds status and number of scaffolds ≥ 30 (n = 103) were used to perform the substitution rates and Bayesian analysis. Using a high stringent BLASTn with an e-value of 1e-20, the coding sequences (CDSs) of SA20 were aligned against all the other strains. All the CDSs which met the following requirements: minimal percentage of identity of 98 %, difference of length between query and subject ≤ 5 bp, lacking paralogous sequences, and present in all the genomes (n = 382 genes) were extracted and concatenated to form the core genome of the species. The core genome of the 142 strains was aligned with MAFFT v7.302b (Katoh and Standley, 2013) with the parameter "–auto" enabled. The BEAUti package of BEAST v1.8.3 (Drummond and Rambaut, 2007) was used to generate the ".xml" file to run the Bayesian analysis with the following parameters: output files of MAFFT as

data, five taxa groups (Taxa I – all the non-typeable strains of this work; Taxa II – all strains of ST-927; Taxa III – all strains of ST-260; Taxa IV – all strains of group I, II, III; and Taxa V – all strains of group I, II, III and ST-261), tip dates, as described in Supplementary table 2, general time-reversible (GTR) model with gamma correction plus invariant sites, strict clock and coalescent exponential growth tree. Markov Chain Monte Carlo (MCMC) was performed on BEAST with 100 million generations with log collection in every 100 generations. The tracer package of BEAST was used to analyze the log collection with a burn-in value of 10 % of generations; TreeAnnotator package of BEAST was used to generate the maximum clade credibility (MCC) tree on a nexus file; and TempEst software v1.5 (Rambaut *et al.*, 2016) was used to create linear regression of root-to-tip distances over the year of sampling and to show the tree previously generated by TreeAnnotator.

Results

Main sequencing and genomic features

The mean fold coverage of the runs on Ion Torrent was 224 ± 88. The main features of the 39 genomes are organized on table 1. All isolates were comprised of one single chromosome with average size of 1,844,131 ± 4460 bp. The number of protein coding sequences (CDS) varied from 1503 to 1729, pseudogenes from 98 to 318 between the

strains and the average G+C content was 35.48%. A search on the genome was performed to assign the capsular serotype by an in-house script and all strains were typed as Ib.

MLST and eBURST profiles

The 39 strains presented two distinct sequence-types, one previously reported in Brazil (ST-260) plus a new ST never described before submitted to PubMLST and named ST-927 (table 1). Three strains (SA85, SA95 and SA97 (Godoy *et al.*, 2013)) previously typed as ST260 were reclassified a ST927. Also, several strains were non-typeable because of a partial deletion on *glcK* locus and the consequential hindrance of allelic typing, an event previously described by Assis *et al.* (2016). The eBURST analysis of all piscine-related STs is shown on figure 2. It grouped strains in two previously established CCs, CC10, with ST-283, ST-491 and ST-739, and CC7, with ST-6, ST-7, ST-500 and ST-735, and the new proposed clonal complexes CC260 and CC261, which have ST-260 and ST-261 as founders, respectively. Those two new CCs had close relationship with the group formed by non-typeable lineage and ST-257. Strain STIR-CD-17 (Genbank access number ALXB01) has no submitted ST, but its ST is derived from ST-260 only by a SNP on *tkt* allele. Neither ST-103 nor ST-258 shared enough alleles with any of the other piscine strains to be considered as genetically related.

Similarity between genomes

The all vs. all analysis of genomic similarity is presented on figure 3. All the Brazilian piscine strains showed high similarity (>98%), but two groups were slightly distinct, one composed of non-typeable isolates and other of ST-260 plus ST-927 strains, showing higher similarity inside the group than between groups.

Phylogenomic analysis based on wgMLST

The NeighborNet phylogenomic tree separated all the isolates on four main groups based on "All_loci" comparison (figure 4a). One group was the single strain 138spar (ST-261, USA), other group was strains STIR-CD-14 (ST-491, Vietnam) and GD201008-001 (ST-7, China). The Brazilian strains were divided in two groups (figure 4b), one group was formed by all non-typeable strains, all arising from Central-South region and the other major group was formed by ST-260 and ST-927 strains, arising mainly from Northeast but with representatives from Southeast. All the groups formed subgroups based on the divergence of alleles.

Bayesian analysis of evolution and emerging of Brazilian *S. agalactiae* clade

The substitution rate and evolution grouped strains in accordance with MLST (Figure 5). A Bayesian analysis was performed using BEAST software to investigate the temporal dynamics on the *S. agalactiae* species over substitution found in the core genome of *S. agalactiae*. The experiment was performed with 100 million generations and showed an Effective Sample Size (ESS) with values ≥ 200 for key parameters (output variables). The substitution rate on this population was 6.21×10^{-7}

substitution/site/year (95 % highest probability density (HPD), 3.42×10^{-7} to 8.97×10^{-7}). One clade formed by ST-261, which strains came from USA, Israel and China, plus ST-927, ST-260 and non-typeable strains of this work, all isolated from fish hosts on Latin America, was named Taxa V (see Material and Methods). This taxa seemed to have emerged at least 1233 years ago (95 % HPD, 738 to 1886), whereas Taxa IV, containing the reported STs of *S. agalactiae* from fish in Brazil, seemed to have emerged around 585 years ago (95 % HPD, 341 to 890). A third group (Taxa III) was composed only by ST-260 from Brazilian isolates, although this ST occurs outside Brazil, and had the time to most recent common ancestor (tMRCA) around 1983 (95 % HPD, 1963 to 1999). And, finally, two group exclusively from Brazilian samples (Taxa II – non typeable strains and Taxa I – ST-927) emerged, respectively, in 1956 (95 % HPD, 1923 to 1984) and 1978 (95 % HPD, 1955 to 1997).

Discussion

Nile tilapia is the main farmed fish in Brazil, but its production is highly affected by streptococcosis (Assis *et al.*, 2016). Besides having a high territorial extension, all Brazilian regions present at least one hot season, when water temperatures exceeds 27°C and outbreaks happen. This work presents isolates from GBS isolated from diseased fish arising of all Brazilian regions except North, where tilapia farming is absent because of the legal prohibition of the introduction of exotic species on the

Amazon basin. Independent on the region of origin, the fish presented the same common clinical signs of streptococcosis, exophthalmia, erratic swimming and corneal opacity (Mian *et al.*, 2009; Pereira *et al.*, 2010).

Through MLST the strains could be discriminated in three STs, being one previously unknown. This typing takes in consideration seven genes which play important functions for the cell (Jones *et al.*, 2003; Maiden, 2006) and whose mutations could possibly drive the cell to a purifying selection. The new clonal complex CC260 was proposed. It has ST-260 as founder and, comparing ST-927 to the founder ST, only the *glnA* allele has obtained a single nucleotide polymorphism (SNP). In relation to ST-260, this new allele changed on the position 474 of the 498 nt portion of the *glnA*, from guanine (G) to thymine (T), what changed the amino acid from glycine (G) to cysteine (C). This nonsynonymous mutation may have not compromised the glutamine synthesis by *glnA*, since the strains SA85, SA95, SA97, SA112 and SA218 are still viable. The isolate SA20 (Pereira *et al.*, 2013) previously typed as ST-553 (Godoy *et al.*, 2013) was not typeable on the present work along with other 24 isolates, which share a partial gene deletion event on *glcK*, previously described in SA20 (Assis *et al.*, 2016). These strains are more closely related to ST-552, as *glcK* partially deleted derives from allele number 26. A PCR reaction targeting the gene *glcK* was performed to all Brazilian strains which were described previously in the work of Godoy *et al.* (2013), and all of the PCRs were negative in electrophoresis on a 1.5% agarose gel (data not shown). These events resulted in the deconstruction of previously described CC552 (Godoy *et*

al., 2013), which used to comprise both ST-260 that, previously shared six alleles with ST-552 (now non-typeable) and ST-259, which shared six alleles with ST-260, and also that led to the construction of the new CC260, with ST-260 as founder and ST-259 and ST-927, and CC261, with ST-261 as founder and the other STs 246 and 891, being ST-261 and ST-246 still a double-locus variant from ST-552 (Godoy *et al.*, 2013).

MLST is the main epidemiological tool for human and animal diseases caused by *Streptococcus agalactiae*, and it has been widely used to obtain evolutionary relationships between strains from different epidemiological bias as hosts species, geographical distance, period of isolation (Evans *et al.*, 2008; Godoy *et al.*, 2013; Lusiastuti *et al.*, 2013; Bergal *et al.*, 2015; Chen *et al.*, 2015; Sun *et al.*, 2016), but it is known that because of using only seven genic loci, many genetic information are neglected and it compromises its resolution on the molecular characterization (Laing *et al.*, 2011), as total SNPs, lateral genetic transfer events or phage-related genes. In the present work, the three obtained STs showed a geographical discrimination, once the non-typeable strains occurred exclusively in Central-South region and the other two closely related STs occur mainly in Northeast, in accordance with a previous study (Godoy *et al.*, 2013). The occurrence of ST-260 and ST-927 (SA245 and SA218, respectively) in the Southeast region shows that these two types are circulating through the country and, as *S. agalactiae* is believed to be transmitted by contact between animals (Evans, 2002; Mian *et al.*, 2009). This circulation may be linked to the inter-regional transport of infected animals, mainly fingerlings produced in commercial

hatcheries. All strains in the present work were serotyped as Ib, and this corroborates the findings by Godoy *et al.* (2013), as this serotype was the predominant in Brazil.

In the analysis of genomic similarity, the average 1.8 mb genomes were fragmented in thousands of pieces of 200 bp and compared all against all. The results showed that each ST comprises strains with similar sequences (>99.8%), and differences between STs were little (<2%). It was previously shown through comparative genomics that *S. agalactiae* has a genome structured with a stable "back-bone" and other elements make difference between many lineages (Brochet *et al.*, 2006). So, little differences between genomes should happen due to polymorphisms on the gene sequences. Fish-associated GBS isolates from previously known CC552 are known to harbour some genetic content which are specific (Delannoy *et al.*, 2016), not shared with other CCs, which may be associated to other hosts. But there is no information about genomic diversity into a population of closely related *S. agalactiae* strains, which could be used to track every source of an outbreak of streptococcosis. The genomic features of the Brazilian isolates of *S. agalactiae* showed a number of pseudogenes varying from 98 to 320 between isolates, even inside each genomic lineage. It is known that piscine GBS is passing through a reductive evolution and it is commonly seen great percentage of pseudogenes, over 10% of the genomes, which is an adaptation to these hosts (Rosinski-Chupin *et al.*, 2013).

The analysis of genomic similarity corroborated with MLST results, but none was sufficient to discriminate the closely related strains. So, to understand the population

structure and map the Brazilian panorama of streptococcosis a whole-genome MLST was performed. The results strengthened the previous comparative analysis, since phylogenomic tree based on all loci comparison revealed that genomic-types are very nearer. The geographical discrimination of the isolates was in agreement with the MLST, but with wgMLST approach it was possible to see the relative genetic distances between each strain. According to Maiden *et al.* (2013) BIGSdb-wgMLST has potential to discriminate very closely related strains, or single clone pathogens. The genetic proximity inside a phylogenomic group seemed linked to the year of isolation. Inside the non-typeable group (figure 4), the lower axis concentred older isolates, and the upper axis, the newest ones, which showed that more contemporary strains had less allelic differences and the mutations which drives the allelic changing were happening through a temporal scale.

On the other phylogenomic group, although ST-260 and ST-927 isolates were neighbours, isolates from each lineage were closer to each other. The ST-260 was predominant on the state of Ceará on Northeast region, but SA245 was isolated from Minas Gerais state, on Southeast. The closer relationship between SA245 to SA256 and 289, both from Ceará state (northeast region), and the fact of all being isolated in 2013, suggest that strain SA245 came from Ceará to Minas Gerais near this period. Strain SA73, from Ceará state (supplementary table 1), was the only one in this study isolated from a diseases fingerling, and it is of concern on epidemiology of streptococcosis, once that fingerlings are commonly transported from hatcheries to

many grow out farms, even interregionally. On the ST-927 group, the two strains from state of Pernambuco, (SA102 and SA97) were closer to each other, and the same occurred to strains SA85 and SA95, from Alagoas state. Moreover, the strain SA218 from Southeast was very close to SA97 (both ST-927), from Northeast, but there were two years of difference on their isolations, what suggests that both diverged from a common ancestor which has not mutated so much since its transportation event. The circulation of different lineages from Northeast to Southeast may be a concern, as there is no information on different pathogenicity or clinical presentation between these STs. The hybrid catfish isolate SA81 was very close to some tilapia strains inside the non-typeable group. The little allelic differences between these strains suggest that, if there is a genomic pattern on the strains from each species of host, it could be linked to other features as, in example, specific genes or pseudogenes, but it would be necessary more closely related genomes of GBS isolated from different fish species to assay host-specific genomic features. For the moment, it is known that strains arising from phylogenetically distant hosts but from the same CCs of fish isolates could experimentally infect fish (Chen *et al.*, 2015), but there is no evidence for natural disruption of host barrier by *S. agalactiae*.

Interestingly, besides this huge discrimination by wgMLST, all the piscine strains from Latin America have emerged from a single branch on the Bayesian evolutionary tree (figure 5), and are grouped on the taxa IV in accordance with MLST, into ST-260, ST-927 and non-typeable taxa. This group, which includes all the Brazilian strains of the

present study, had an emergence of about 585 years, which is surprisingly more recent than the group V, which includes the other American, Chinese and Israelite piscine strains from ST-261 and had an emergence around 1234 years ago. It´s possible to see that since the divergence from ST-261 and group IV, this last group, which comprises the previously known CC552, is evolving more rapidly and diversifying, what led to the intra-groups I, II and III. These results predicated that, in Brazil, *S. agalactiae* is passing through a recent process of adaptation to the piscine hosts and it could be linked to the increase of Brazilian aquaculture and a bigger spreading and pressure of selection of this pathogenic species. It is known that in the past two decades it has been introduced four lineages of Nile tilapia coming from different World origins, and maybe they could have different host-pathogen interactions and represent subpopulations of *S. agalactiae*.

According to the Bayesian analysis of evolution (figure 5), inside the group I formed by non-typeable strains, SA81 (marked with a star) emerged in a different branch from the tilapia strains, which suggests that there is a microevolution inside this group which is leading to different host tropisms. In the evolutionary analysis (figure 5), the human ST-7 strain A909 isolated in 1975 (GenBank access number CP000114, marked with a star) emerged previously on a single branch in contrast to all the other six ST-7 strains isolated between 2009-2015 (supplementary table 2), which were from fish origin and emerged more contemporaneously to the Brazilian strains in 1981 (1964-1996) and share a single branch. It showed that since ST-7 emergence, there has been

a diversification to different host tropisms, and besides previous genomic studies revealed a close genomic relationship between these isolates (Liu *et al.*, 2013; Rosinski-Chupin, 2013), the present work also highlights a common ancestor for some human and piscine strains.

Interestingly, there were no clones in this population, even in farms in the same city or region. WgMLST results showed that piscine GBS population of Brazil are relatively heterogeneous on genomic term, the closest strains (SA623 and SA627) diverged by 22 alleles, so each isolate is a single clone. All the genomes in question are accumulating mutations, which, in different alleles, make up the bigger genomic distance visible by wgMLST phylogenomics. The evolutionary tree of *S. agalactiae* hinted that along the expansion of Brazilian aquaculture in the past few decades, the population of *S. agalactiae* evolved rising up the new taxa II (figure 5) and also diversifying into the different taxa.

The study revealed that Brazilian *S. agalactiae* population of piscine isolates is diverse and spatially distributed in accord to the sequence type, and many evolutionary events are recent leading to new groups. The whole-genome MLST showed that Brazilian fish lineages are evolving quickly in each outbreak so much that there are no clones, but only single clone strains on this population, and the evolutionary study revealed that there are contemporary events which are leading to a bigger diversity inside the preponderant group of piscine *S. agalactiae* in Brazil. The gene-by-gene approach on

wgMLST is a powerful tool to track outbreaks of streptococcosis by *S. agalactiae* farm-to-farm, and to establish spatial links between disease events.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank the Coordination for the Improvement of Higher Education Personnel (CAPES) and National Counsel of Technological and Scientific Development (CNPq) for the financing of this study.

References

AGREN, J.; SUNDSTROM, A.; HAFSTROM, T. *et al.* Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. *PLoS One*, vol.7, no.Number, p.e39107, 2012.

ALMEIDA, A.; ALVES-BARROCO, C.; SAUVAGE, E. *et al.* Persistence of a dominant bovine lineage of group B Streptococcus reveals genomic signatures of host adaptation. *Environ Microbiol*, vol.18, no.Number, p.4216-4229, 2016.

ALMEIDA, A.; VILLAIN, A.; JOUBREL, C. *et al.* Whole-Genome Comparison Uncovers Genomic Mutations between Group B Streptococci Sampled from Infected Newborns and Their Mothers. *J Bacteriol*, vol.197, no.Number, p.3354-3366, 2015.

AMAL, M.N.; ZAMRI-SAAD, M.; IFTIKHAR, A.R. *et al.* An outbreak of Streptococcus agalactiae infection in cage-cultured golden pompano, Trachinotus blochii (Lacepede), in Malaysia. *J Fish Dis*, vol.35, no.Number, p.849-852, 2012.

ANDREWS, S. - - FastQC A quality control application for high throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/download.html#fastqc. Accessed in: 01/11.

ASSIS, G.B.; TAVARES, G.C.; PEREIRA, F.L. *et al.* Natural coinfection by Streptococcus agalactiae and Francisella noatunensis subsp. orientalis in farmed Nile tilapia (Oreochromis niloticus L.). *J Fish Dis*, no.Number, 2016.

BELÉN, A.; PAVÓN, I.; MAIDEN, M.C.J. Multilocus Sequence Typing. *Methods in molecular biology (Clifton, N.J.)*, vol.551, no.Number, p.129-140, 2009.

BERGAL, A.; LOUCIF, L.; BENOUARETH, D.E. *et al.* Molecular epidemiology and distribution of serotypes, genotypes, and antibiotic resistance genes of Streptococcus agalactiae clinical isolates from Guelma, Algeria and Marseille, France. *Eur J Clin Microbiol Infect Dis*, vol.34, no.Number, p.2339-2348, 2015.

BERGH, K.; STOELHAUG, A.; LOESETH, K. *et al.* Detection of group B streptococci (GBS) in vaginal swabs using real-time PCR with TaqMan probe hybridization. *Indian J Med Res*, vol.119 Suppl, no.Number, p.221-223, 2004.

BERRIDGE, B.R.; BERCOVIER, H.; FRELIER, P.F. Streptococcus agalactiae and Streptococcus difficile 16S-23S intergenic rDNA: genetic homogeneity and species-specific PCR. *Vet Microbiol*, vol.78, no.Number, p.165-173, 2001.

BOHNSACK, J.F.; WHITING, A.; GOTTSCHALK, M. *et al.* Population structure of invasive and colonizing strains of Streptococcus agalactiae from neonates of six U.S. Academic Centers from 1995 to 1999. *J Clin Microbiol*, vol.46, no.Number, p.1285-1291, 2008.

BOWATER, R.O.; FORBES-FAULKNER, J.; ANDERSON, I.G. *et al.* Natural outbreak of Streptococcus agalactiae (GBS) infection in wild giant Queensland grouper, Epinephelus lanceolatus (Bloch), and other wild fish in northern Queensland, Australia. *J Fish Dis*, vol.35, no.Number, p.173-186, 2012.

BRAZIL - Ministério da Pesca e Aquicultura. BOLETIM ESTATÍSTICO DA PESCA E AQUICULTURA 2011. 2013. <http://www.icmbio.gov.br/cepsul/images/stories/biblioteca/download/estatistica/est_2011_bol__bra.pdf>. Accessed in: 01/20/2017.

BROCHET, M.; COUVE, E.; ZOUINE, M. *et al.* Genomic diversity and evolution within the species Streptococcus agalactiae. *Microbes Infect*, vol.8, no.Number, p.1227-1243, 2006.

CARVALHO-CASTRO, G.A.; LOPES, C.O.; LEAL, C.A. *et al.* Detection of type III secretion system genes in Aeromonas hydrophila and their relationship with virulence in Nile tilapia. *Vet Microbiol*, vol.144, no.Number, p.371-376, 2010.

CHEN, M.; WANG, R.; LUO, F.G. *et al.* Streptococcus agalactiae isolates of serotypes Ia, III and V from human and cow are able to infect tilapia. *Vet Microbiol*, vol.180, no.Number, p.129-135, 2015.

CODY, A.J.; BENNETT, J.S.; MAIDEN, M.C.J. Multi-Locus Sequence Typing and the Gene-by-Gene Approach to Bacterial Classification and Analysis of Population Variation. vol.41, no.Number, p.201-219, 2014.

CREPALDI, D.V., FARIA, P.M.C., TEIXEIRA, E.A., RIBEIRO, L.P., COSTA, A.A.P., MELO, D.C., CINTRA, A.P., PRADO, S.A., COSTA, F.A.A., DRUMOND, M.L., LOPES, V.E., MORAES, V.E. A situação da Aquacultura no Brasil e no Mundo. *Revista Brasileira Reprodução Animal*, vol.30, no.Number, p.81-85, 2006.

CROUCHER, N.J.; HARRIS, S.R.; GRAD, Y.H. *et al.* Bacterial genomes in epidemiology--present and future. *Philos Trans R Soc Lond B Biol Sci*, vol.368, no.Number, p.20120202, 2013.

DELANNOY, C.M.; ZADOKS, R.N.; CRUMLISH, M. *et al.* Genomic comparison of virulent and non-virulent Streptococcus agalactiae in fish. *J Fish Dis*, vol.39, no.Number, p.13-29, 2016.

DELANNOY, C.M.J., CRUMLISH, M., FONTAINE, M. C., POLLOCK, J., FOSTER, G., DAGLEISH, M. P., TURNBULL, J. F., ZADOKS, R. N. Human Streptococcus agalactiae strains in aquatic mammals and fish. *Bmc Microbiology*, vol.13, no.Number, 2013.

DERMER, P.; LEE, C.; EGGERT, J. *et al.* A history of neonatal group B streptococcus with its related morbidity and mortality rates in the United States. *J Pediatr Nurs*, vol.19, no.Number, p.357-363, 2004.

DOW, S.W.; JONES, R.L.; THOMAS, T.N. *et al.* Group B streptococcal infection in two cats. *J Am Vet Med Assoc*, vol.190, no.Number, p.71-72, 1987.
DRUMMOND, A.J.; RAMBAUT, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, vol.7, no.Number, p.214, 2007.
DUREMDEZ, R.; AL-MARZOUK, A.; QASEM, J.A. *et al.* Isolation of Streptococcus agalactiae from cultured silver pomfret, Pampus argenteus (Euphrasen), in Kuwait. *J Fish Dis*, vol.27, no.Number, p.307-310, 2004.
EDELSTEIN, R.M.; PEGRAM, R.G. Contagious skin necrosis of Somali camels associated with Streptococcus agalactiae. *Trop Anim Health Prod*, vol.6, no.Number, p.255-256, 1974.
EDWARDS, M.S.; BAKER, C.J. Group B streptococcal infections in elderly adults. *Clin Infect Dis*, vol.41, no.Number, p.839-847, 2005.
ELDAR, A.; BEJERANO, Y.; LIVOFF, A. *et al.* Experimental streptococcal meningo-encephalitis in cultured fish. *Veterinary Microbiology*, vol.43, no.Number, 1995.
ELDAR, A., BEJERANO, Y., LIVOFF, A., HOROVITCZ, A., BERCOVIER, H. Experimental streptococcal meningo-encephalitis in cultured fish. *Vet Microbiol*, vol.43, no.Number, p.33-40, 1995
ELLIOTT, J.A.; FACKLAM, R.R.; RICHTER, C.B. Whole-cell protein patterns of nonhemolytic group B, type Ib, streptococci isolated from humans, mice, cattle, frogs, and fish. *J Clin Microbiol*, vol.28, no.Number, p.628-630, 1990.
EVANS, J.J.; BOHNSACK, J.F.; KLESIUS, P.H. *et al.* 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 is also associated with human neonatal infections in Japan. *Journal of Medical Microbiology*, vol.57, no.Number, 2008.
EVANS, J.J.; BOHNSACK, J.F.; KLESIUS, P.H. *et al.* 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 is also associated with human neonatal infections in Japan. *J Med Microbiol*, vol.57, no.Number, p.1369-1376, 2008.
EVANS, J.J.; KLESIUS, P.H.; GILBERT, P.M. *et al.* Characterization of β-haemolytic Group B Streptococcus agalactiae in cultured seabream, Sparus auratus L., and wild mullet, Liza klunzingeri (Day), in Kuwait. *J Fish Dis*, vol.25, no.Number, p.505-513, 2002.
EVANS, J.J., KLESIUS, P.H., GILBERT, P.M., SHOEMAKER, C.A., SARAWI, M.A.AL., LANDSBERG, J., DUREMDEZ, R., MARZOUK, A.AL.., ZENKI, S.AL. Characterization of β-haemolytic Group B Streptococcus agalactiae in cultured seabream, Sparus auratus L., and wild mullet, Liza klunzingeri (Day), in Kuwait. *J Fish Dis*, vol.25, no.Number, p.505-513, 2002.
FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. Cultured Aquatic Species Information Programme - Oreochromis niloticus (Linnaeus, 1758). 2017. http://www.fao.org/fishery/culturedspecies/Oreochromis_niloticus/en#tcNA00EA. Accessed in: 01/10/2017.
FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. The State of World Fisheries and Aquaculture 2014. 2014. <http://www.fao.org/3/a-i3720e.pdf>. Accessed in: 01/10/2017.
FAO - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. The State of World Fisheries and Aquaculture 2016. 2016. <http://www.fao.org/3/a-i5555e.pdf>. Accessed in: 01/10/2017.

FEIL, E.J.; LI, B.C.; AANENSEN, D.M. *et al.* eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data. *Journal of Bacteriology*, vol.186, no.Number, p.1518-1530, 2004.
FIGUEIREDO, H.C.; KLESIUS, P.H.; ARIAS, C.R. *et al.* Isolation and characterization of strains of Flavobacterium columnare from Brazil. *J Fish Dis*, vol.28, no.Number, p.199-204, 2005.
FIGUEIREDO, H.C.P.; SOARES, S.C.; PEREIRA, F.L. *et al.* Comparative genome analysis of Weissella ceti, an emerging pathogen of farm-raised rainbow trout. *Bmc Genomics*, vol.16, no.Number, 2015.
GALARDINI, M., BIONDI, E.G., BAZZICALUPO, M., MENGONI, A. CONTIGuator a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code for Biology and Medicine*, vol.6, no.Number, 2011.
GODOY, D.T.; CARVALHO-CASTRO, G.A.; LEAL, C.A. *et al.* Genetic diversity and new genotyping scheme for fish pathogenic Streptococcus agalactiae. *Lett Appl Microbiol*, vol.57, no.Number, p.476-483, 2013.
HERNANDEZ, E.; FIGUEROA, J.; IREGUI, C. Streptococcosis on a red tilapia, Oreochromis sp., farm: a case study. *J Fish Dis*, vol.32, no.Number, p.247-252, 2009.
HILLERTON, J.E.; BERRY, E.A. Treating mastitis in the cow--a tradition or an archaism. *J Appl Microbiol*, vol.98, no.Number, p.1250-1255, 2005.
HUSON, D.H.; BRYANT, D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*, vol.23, no.Number, p.254-267, 2006.
IBGE - Instituto Brasileiro de Geografia e Estatística. Brazil. Produção da Pecuária Municipal 2015. 2015. <http://www.ibge.gov.br/home/estatistica/economia/ppm/2015/default.shtm>. Accessed in: 12/12/2016.
ISMAIL, A.Q.; ANTHONY, M. Cross-species transfer of group B streptococcus via ingestion? *J Perinat Med*, vol.40, no.Number, p.201-202, 2012.
JOHRI, A.K.; PAOLETTI, L.C.; GLASER, P. *et al.* Group B Streptococcus: global incidence and vaccine development. *Nat Rev Microbiol*, vol.4, no.Number, p.932-942, 2006.
JOLLEY, K.A.; MAIDEN, M.C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, vol.11, no.Number, p.595, 2010.
JONES, N.; BOHNSACK, J.F.; TAKAHASHI, S. *et al.* Multilocus Sequence Typing System for Group B Streptococcus. *J Clin Microbiol*, vol.41, no.Number, p.2530-2536, 2003.
JORGENSEN, H.J.; NORDSTOGA, A.B.; SVILAND, S. *et al.* Streptococcus agalactiae in the environment of bovine dairy herds--rewriting the textbooks? *Vet Microbiol*, vol.184, no.Number, p.64-72, 2016.
KATO, K. Impact of the next generation DNA sequencers. *International Journal of Clinical and Experimental Medicine*, vol.2, no.Number, p.193-202, 2009.
KATOH, K.; STANDLEY, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, vol.30, no.Number, p.772-780, 2013.
KEEFE, G.P. Streptococcus agalactiae mastitis: a review. *The Canadian Veterinary Journal*, vol.38, no.Number, p.429-437, 1997.
LAGO, A.; GODDEN, S.M.; BEY, R. *et al.* The selective treatment of clinical mastitis based on on-farm culture results: I. Effects on antibiotic use, milk withholding time, and short-term clinical and bacteriological outcomes. *Journal of Dairy Science*, vol.94, no.Number, p.4441-4456, 2011.
LAGO, A.; GODDEN, S.M.; BEY, R. *et al.* The selective treatment of clinical mastitis based on on-farm culture results: II. Effects on lactation performance, including clinical mastitis recurrence,

somatic cell count, milk production, and cow survival. *J Dairy Sci*, vol.94, no.Number, p.4457-4467, 2011.
LAING, C.R.; ZHANG, Y.; THOMAS, J.E. *et al.* Everything at once: comparative analysis of the genomes of bacterial pathogens. *Veterinary Microbiology*, vol.153, no.Number, p.13-26, 2011.
LEAL, C.A.; TAVARES, G.C.; FIGUEIREDO, H.C. Outbreaks and genetic diversity of Francisella noatunensis subsp orientalis isolated from farm-raised Nile tilapia (Oreochromis niloticus) in Brazil. *Genet Mol Res*, vol.13, no.Number, p.5704-5712, 2014.
LIU, G.; ZHANG, W.; LU, C. Comparative genomics analysis of Streptococcus agalactiae reveals that isolates from cultured tilapia in China are closely related to the human strain A909. *BMC Genomics*, vol.14, no.Number, p.775, 2013.
LIU, G.; ZHANG, W.; LU, C. Comparative genomics analysis of Streptococcus agalactiae reveals that isolates from cultured tilapia in China are closely related to the human strain A909. *BMC Genomics*, vol.14, no.Number, 2013.
LOPES JÚNIOR, J.E.F.; LANGE, C.C.; BRITO, M.A.V.P. *et al.* Relationship between total bacteria counts and somatic cell counts from mammary quarters infected by mastitis pathogens. *Ciência Rural*, vol.42, no.Number, p.691-696, 2012.
LUSIASTUTI, A.M.; SEEGER, H.; INDRAWATI, A. *et al.* The Comparison of Streptococcus agalactiae Isolated from Fish and Bovine using Multilocus Sequence Typing. *HAYATI Journal of Biosciences*, vol.20, no.Number, p.157-162, 2013.
MAIDEN, M.C. Multilocus sequence typing of bacteria. *Annu Rev Microbiol*, vol.60, no.Number, p.561-588, 2006.
MAIDEN, M.C.; JANSEN VAN RENSBURG, M.J.; BRAY, J.E. *et al.* MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol*, vol.11, no.Number, p.728-736, 2013.
MAIDEN, M.C.J. Multilocus sequence typing of bacteria. *Annu Rev Microbiol*, vol.60, no.Number, p.561-588, 2006.
MAIDEN, M.C.J.; BYGRAVES, J.A.; FEIL, E. *et al.* Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A*, vol.95, no.Number, p.3140-3145, 1998.
MAIONE, D.; MARGARIT, I.; RINAUDO, C.D. *et al.* Identification of a universal Group B streptococcus vaccine by multiple genome screen. *Science*, vol.309, no.Number, p.148-150, 2005.
MATA, A.I.; GIBELLO, A.; CASAMAYOR, A. *et al.* Multiplex PCR assay for detection of bacterial pathogens associated with warm-water Streptococcosis in fish. *Appl Environ Microbiol*, vol.70, no.Number, p.3183-3187, 2004.
MCDONALD, J.S. Streptococcal and staphylococcal mastitis. *J Am Vet Med Assoc*, vol.170, no.Number, p.1157-1159, 1977.
MIAN, G.F.; GODOY, D.T.; LEAL, C.A. *et al.* Aspects of the natural history and virulence of S. agalactiae infection in Nile tilapia. *Vet Microbiol*, vol.136, no.Number, p.180-183, 2009.
MWEU, M.M.; TOFT, N.; KATHOLM, J. *et al.* Evaluation of two herd-level diagnostic tests for Streptococcus agalactiae using a latent class approach. *Vet Microbiol*, vol.159, no.Number, p.181-186, 2012.
NETTO, L.N.; LEAL, C.A.; FIGUEIREDO, H.C. Streptococcus dysgalactiae as an agent of septicaemia in Nile tilapia, Oreochromis niloticus (L.). *J Fish Dis*, vol.34, no.Number, p.251-254, 2011.

ONMUS-LEONE, F., HANG, J., CLIFFORD, R.J., YANG, Y., RILEY, M.C., KUSCHNER, R.A., WATERMAN, P.E., LESHO, E.P. Enhanced De Novo Assembly of High Throughput Pyrosequencing Data Using Whole Genome Mapping. *Plos One*, vol.8, no.Number, 2013.
ORTEGA ASENCIOS, Y.; BARREIRO SÁNCHEZ, F.; BUENO MENDIZÁBAL, H. *et al.* First report of Streptococcus agalactiae isolated from Oreochromis niloticus in Piura, Peru: Molecular identification and histopathological lesions. *Aquaculture Reports*, vol.4, no.Number, p.74-79, 2016.
PALLEN, M.J.; WREN, B.W. Bacterial pathogenomics. *Nature*, vol.449, no.Number, p.835-842, 2007.
PEREIRA, F.L.; SOARES, S.C.; DORELLA, F.A. *et al.* Evaluating the efficacy of the new Ion PGM Hi-Q Sequencing Kit applied to bacterial genomes. *Genomics*, vol.107, no.Number, p.189-198, 2016.
PEREIRA, U.D.P.; RODRIGUES DOS SANTOS, A.; HASSAN, S.S. *et al.* Complete genome sequence of Streptococcus agalactiae strain SA20-06, a fish pathogen associated to meningoencephalitis outbreaks. *Standards in Genomic Sciences*, vol.8, no.Number, p.188-197, 2013.
PEREIRA, U.P.; MIAN, G.F.; OLIVEIRA, I.C. *et al.* Genotyping of Streptococcus agalactiae strains isolated from fish, human and cattle and their virulence potential in Nile tilapia. *Vet Microbiol*, vol.140, no.Number, p.186-192, 2010.
PEREIRA, U.P.; SOARES, S.C.; BLOM, J. *et al.* In silico prediction of conserved vaccine targets in Streptococcus agalactiae strains isolated from fish, cattle, and human samples. *Genet Mol Res*, vol.12, no.Number, p.2902-2912, 2013.
PETZER, I.-M.; KARZIS, J.; LESOSKY, M. *et al.* Host adapted intramammary infections in pregnant heifers which were co-housed and reared on fresh milk as calves. *BMC Veterinary Research*, vol.9, no.Number, p.49, 2013.
PLUMB, J.A.; SCHACHTE, J.H.; GAINES, J.L. *et al.* Streptococcus sp. from Marine Fishes Along the Alabama and Northwest Florida Coast of the Gulf of Mexico. *Transactions of the American Fisheries Society*, vol.103, no.Number, p.358-361, 1974.
POURGHOLAM, R., LALUEI, F., SAEEDI, A.A., ZAHEDI, R., SAFARI, R., TAGHAVI, M.J., SARAVI, H.N., POURGHOLAM, H. Distribution and Molecular identification of some causative agents of streptococcosis isolated from farmed rainbow trout (*Oncorhynchus mykiss,* Walbaum) in Iran. *Iranian Journal of Fisheries Sciences*, vol.10, no.Number, p.109-122, 2011.
PUYMEGE, A.; BERTIN, S.; GUEDON, G. *et al.* Analysis of Streptococcus agalactiae pan-genome for prevalence, diversity and functionality of integrative and conjugative or mobilizable elements integrated in the tRNA(Lys CTT) gene. *Mol Genet Genomics*, vol.290, no.Number, p.1727-1740, 2015.
RAJAGOPAL, L. Understanding the regulation of Group B Streptococcal virulence factors. *Future Microbiol*, vol.4, no.Number, p.201-221, 2009.
RAMBAUT, A.; LAM, T.T.; MAX CARVALHO, L. *et al.* Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution*, vol.2, no.Number, 2016.
RICHARDS, V.P.; LANG, P.; BITAR, P.D. *et al.* Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted Streptococcus agalactiae. *Infect Genet Evol*, vol.11, no.Number, p.1263-1275, 2011.
ROBINSON, J.A.; MEYER, F.P. Streptococcal fish pathogen. *J Bacteriol*, vol.92, no.Number, p.512, 1966.

ROSINSKI-CHUPIN, I.; SAUVAGE, E.; MAIREY, B. *et al.* Reductive evolution in Streptococcus agalactiaeand the emergence of a host adapted lineage. *Bmc Genomics*, vol.14, no.Number, p.1-15, 2013.
ROSINSKI-CHUPIN, I., SAUVAGE, E., MAIREY, B., MANGENOT, S., MA, L., DA CUNHA, V., RUSNIOK, C., BOUCHIER, C., BARBE, V., GLASER, P. Reductive evolution in Streptococcus agalactiae and the emergence of a host adapted lineage. *BMC Genomics*, vol.14, no.Number, 2013.
RUANE, N.M.; COLLINS, E.M.; GEARY, M. *et al.* Isolation of Streptococcus agalactiae and an aquatic birnavirus from doctor fish Garra rufa L. *Irish Veterinary Journal*, vol.66, no.Number, p.16-16, 2013.
RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M.A., BARRELL, B. Artemis: sequence visualization and annotation. *Bioinformatics*, vol.16, no.Number, p.944-945, 2000.
SCHUCHAT, A. Epidemiology of group B streptococcal disease in the United States: shifting paradigms. *Clin Microbiol Rev*, vol.11, no.Number, p.497-513, 1998.
SEBASTIÃO, F.A.; NOMURA, D.; SAKABE, R. *et al.* Hematology and Productive Performance of Nile Tilapia (Oreochromis niloticus) Naturally Infected With Flavobacterium columnare. *Brazilian Journal of Microbiology*, vol.42, no.Number, p.282-289, 2011.
SEEMANN, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, vol.30, no.Number, p.2068-2069, 2014.
SHEPPARD, A.E.; VAUGHAN, A.; JONES, N. *et al.* Capsular Typing Method for Streptococcus agalactiae Using Whole-Genome Sequence Data. *J Clin Microbiol*, vol.54, no.Number, p.1388-1390, 2016.
SOARES, S.C.; GEYIK, H.; RAMOS, R.T. *et al.* GIPSy: Genomic island prediction software. *Journal of Biotechnology*, no.Number, 2015.
SORENSEN, U.B.; POULSEN, K.; GHEZZO, C. *et al.* Emergence and global dissemination of host-specific Streptococcus agalactiae clones. *MBio*, vol.1, no.Number, 2010.
SOTO, E.; WANG, R.; WILES, J. *et al.* Characterization of Isolates of Streptococcus agalactiae from Diseased Farmed and Wild Marine Fish from the U.S. Gulf Coast, Latin America, and Thailand. *J Aquat Anim Health*, vol.27, no.Number, p.123-134, 2015.
SPRINGMAN, A.C.; LACHER, D.W.; WU, G. *et al.* Selection, recombination, and virulence gene diversity among group B streptococcal genotypes. *J Bacteriol*, vol.191, no.Number, p.5419-5427, 2009.
SU, Y.L.; FENG, J.; LI, Y.W. *et al.* Development of a quantitative PCR assay for monitoring Streptococcus agalactiae colonization and tissue tropism in experimentally infected tilapia. *J Fish Dis*, vol.39, no.Number, p.229-238, 2016.
SUN, J.; FANG, W.; KE, B. *et al.* Inapparent Streptococcus agalactiae infection in adult/commercial tilapia. *Sci Rep*, vol.6, no.Number, p.26319, 2016.
TETTELIN, H.; MASIGNANI, V.; CIESLEWICZ, M.J. *et al.* Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A*, vol.102, no.Number, p.13950-13955, 2005.
VANDAMME, P.; DEVRIESE, L.A.; POT, B. *et al.* Streptococcus difficile is a nonhemolytic group B, type Ib Streptococcus. *Int J Syst Bacteriol*, vol.47, no.Number, p.81-85, 1997.
WANG, R.; LI, L.; HUANG, Y. *et al.* Comparative genome analysis identifies two large deletions in the genome of highly-passaged attenuated Streptococcus agalactiae strain YM001

compared to the parental pathogenic strain HN016. *BMC Genomics*, vol.16, no.Number, p.897, 2015.
YAO, K.; POULSEN, K.; MAIONE, D. *et al.* Capsular gene typing of Streptococcus agalactiae compared to serotyping by latex agglutination. *J Clin Microbiol*, vol.51, no.Number, p.503-507, 2013.
YE, X.; LI, J.; LU, M. *et al.* Identification and molecular typing of Streptococcus agalactiae isolated from pond-cultured tilapia in China. *Fisheries Science*, vol.77, no.Number, p.623-632, 2011.

Uncategorized References

TABLES

Table 1. Main genomic features and sequence-type of Brazilian GBS isolates

Supplementary Table 1. Geographical origin of the 39 strains of *S. agalactiae* isolated from outbreaks of streptococcosis in Brazil

Supplementary Table 2. Isolation dates of the strains from the Bayesian evolutionary analysis.

FIGURES

Figure 1. Map of the distribution of the *Streptococcus agalactiae* sequence-types through Brazilian territory.

Figure 2. Reconstruction of evolutionary relationships between piscine *Streptococcus agalactiae* through eBURST analysis from MLST. Points represent a ST, circles represent clonal complexes or closely related STs, blue lines represent double-locus variants and isolated points represent singletons. Clonal complexes arising exclusively from piscine strains are CC260, CC261 and the CC formed by ST-257 and non-typeable strains. CC10 and CC7 arise from piscine and human strains. Dashed area encompasses STs the previously known CC552 with the new ST-927.

Figure 3. All vs. all similarity between genomes of study, from red to green it increases similarity and dashed-delimited areas harbor intra-ST similarities.

Figure 4. Phylogenomic NeighborNet tree of whole-genome MLST data. Scale bar measures 100 alleles of difference. (a) The four major phylogenomic groups, with a focus on the brazilian strains. (b) Ampliation of the brazilian phylogenomic splits. The left side of the tree harbors all ST-927 and ST-260 strains, all from grow out farms, with the exception of the marked SA73, from a hatchery. All the strains from those two STs arore from Northeast region, with the exception of the marked isolates SA218 and SA245, both emerged from Southeast region. The right side of the tree harbors all non-typeable strains, all arising from Central-South macroregion. The marked strain SA81 was obtained from a diseased catfish, while all the remaining strains were obtained from Nile tilapia.

Figure 5. Bayesian evolutionary analysis of *Streptococcus agalactiae*. The upper branch of the tree give origin to the taxa IV, which contains Latin-American piscine isolates of *S. agalactiae* and taxa V, which is taxa IV in addition to ST-261. The marked strain SA81 is an exclusive Brazilian catfish isolate, while all the other Brazilian isoltes are from tilapia source. The time of most recent common ancestor (tMRCA) of each taxa of interest is provided and marked with a yellow line. On the lower region of the tree, inside ST-7, the marked strain A909 was isolated from human, while the other ST-7 isolates were from fish, and they have different branches of origin. Fish ST-7 isolates have a tMRCA of 35 years.

Figure 1

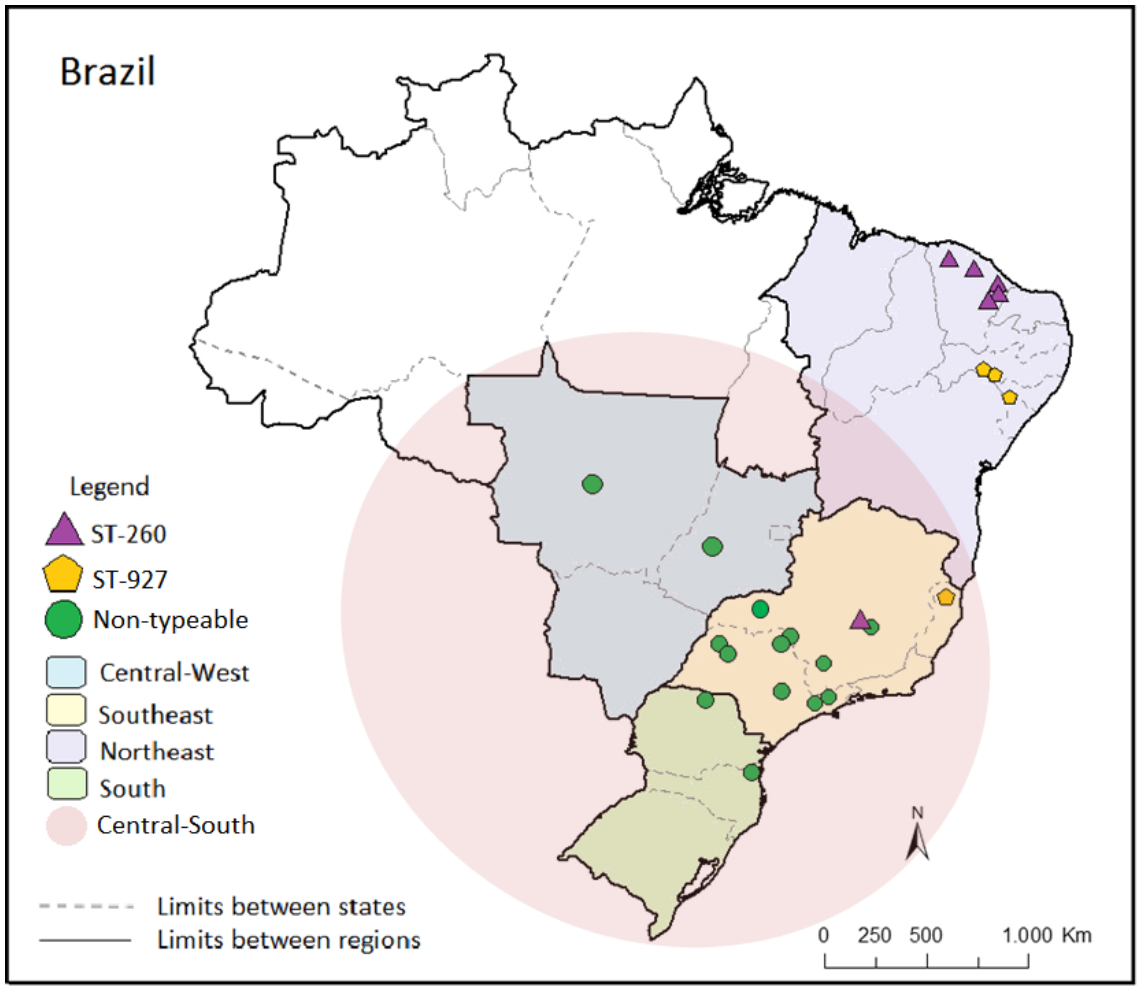
Brazil
Legend
ST-260
ST-927
Non-typeable
Central-West
Southeast
Northeast
South
Central-South
Limits between states
Limits between regions
N
0 250 500 1.000 Km

Figure 2

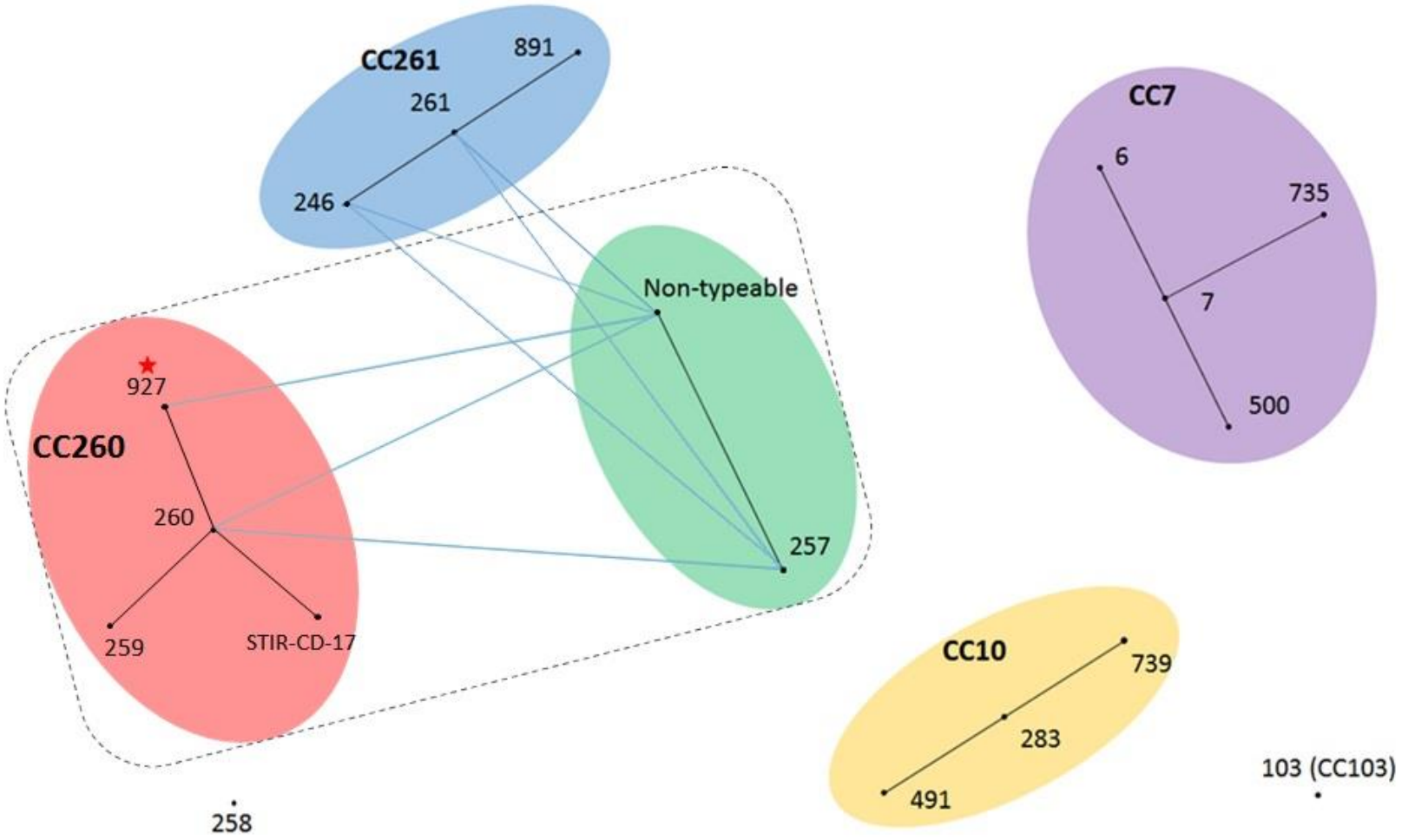
CC261
891
261
246
Non-typeable
CC7
6
735
7
500
927
CC260
260
257
259
STIR-CD-17
CC10
739
283
491
103 (CC103)
258

Figure 3

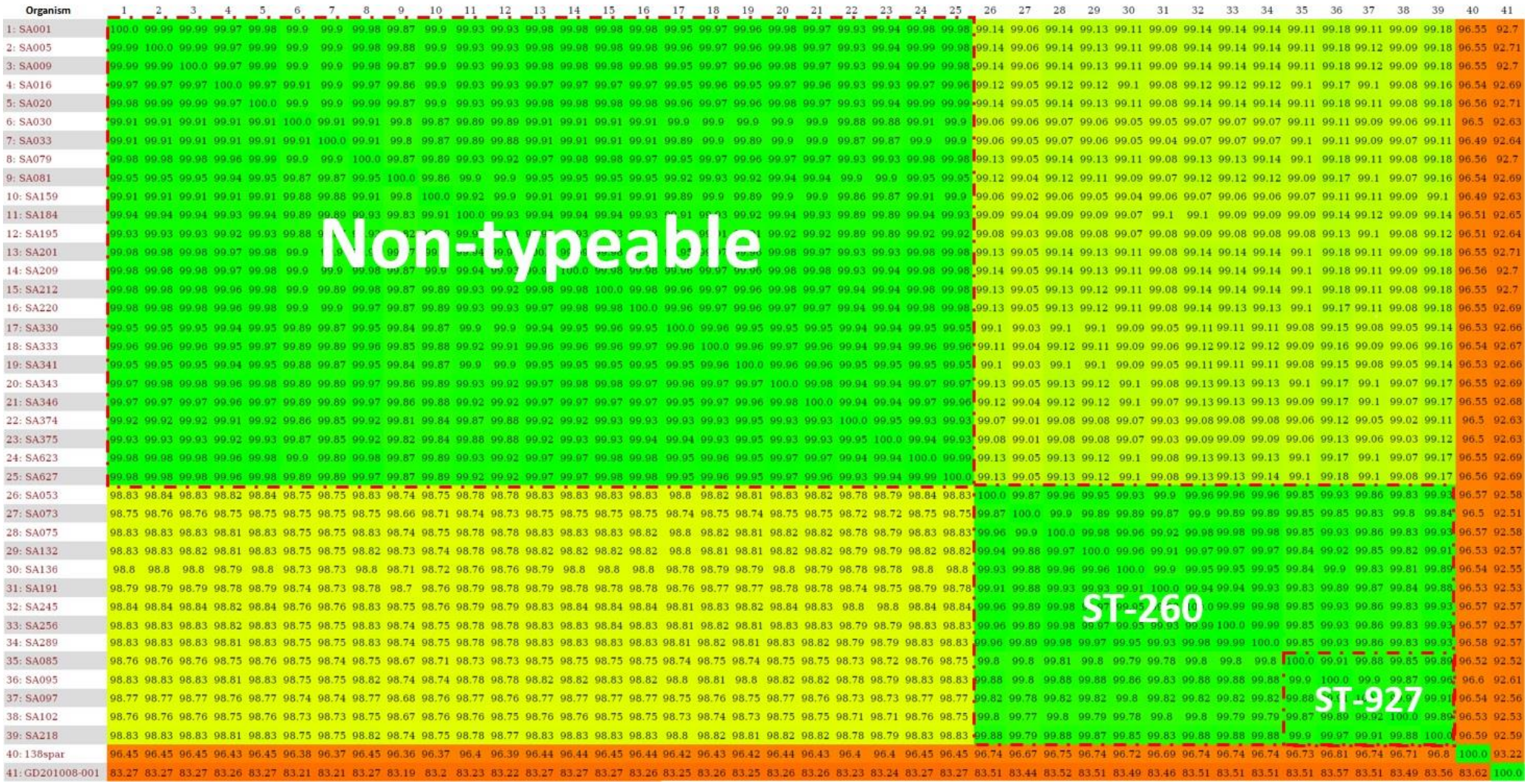

Figure 4

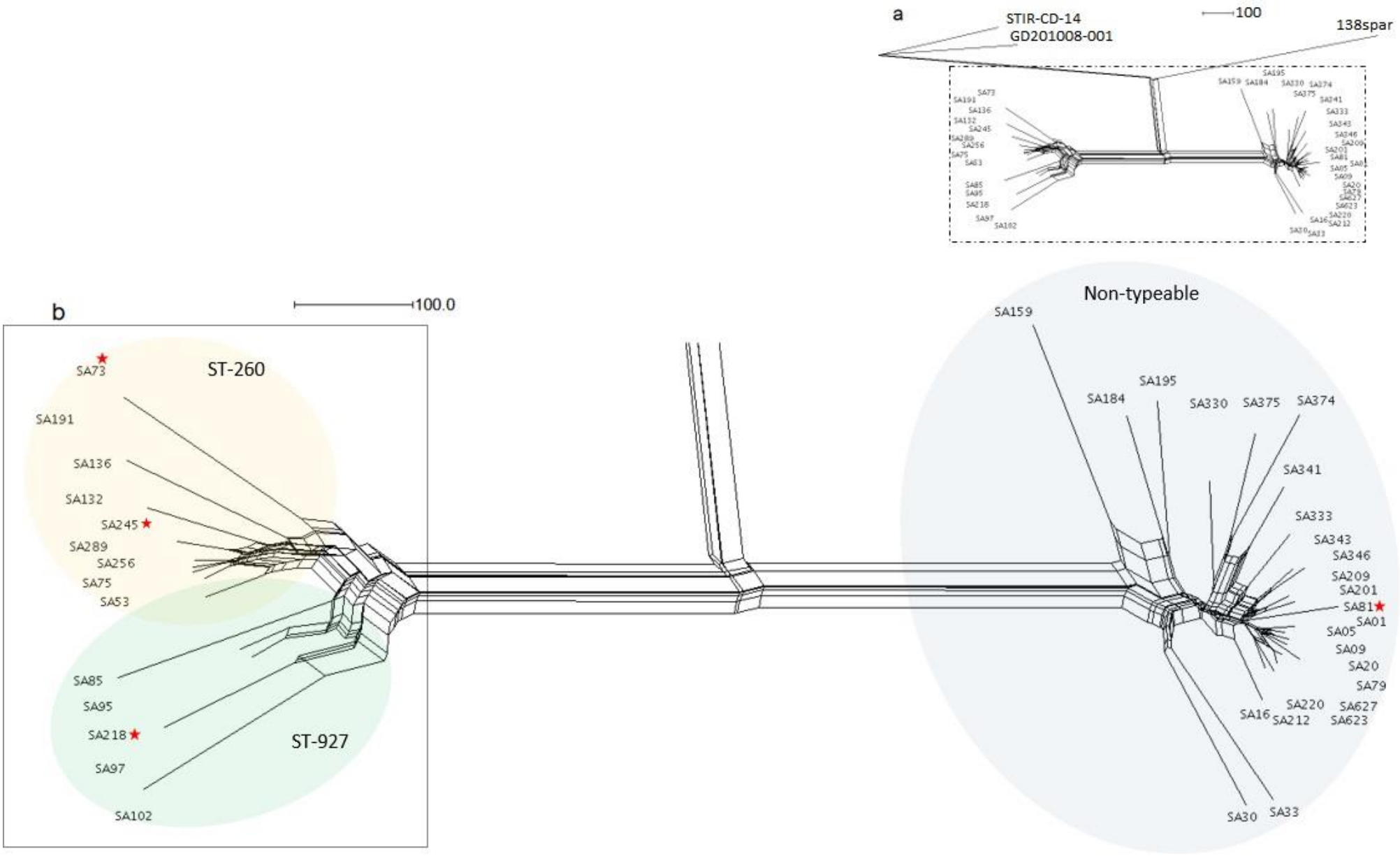
a
STIR-CD-14
GD201008-001
100
138spar
b
100.0
ST-260
ST-927
Non-typeable
SA73
SA191
SA136
SA132
SA245
SA289
SA256
SA75
SA53
SA85
SA95
SA218
SA97
SA102
SA159
SA195
SA184
SA330
SA375
SA374
SA341
SA333
SA343
SA346
SA209
SA201
SA81
SA01
SA05
SA09
SA20
SA79
SA16
SA220
SA627
SA212
SA623
SA30
SA33

Figure 5

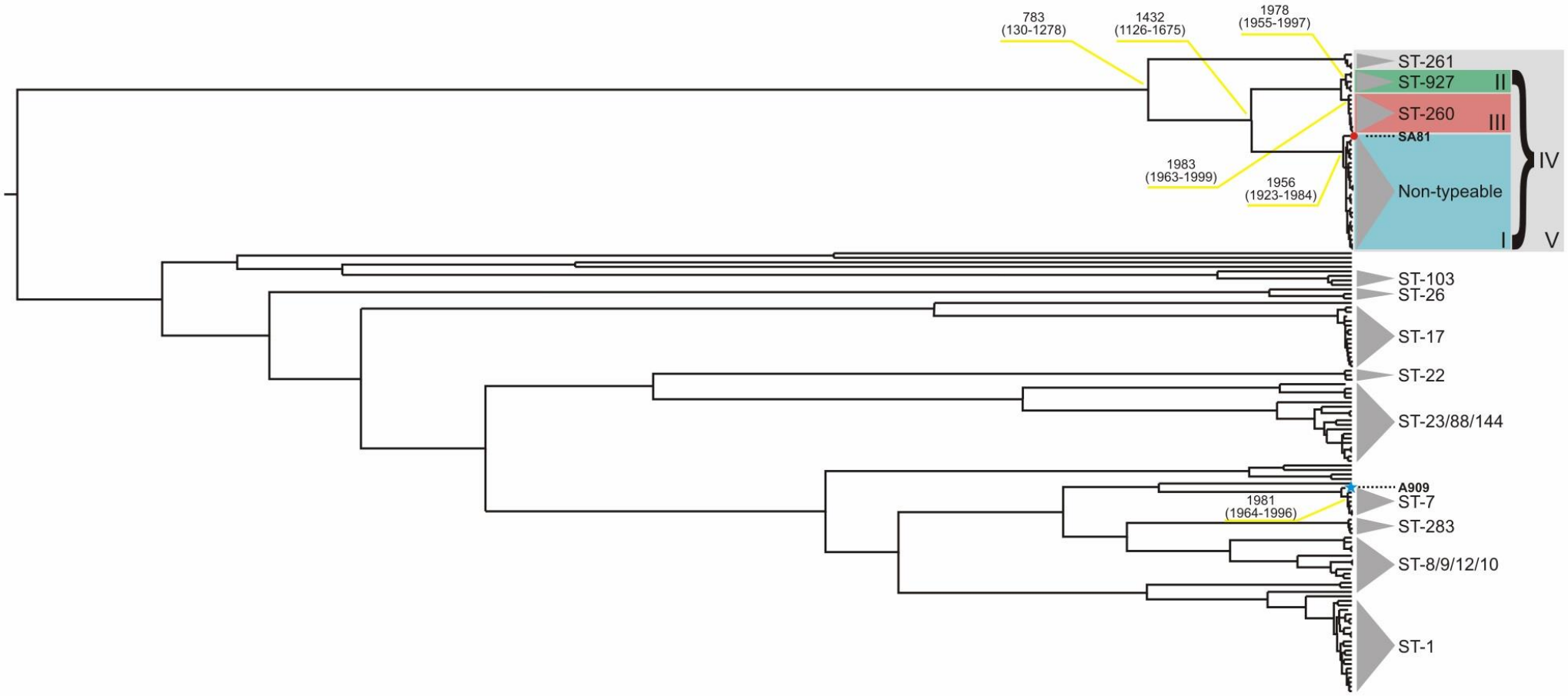
783
(130-1278)
1432
(1126-1675)
1978
(1955-1997)
1983
(1963-1999)
1956
(1923-1984)
ST-261
ST-927
II
ST-260
III
SA81
Non-typeable
I
IV
V
ST-103
ST-26
ST-17
ST-22
ST-23/88/144
A909
1981
(1964-1996)
ST-7
ST-283
ST-8/9/12/10
ST-1

Table 1

Isolate	Size (bp)	CDS[1]	Pseudogenes	tRNA	Coverage	Sequence-type
SA01	1841943	1656	172	62	76	NT[2]
SA05	1841945	1652	172	62	207	NT
SA09	1841929	1638	189	62	69	NT
SA16	1841859	1690	129	62	83	NT
SA20	1841952	1678	233	62	1443	NT
SA30	1841729	1592	234	62	163	NT
SA33	1841628	1599	224	62	161	NT
SA53	1848970	1700	123	64	187	260
SA73	1848838	1588	244	64	154	260
SA75	1849016	1667	159	64	233	260
SA79	1841946	1654	171	62	227	NT
SA81	1840363	1687	129	62	240	NT
SA85	1849989	1592	238	64	210	927
SA95	1856590	1707	122	64	304	927
SA97	1856410	1580	260	64	158	927
SA102	1849521	1510	320	64	135	927
SA132	1852032	1657	170	64	344	260
SA136	1849103	1643	180	64	377	260
SA159	1841483	1503	318	62	130	NT
SA184	1841893	1608	218	62	218	NT
SA191	1848676	1599	232	64	157	260
SA195	1841715	1596	232	62	175	NT
SA201	1841834	1691	144	62	239	NT
SA209	1841835	1684	141	62	268	NT
SA212	1841962	1683	141	62	256	NT
SA218	1849985	1706	121	64	269	927
SA220	1841963	1672	149	62	188	NT
SA245	1848955	1690	135	64	318	260
SA256	1848972	1687	138	64	206	260
SA289	1848987	1705	122	64	235	260
SA330	1842081	1694	132	62	222	NT
SA333	1842037	1705	119	62	226	NT
SA341	1842113	1702	129	62	322	NT
SA343	1841977	1702	118	62	464	NT
SA346	1841984	1698	122	62	424	NT
SA374	1842255	1644	184	62	162	NT
SA375	1842219	1684	149	62	192	NT
SA623	1842115	1729	98	64	134	NT
SA627	1842066	1729	99	64	227	NT

Supplementary table 1

Isolate	City	State	Region	Year
SA01	Esmeraldas	MG	Southeast	2003
SA05	Linhares	MG	Southeast	2003
SA09	Linhares	ES	Southeast	2005
SA16	Tietê	SP	Southeast	2006
SA20	Itambaracá	PR	South	2006
SA30	Tietê	SP	Southeast	2006
SA33	Zacarias	SP	Southeast	2006
SA53	Jaibaras	CE	Northeast	2007
SA73	Russas	CE	Northeast	2008
SA75	Jaguaribara	CE	Northeast	2008
SA79	Joinville	SC	South	2009
SA81	Lucas do Rio Verde	MT	Central-West	2009
SA85	Piranhas	AL	Northeast	2010
SA95	Piranhas	AL	Northeast	2010
SA97	Jatobá	PE	Northeast	2010
SA102	Petrolândia	PE	Northeast	2010
SA132	General Sampaio	CE	Northeast	2011
SA136	Jaguaribara	CE	Northeast	2011
SA159	Alfenas	MG	Southeast	2011
SA184	Pinheiros	ES	Southeast	2011
SA191	Jaguaruana	CE	Northeast	2011
SA195	São José dos Campos	SP	Southeast	2011
SA201	Linhares	ES	Southeast	2012
SA209	Buritana	SP	Southeast	2012
SA212	Rifaina	SP	Southeast	2012
SA218	Linhares	ES	Southeast	2012
SA220	Rifaina	SP	Southeast	2013
SA245	Betim	MG	Southeast	2013
SA256	Jaguaribara	CE	Northeast	2013
SA289	Jaguaribara	CE	Northeast	2013
SA330	Franca	SP	Southeast	2013
SA333	Goiás	GO	Central-West	2014
SA341	Alfenas	MG	Southeast	2014
SA343	Alfenas	MG	Southeast	2014
SA346	Alfenas	MG	Southeast	2014
SA374	São Paulo	SP	Southeast	2014
SA375	São Paulo	SP	Southeast	2014
SA623	Uberlândia	MG	Southeast	2015
SA627	Uberlândia	MG	Southeast	2015

Supplementary table 2

ST	Strain	Year	Year precision[1]
1	09mas018883	2003	10
1	BG_NI_011	2010	
1	CZ_NI_004	2008	
1	CZ_NI_006	2008	
1	CZ_NI_008	2008	
1	CZ_NI_009	2008	
1	CZ_NI_013	2009	
1	CZ_NI_015	2009	
1	DE_NI_001	2007	
1	DK_NI_012	2010	
1	DK_NI_013	2010	
1	DK_NI_022	2011	
1	GB_NI_009	2010	
1	GB_NI_010	2010	
1	GBS_ST_1	2015	
1	IT_NI_0031	2010	
1	IT_NI_028	2009	
1	MRI_Z1_212	2002	10
1	RBH05	2008	
1	SS1	1992	
3	DE_NI_022	2009	
7	A909	1975	5
7	GD201008	2010	
7	GX064	2011	
7	HN016	2011	
7	WC1535	2015	
7	YM001	2011	
7	ZQ0910	2009	2
8	BE_NI_005	2010	
8	PSS_7736	2005	10
9	DK_NI_008	2009	
10	DE_NI_004	2008	
10	DK_NI_015	2008	
12	BG_NI_007	2009	
12	BG_NI_010	2009	
17	BG_NI_002	2009	
17	COH1	1985	5
17	DE_NI_013	2009	
17	DE_NI_036	2010	

Supplementary table 2 continued

ST	Strain	Year	Year precision[1]
17	DE_NI_037	2010	
17	DK_NI_001	2009	
17	DK_NI_007	2009	
17	GB_NI_003	2009	
17	GB_NI_004	2010	
17	IT_NI_009	2008	
17	NGBS128	2010	
22	GBS1_NY	2012	
22	GBS2_NM	2012	
22	GBS6	2009	
23	759_SAGA	2013	
23	BE_NI_001	2009	
23	BG_NI_004	2010	
23	BG_NI_005	2009	
23	CZ_NI_001	2008	
23	CZ_NI_005	2008	
23	DE_NI_014	2009	
23	DE_NI_033	2010	
23	DE_NI_040	2010	
23	DK_NI_002	2009	
23	DK_NI_005	2009	
23	GB_NI_006	2010	
23	NEM316	1975	5
26	CNCTC_10_84	1970	6
26	IT_NI_036	2010	
26	IT_NI_037	2010	
28	BG_NI_009	2008	
61	SA111	2013	
88	DE_NI_012	2009	
88	DK_NI_014	2009	
88	DK_NI_016	2010	
103	GBS85147	1995	
103	M19	2010	
103	SA07	2005	
103	SA172	2011	
110	2603V_R	1993	
144	BG_NI_001	2009	
144	DE_NI_006	2008	
196	DK_NI_019	2010	
255	CZ_NI_007	2008	

Supplementary table 2 continued

ST	Strain	Year	Year precision[1]
260	SA132	2011	
260	SA136	2011	
260	SA191	2011	
260	SA245	2013	
260	SA256	2013	
260	SA289	2013	
260	SA53	2007	
260	SA73	2008	
260	SA75	2008	
261	138p	2007	
261	138SPAR	2011	
261	2_22	1986	
261	GX026	2011	
283	CU_GBS_08	2008	
283	CU_GBS_98	1998	
283	SG_M1	2015	
297	NGBS357	2011	
315	BE_NI_007	2010	
452	NGBS572	2012	
459	NGBS061	2010	
479	CZ_NI_014	2009	
609	ILRI005	2004	
617	ILRI112	2002	
739	FWL1402	2014	
927	SA102	2010	
927	SA218	2012	
927	SA85	2010	
927	SA95	2010	
927	SA97	2010	
NT	ATCC13813	1949	
NT	BE_NI_008	2010	
NT	DK_NI_021	2009	
NT	S25	2015	
NT	SA01	2003	
NT	SA020	2006	
NT	SA05	2003	
NT	SA09	2005	
NT	SA159	2011	
NT	SA16	2006	
NT	SA184	2011	

Supplementary table 2 continued

ST	Strain	Year	Year precision[1]
NT	SA195	2011	
NT	SA201	2012	
NT	SA209	2012	
NT	SA212	2012	
NT	SA220	2013	
NT	SA30	2006	
NT	SA330	2013	
NT	SA33	2006	
NT	SA333	2014	
NT	SA341	2014	
NT	SA343	2014	
NT	SA346	2014	
NT	SA374	2014	
NT	SA375	2014	
NT	SA623	2015	
NT	SA627	2015	
NT	SA79	2009	
NT	SA81	2009	
ND ST	BG_NI_006	2009	
ND ST	DE_NI_003	2008	
ND ST	GB_NI_005	2009	
ND ST	Gottschalk_992B	2006	6
ND ST	H002	2011	

5. FINAL CONSIDERATIONS

This work showed that Brazilian population of *Streptococcus agalactiae* is diverse, and only high discriminatory molecular methods can separate closely related strains. In this work, the wgMLST could show that each of the 39 strains involved in outbreaks in Brazilian fish farms is a single clone, and there are geographical correlations between isolates in terms of genetic proximity and there is a movement of different microbial lineages between regions. An evolutionary study showed that different lineages of piscine GBS strains had distinct emergences in the time, and Brazilian strains emerged into a group which hosts exclusively fish species.

Future studies will aim to discover other features on the genomes of the Brazilian GBS isolates, to help in the knowledge of this pathogen and to improve its control in the field.