U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

# DNA markers for detection and infrasubspecific discrimination of mastitis-causing Streptococcaceae

Dissertação de Mestrado em Genética Forense

Alexandre Miguel Santos Almeida

**Faculdade de Ciências da Universidade do Porto**

**2012**

# DNA markers for detection and infrasubspecific discrimination of mastitis-causing Streptococcaceae

Dissertação submetida à Faculdade de Ciências da Universidade do Porto para obtenção do grau de Mestre em Genética Forense.




**Local de realização:**
Departamento de Biologia
Faculdade de Ciências da Universidade do Porto

**Orientadores:**
Professor Doutor Fernando Tavares
Faculdade de Ciências da Universidade do Porto
Centro de Investigação em Biodiversidade e Recursos Genéticos

Doutor Ricardo Araújo
Instituto de Patologia e Imunologia Molecular da Universidade do Porto

"*A scientific truth does not triumph*
*by convincing its opponents and making them see the light,*
*but rather because its opponents eventually die*
*and a new generation grows up that is familiar with it.*"
- Max Planck

# Acknowledgements

Individual effort and skill can only take you so far. Without the help and dedication of colleagues, friends or family you'll never accomplish much in life. Throughout this past year, several people in one way or another were essential for the outcome of this work.

First and foremost, I would like to thank my supervisor Prof. Fernando Tavares for all the dedication, cooperation and friendship granted during my thesis. In spite of being very busy with numerous responsibilities, he always managed to teach and guide me throughout this endeavour and make sure the work was successful. With his articulate speeches and guidance, we always managed to solve any challenge we faced.

I also have to issue a very special thanks to my laboratory partner Pedro Albuquerque. His tremendous kindness and willingness to always help and teach was one of the main driving forces for this work. He taught me all the techniques and protocols used throughout this year and even during his final months of his PhD he was always available to help and offer his expertise on the theoretical and technical aspects of the work. Without him this experience would have been extremely more difficult.

A special thanks to my co-supervisor Ricardo Araújo who, in spite of being institutionally apart, was always on par with the work, available to help and offer his interesting perspective and collaboration.

I would also like to thank Prof. Niza Ribeiro for offering his expertise as a more directly involved researcher in the subject of bovine mastitis. A special thanks as well to Paula Santos and João Sousa from SEGALAB for providing the bacterial isolates used during this study, which were essential for the experimental analyses.

Also, thank you to Prof. António Amorim, leader of the Population Genetics group of IPATIMUP, for giving me the opportunity to develop my thesis in the Sciences Faculty with Prof. Fernando Tavares.

Although these were the people directly involved in my dissertation, I would also like to thank my friends and colleagues from the Population Genetics group of IPATIMUP, Ana Carolina, Catarina Seabra, Catarina Xavier, Filipa Cabral, Inês Martins, Marisa Oliveira and Sofia Marques for their friendship and numerous fun times shared these

past couple of years. A special word to Ana Carolina who has a very special place in my heart for always supporting me during the good and the bad, and who has given me some of the best moments of my life.

A special thanks also to Maria João Fonseca and Eduarda Almeida from the Microbial Diversity and Evolution group of CIBIO, who welcomed me kindly to the group and were great companions in the laboratory during my work.

Last but not least, thank you to my parents for continuing to allow me to pursue my academic goals, and the rest of my friends and family who I may have forgotten during these busy times but that still hold a precious place in my life.

# Abstract

Molecular methods have shown to be a fast and reliable approach for bacterial detection and identification. Among these, DNA-based methods encompass some of the most promising approaches, due to the development of several efficient techniques. Moreover, increasingly larger and more informative genome databases allow effective *in silico* analyses for the selection of taxa-specific DNA signatures. On the other hand, biochemical and phenotypic tests, frequently employed in routine diagnostic laboratories, present a limited accuracy and are inherently biased towards culturable organisms.

The applicability of these methods range from disease diagnosis and identification of biological agents to metagenomic studies and community profiling. In the field of veterinary medicine, bovine mastitis, an inflammatory disease in the mammary gland, is currently a major concern affecting dairy herds worldwide. Due to changes in milk quality and composition, this disease is responsible for significant financial losses in the dairy industry. Over 150 pathogenic agents have been identified, with particular prevalence of those belonging to the Streptococcaceae family. Thus, efficient detection and typing methods are required for disease prevention, source tracking, treatment assessment and control.

The aim of this work was to develop a rapid, reliable and inexpensive platform for the detection of prevalent bovine mastitis pathogens within the Streptococcaceae family and gain additional insight into the infrasubspecific diversity of this group to improve epidemiological characterisation.

Using Insignia and the Protein Family Database (Pfam), DNA signatures were selected for well-known bovine mastitis-causing taxonomic ranks: a broad-spectrum marker for the Streptococcaceae family, taxa-specific markers for the *Lactococcus* and *Streptococcus* genera, and specific markers for *Streptococcus agalactiae* and *Streptococcus uberis*, two particularly prevalent mastitis pathogens. Additionally, markers of functional traits associated with the virulence potential of bovine mastitis strains were used, based on the fructose and nisin operon. The virulence-associated genes *hasC*, *gapC* and *oppF*, frequently described in *S. uberis* strains, were also selected.

Experimental validation was carried out by PCR and dot blot hybridisation, and an image algorithm was used to allow an operator-independent interpretation of the results. A set of 44 reference strains and isolates, representative of the Streptococcaceae family, of closely related species and of organisms with matching hosts, was tested with the selected DNA markers. The isolates used were obtained from different mastitic milk samples of distinct locations within Portugal and previously identified by the automated identification system VITEK 2. Sequencing analysis of the 16S rRNA gene revealed an incorrect identification of some of these isolates, emphasising the increased reliability and accuracy of molecular methods.

Based on the results obtained, the broad-spectrum taxonomic marker was specific to the *Streptococcus* genus and the markers selected for *Lactococcus*, *S. agalactiae* and *S. uberis* were shown to be specific to the corresponding taxa. The functional markers allowed increased discrimination of strain-specific patterns of *S. agalactiae* and *S. uberis*: the fructose operon markers were specific to bovine isolates of *S. agalactiae* and the nisin operon markers were present in a particular cluster of strains with a common origin. Furthermore, dot blots using the virulence-associated markers revealed specific patterns that were able to discriminate additional species, such as *Streptococcus bovis* and *Streptococcus parauberis*, and detect other organisms closely related to the Streptococcaceae family.

This work suggests that the combined use of taxa-specific and functional markers presents a promising approach for the reliable, rapid and cost-effective detection and typing of bovine mastitis-causing pathogens, for the treatment and control of this disease.

# Resumo

Os métodos moleculares têm vindo a adquirir um papel de destaque para a deteção e identificação bacteriana. As técnicas baseadas em DNA em particular englobam algumas das metodologias mais promissoras devido ao desenvolvimento de diversas técnicas eficazes. Além disso, a disponibilidade de bases de dados de genomas de microrganismos cada vez mais informativas permitem análises *in silico* fidedignas para a obtenção de assinaturas de DNA taxa-específicas. Por outro lado, testes bioquímicos e fenotípicos, frequentemente utilizados em laboratórios de rotina, apresentam uma precisão limitada e são enviesados inerentemente para organismos cultiváveis.

A aplicabilidade destes métodos vai desde o diagnóstico de doenças e agentes biológicos a estudos de metagenómica e caracterização de comunidades. A mastite bovina, uma doença inflamatória na glândula mamária, é atualmente uma das grandes preocupações para a medicina veterinária, afetando o gado bovino por todo o mundo. Isto leva a alterações significativas quer na composição, quer na qualidade do leite, resultando em grandes prejuízos económicos para a indústria. Mais de 150 agentes patogénicos já foram identificados, com especial prevalência de organismos pertencentes à família Streptococcaceae. Deste modo, métodos eficazes de deteção e tipagem são essenciais para prevenir, controlar e avaliar a eficácia do tratamento desta doença.

O objetivo deste trabalho foi desenvolver uma plataforma rápida, barata e fidedigna para a deteção de agentes patogénicos responsáveis por mastites bovinas pertencentes à família Streptococcaceae e inferir alguma diversidade infrasubespecífica para melhorar a caraterização epidemiológica.

Através das ferramentas Insignia e Protein Family Database (Pfam), foram selecionadas assinaturas de DNA para níveis taxonómicos frequentemente associados a organismos responsáveis por mastites bovinas: um marcador de largo espetro para a família Streptococcaecae, marcadores para os géneros *Lactococcus* e *Streptococcus*, e marcadores específicos para dois patogénicos com especial interesse, *Streptococcus agalactiae* e *Streptococcus uberis*. Além disso, foram utilizados marcadores funcionais de caraterísticas fenotípicas associadas à virulência de estirpes de mastites, baseados no operão da frutose e da nisina. Foram

selecionados também três genes de virulência descritos frequentemente em *S. uberis*: *hasC*, *gapC* e *oppF*.

A validação experimental foi realizada por PCR e hibridação em dot blot, complementada com um programa de análise automática de imagem para uma interpretação objetiva dos resultados obtidos. Os marcadores selecionados foram testados com um conjunto de 44 isolados e estirpes de coleção, representativo da família Streptococcaceae, de organismos filogeneticamente próximos e de outros que partilham um habitat comum. Os isolados foram obtidos a partir de amostras de leite contaminadas provenientes de diferentes localidades de Portugal e previamente identificados pelo sistema de identificação bacteriano VITEK 2. A sequenciação do gene 16S rRNA revelou que alguns isolados tinham sido mal identificados, o que realça o facto de os métodos moleculares serem alternativas mais exatas e precisas.

Com base nos resultados obtidos, o marcador taxonómico de largo espetro foi específico para o género *Streptococcus* e os marcadores selecionados para *Lactococcus*, *S. agalactiae* e *S. uberis* mostraram ser específicos para os grupos pretendidos. Os marcadores funcionais permitiram inferir alguns padrões específicos de linhagens de *S. agalactiae* e *S. uberis*: os marcadores do operão da frutose demonstraram ser específicos para isolados bovinos de *S. agalactiae*, enquanto os marcadores do operão da nisina foram detetados num conjunto de isolados de *S. agalactiae* com uma origem comum. Além disso, os ensaios com os marcadores de genes de virulência permitiram discriminar outras espécies, incluindo *Streptococcus bovis*, *Streptococcus parauberis* e outros organismos filogeneticamente próximos da família Streptococcaceae.

Este trabalho demonstra que o uso simultâneo de marcadores taxonómicos e funcionais é uma estratégia promissora para a deteção e tipagem eficazes de patogénicos responsáveis por mastites bovinas, podendo assim contribuir para um melhor tratamento e controlo desta doença.

# List of tables

# List of figures

# List of abbreviations

**API -** Analytical Profile Index

**ATCC** - American Type Culture Collection

**BHI** - Brain Heart Infusion Medium

**BLAST** - Basic Local Alignment Search Tool

**bp** - Base pair

**CAI** - Codon Adaptation Index

**CDS** - Coding DNA Sequence

**CMT** - California Mastitis Test

**CUPID** - Core and Unique Protein Identification

**DGGE** - Denaturing gradient gel electrophoresis

**DNA** - Deoxyribonucleic acid

**dNTPs** - Deoxyribonucleotide triphosphates

**EC** - Electrical conductivity

**eCAI** - Expected Codon Adaptation Index

**ELISA** - Enzyme Linked Immunosorbent Assay

**FISH** - Fluorescence *in situ* hybridisation

**GAPDH** - Glyceraldehyde-3-phosphate dehydrogenase

**GC content** - Percentage of guanine or cytosine bases

**MRS** - de Man, Rogosa and Sharpe Medium

**MLST** - Multilocus sequence typing

**MLVA** - Multilocus variable number of tandem repeats analysis

**ORF** - Open Reading Frame

**PCR** - Polymerase Chain Reaction

**Pfam** - Protein Family Database

**PFGE** - Pulse field gel electrophoresis

**PTS** - Phosphotransferase system

**RAPD** - Randomly amplified polymorphic DNA

**rRNA** - Ribosomal ribonucleic acid

**RFLP** - Restriction fragment length polymorphism

**RNA** - Ribonucleic acid

***Taq*** - DNA polymerase from *Thermus aquaticus*

# Table of contents

# Introduction

## 1. Bacterial identification, detection and typing

The germ theory of disease, proposed and validated between the 17th and 19th century, marks the beginning of clinical microbiology and its relevance for modern medicine. The work of Anton van Leeuwenhoek, Louis Pasteur, Robert Koch and other notable figures contributed to the discovery that the presence and action of microorganisms in the human body are the cause of many diseases. Thus, bacterial identification and detection became essential throughout the following years to fields such as medicine, forensics, biotechnology and agriculture (Nakanishi *et al.*, 2009; Trevors and Masson, 2010; Shome *et al.*, 2011). Bacterial species are responsible for numerous diseases, including pneumonia, meningitis and tuberculosis (Hernandez-Pando *et al.*, 2000; Aguilar *et al.*, 2010; Zancolli *et al.*, 2011), and their diagnosis and treatment is dependent on the accurate identification of the causative agent. In addition, community profiling can be used as a complementary tool for forensic identification (Fierer *et al.*, 2010), whereas the advent of bioterrorist acts also highlights the need for simple and rapid methods of bacterial identification, detection and typing (Dance, 2008). Currently, the available tools for the identification and detection of bacterial species are based on phenotypic assays, serological methods or molecular biology techniques.

Phenotypic methods of bacterial identification are based on the ability to distinguish metabolic and morphological features of known bacteria. Culture in selective or differential media, biochemical tests, such as API 20, and automated identification systems like VITEK are some of the main phenotypic approaches currently used, primarily due to their accessible cost (Torsvik *et al.*, 1990; O'Hara *et al.*, 2000). However, culture-based methodologies are generally time-consuming, technically demanding and less reliable (Fortin *et al.*, 2003; Saini *et al.*, 2012).

Serological techniques, such as the Enzyme Linked Immunosorbent Assay (ELISA), allow the identification and detection of bacterial species based on the antibody-antigen interaction (Engvall and Perlmann, 1972). Nevertheless, these methods cannot be used in immunosuppressed individuals for the diagnosis of several diseases, are less specific to particular organisms and can only be applied to known bacteria (Jacobs, 1993; Daleine and Lagrange, 1995).

Therefore, culture-independent tools may stand out as more reliable, specific and cost-effective techniques for bacterial identification and detection.

## 1.1. Molecular methods

### 1.1.1. Identification and detection techniques

DNA-based methods are now the leading technology used for identification and detection purposes. PCR-based techniques specifically amplify a part of the DNA sequence to be subsequently analysed and compared. With its high specificity and sensitivity it can accurately detect a selected target organism even in dead or growth-inhibited bacterial cells (Forsman *et al.*, 1997).

Furthermore, real-time PCR can also be used for quantification of bacterial cells, in which the amount of DNA specifically amplified is quantified throughout the reaction (Taponen *et al.*, 2009).

In spite of these many advantages, PCR technology has its limitations. Due to its high sensitivity, contamination from non-template DNA present in the work environment can lead to deceptive results, requiring several precautions to minimise this problem. Moreover, sequencing analyses of PCR amplicons are hampered by conventional polymerase errors that can occur, due to the enzyme's inability to correct misincorporated nucleotides. PCR-based technology is also dependent on primer specificity, especially at the 3' end. In addition, false negative results can occur due to PCR inhibitors, or nonspecific amplification due to less restrictive PCR conditions (Wilson, 1997).

Hybridisation-based techniques are a viable alternative for detection purposes. Unlike PCR technology, a DNA probe is specifically hybridised against the DNA sequence of the target organism, under high-stringency conditions. Therefore, the hybridisation signals detected are highly specific. Nevertheless, these methods are, in general, more laborious, time-consuming and costly, usually requiring a previous PCR amplification to obtain the DNA probe.

DNA microarrays, fluorescence *in situ* hybridisation (FISH), southern blot and dot blot techniques have shown reliable and promising results as detection platforms (Amann *et al.*, 2001; Volokhov *et al.*, 2006; Vieira *et al.*, 2007; Zhang *et al.*, 2007).

Microarray assays consist of hybridisation between hundreds of DNA targets immobilized in a microarray slide, and the DNA probe conjugated with a fluorescent or chemiluminescent dye. However, this particular method is technically demanding and based on the use of a high number of reduced sized markers (less than 100 bp), which poses added costs to routine laboratories (Perez *et al.*, 2004).

Dot blot hybridisation, a more viable and inexpensive alternative, has been successfully used for the detection of a number of bacterial species (Wirawan *et al.*, 2006; Vieira *et al.*, 2007; Albuquerque *et al.*, 2011). Instead of a microarray slide, a nylon membrane is used as support for either the DNA of the targeted organisms (traditional dot blot) or multiple DNA markers (inverted dot blot), enabling the simultaneous detection of selected markers with increased length.

### 1.1.2. Typing techniques

Identification and detection techniques are essential for determining the pathogenic agent of interest. On the other hand, genotyping techniques, capable of discriminating organisms to the infrasubspecific level, allow identification of different strains, source tracking and identification of transmission routes (Zadoks and Schukken, 2006).

A number of established methods have been used as genotyping techniques. Pulse field gel electrophoresis (PFGE) consists in separating DNA band fragments, after enzymatic digestion, by electrophoresis of increased resolution. Multilocus sequence typing (MLST) is based on the comparative sequence analysis of housekeeping genes, which allows improved investigation of population structure and evolution (Maiden *et al.*, 1998; Rato *et al.*, 2008). Multilocus variable number tandem repeat analysis (MLVA) is another viable typing technique, discriminating different polymorphisms based on the number of repeats of specific loci. Other techniques, like restriction fragment length polymorphism (RFLP) and randomly amplified polymorphic DNA (RAPD) are also reliable alternatives, but usually require the previous species identification. These methods coupled with reliable diagnostic approaches can be used as efficient tools for the identification, detection and typing of bacterial species in complex environments.

Alternatively, for community profiling, denaturing gradient gel electrophoresis (DGGE) can be used, discriminating small differences in DNA fragments due to their relative migration in denaturant conditions.

The development of effective and reliable methods for bacterial identification, detection and genotyping will help in the study of microbial populations in the areas of medical science, food microbiology and forensics (Gunzburg *et al.*, 1995; Ercolini, 2004; Kuang *et al.*, 2009).

## 1.1.3. DNA signatures

The concept of DNA signatures was first proposed by Phillipy (Phillippy *et al.*, 2007), as a sequence of nucleotides present in a particular organism or group of organisms and absent from all other species. However, the selection of DNA signatures for bacteria discrimination has been made mostly within phylogenetic or functional genes associated with bacteria virulence (Gunzburg *et al.*, 1995; McDonald *et al.*, 2005). Phylogenetic markers, such as the 16S rRNA gene, can present high sequence similarities in closely related species and considerable intragenomic heterogeneity, leading to a low discriminatory potential (Michon *et al.*, 2010). Virulence-related functional markers, on the other hand, require a comprehensive knowledge of the bacterial metabolism, a particularly difficult endeavour in unculturable or poorly characterised bacteria, and are usually within highly dynamic and variable DNA regions. Nevertheless, in spite of being less specific, functional markers can help gain insight into strain-specific patterns and traits of particularly virulent organisms (Ote *et al.*, 2011; Reinoso *et al.*, 2011).

The increasingly larger and more reliable genomic databases allow accurate and efficient *in silico* analyses for the selection of discriminatory taxa-specific markers within the entire genome. Nevertheless, these bioinformatics tools are inherently biased towards fully sequenced organisms, requiring additional validation by reliable molecular methods.

## 1.1.4. Bioinformatics tools

Insignia (Phillippy *et al.*, 2007) is an online utility that calculates target-specific DNA regions based on user-defined organisms. Previous studies have shown the reliability of this database as a preliminary tool for the selection of specific DNA signatures (Albuquerque *et al.*, 2012). Most of the information is already calculated, so results can be quickly obtained. One of the main advantages of this application is the amount of flexibility it delivers for inputting experimental constraints: signature length, melting temperature and GC content can all be tweaked for a specific assay.

The Protein Family Database (Pfam) (Finn *et al.*, 2006) is an online application suited for genomic and proteomic analysis, consisting of a large collection of protein families associated with sequenced taxonomic groups. This enables the user to search for taxa-specific domains of selected organisms (Vieira *et al.*, 2007).

CUPID is a freely available database of taxa-specific proteins. Genus, species or strain levels' specificity of available Open Reading Frames (ORF) are calculated by identifying the most closely related organism (Mazumder *et al.*, 2005).

Complementing these tools for signature selection is the BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1990) application. This utility uses a specific algorithm to assess the specificity of a selected sequence against the NCBI database, the largest available nucleotide databank. Nevertheless, it can also present some limitations in finding sequence similarities, due to algorithm simplifications for faster outputs (Nordberg, 2005).

Other bioinformatics tools have also been successfully used as preliminary *in silico* diagnostic methods (Albuquerque *et al.*, 2009), but their reliability and efficiency is dependent on the available information in each database. Therefore, *in silico* analyses require further experimental validation through accurate and reliable molecular techniques.

The selection and validation of taxa-specific DNA signatures can help solve numerous challenges in clinical microbiology, veterinary medicine, microbial ecology and other biological sciences.

## 2. Bovine mastitis

Derived from the Greek word *mastos* (breast), mastitis refers to an inflammatory disease in the mammary gland, affecting dairy herds worldwide. Based on the severity of disease, mastitis is divided into clinical (symptomatic) and subclinical (asymptomatic) mastitis (McDonald, 1979; Jones and Bailey, 2009).

Causes of inflammation range from physical trauma to chemical irritants, but the most common cause of disease is pathogenic microorganisms. The teat skin cells act as a first line of defence against these infectious agents by producing keratin, a fibrous protein combined with lipid components that have bacteriostatic effects. However,

during inadequate milking procedures, lesions can occur and the teat canal becomes highly vulnerable. Furthermore, after milking, the teat canal remains dilated for 1-2 hours, increasing the likelihood of bacterial infection (Jones and Bailey, 2009). When pathogens enter the teat canal, they multiply and release toxins, enzymes and surface proteins which are responsible for adherence to the host's extracellular matrix. Altogether these induce an inflammatory response from the host, increasing the number of polymorphonuclear neutrophils, phagocytes and other leukocytes (Fig. 1) (Jones and Bailey, 2009; Ote *et al.*, 2011). The immune response can vary greatly, depending on the causative agent, lactation stage, age and health status of the cow (Harmon, 1994). Due to this somatic cell increase, milk quality and composition is significantly altered, reducing its economic value. Thus, mastitis is considered one of the costliest diseases of the dairy industry (Kitchen, 1981), causing, for instance, an annual financial loss of 1.7-2 billion dollars in the U.S (Jones and Bailey, 2009).

A significant number of microbial organisms have been isolated from the bovine mammary gland, indicating that mastitis infections can be caused by over 150 different species, belonging mostly to three major groups of organisms: *Staphylococcus, Streptococcus* and coliforms. Other mastitis-causing agents have been identified, albeit less frequently, including *Enterococcus*, *Mycoplasma*, pseudomonads, algae and yeasts (Hale *et al.*, 1962; Watts, 1988; Zaror *et al.*, 2011).

Pathogenic agents can be found either in the udder (contagious pathogens) or in the cow's surroundings (environmental pathogens) and this distinction is correlated to their behaviour in dairy herds. Longer and more prevalent infections are caused by contagious organisms that spread during the milking process, whereas environmental agents typically cause a more clinically severe case of mastitis. The most common contagious pathogens are *Staphylococcus aureus*, *Streptococcus agalactiae* and *Streptococcus dysgalactiae,* whereas the most prominent environmental agents are *Streptococcus uberis, Escherichia coli* and *Klebsiella pneumonia* (McDonald, 1979).

Recently, with the development of more precise diagnostic techniques, the classical distinction between contagious and environmental agents is in question. Studies indicate that some bacterial strains within a species can display a contagious transmission pattern while others present an environmental origin (Munoz *et al.*, 2007; Zadoks *et al.*, 2011). This suggests that a more comprehensive diagnosis and assessment is required to classify a mastitis-causing organism with a contagious or environmental etiology.

The Streptococcaceae family plays a significant role in bovine mastitis with a high number of identified species from this taxon directly responsible for disease, including *Lactococcus lactis*, *S. agalactiae* and *S. uberis* (Kuang et al., 2009). Further research will help gain an insight into the importance of other organisms belonging to this family for the study of bovine infections.

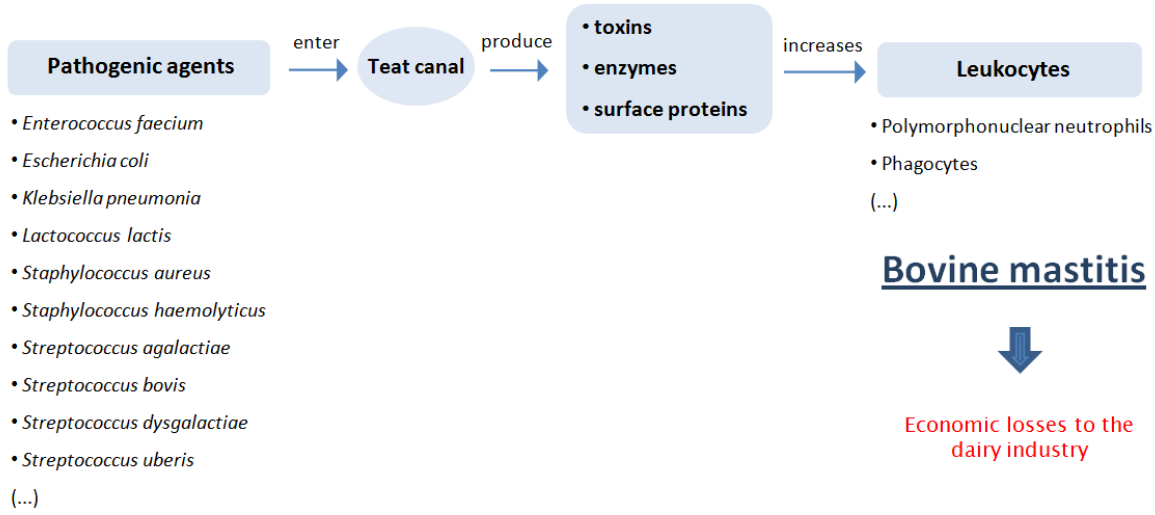


**Figure 1**. Mechanism of the bovine mastitis disease.

## 2.1. Streptococcaceae

This family of gram-positive bacteria, placed within the order of Lactobacillales consists essentially of two main genera: *Lactococcus* and *Streptococcus*. Based on previous studies, both groups have been associated with mastitis-causing agents (Kuang *et al.*, 2009). An additional genus, *Lactovum*, has been identified belonging to the Streptococcaceae family (Fig. 2), without known etiological relation to bovine mastitis.

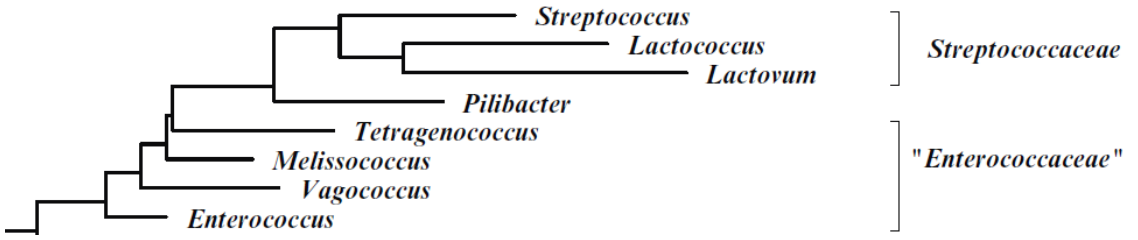


**Figure 2**. Phylogenetic tree, based on small-subunit rRNA gene sequences, of the Streptococcaceae family and closely related taxa (Ludwig et al., 2009).

### 2.1.1. *Lactococcus*

Lactococci, a group of gram-positive microaerophilic bacteria, have been involved in the dairy industry throughout the years. They can be identified as spherical or ovoid cells with 0.5-1 µm diameter (Teuber and Geis, 2006). As the first bacteria to be purely cultured, they have particular significance for microbiology (Lister, 1873). Initially identified as *Streptococcus lactis,* lactococci were later separated into a new genus (Schleifer *et al.*, 1985). Currently, they are used in industrial fermentations as manufacturers of dairy products.

Animal environments are the main habitats of lactococci. Originally present only in the dairy cow and raw milk, colonisation and expansion to other animal species, due to evolutionary changes, has been suggested (Teuber and Geis, 2006).

Lactococci are nutritionally fastidious organisms, requiring complex media for optimal growth. Isolation from dairy products is even more troublesome due to the abundance of solid or semi-solid fat containing material, but there are a number of available and published isolation methods (Endo *et al.*, 2011; Pavlidou *et al.*, 2011; Yu *et al.*, 2011).

In regards to bovine mastitis, *L. lactis* and *Lactococcus garviae* stand out as the most prevalent agents (Teuber and Geis, 2006; Kuang *et al.*, 2009; Wyder *et al.*, 2011). *L. lactis*, a mastitis causing agent, has also been studied for its production of bacteriocins with antimicrobial effects against more significant mastitis pathogens (Lee *et al.*, 2011).

### 2.1.2. *Streptococcus*

This genus of gram-positive cocci comprises a diverse group of species, normally colonising mammalian skin membranes as commensal organisms. Although *Streptococcus* abundantly inhabit the mucosa and skin surface of mammals, they are also a cause of disease and some are even considered primary pathogens (Cleary and Cheng, 2006).

Regarding bovine mastitis, *S. agalactiae*, *Streptococcus bovis*, *S. dysgalactiae*, *S. uberis* and other streptococci have been studied and identified as causative agents, with *S. agalactiae* and *S. uberis* standing out as two particularly significant mastitis pathogens (McDonald *et al.*, 2005; Jones and Bailey, 2009; Kuang *et al.*, 2009; Unnerstad *et al.*, 2009; Wyder *et al.*, 2011).

The species *S. agalactiae*, causing mostly subclinical infections, is one of the most prevalent contagious pathogens. Since a single strain is able to infect multiple animals in a herd, particular importance is given to the study of the epidemiology of these pathogens (Jones *et al.*, 2003; Zadoks *et al.*, 2011). The specific microenvironment of the udder is necessary for the growth of this species, and differences in pathogenicity between strains is linked to factors determining their ability to adhere to the mammary epithelium. Infections caused by *S. agalactiae* are generally low-grade and persistent, but can be readily eliminated with intramammary therapy (Keefe, 1997). Recent studies of *S. agalactiae* have identified distinct populations between human and bovine mastitis isolates (Zadoks *et al.*, 2011). These differences have been correlated to the bovine strains' acquisition of genes through interspecies horizontal transfer, resulting in environmental adaptation and a competitive advantage of these strains during infection of bovine hosts (Richards *et al.*, 2011).

The pathogen *S. bovis*, commonly identified as a mastitis agent, has been described as a genotypically heterogeneous group and is found primarily in the intestinal tract of bovines (Wyckoff *et al.*, 1997).

In regards to *S. dysgalactiae*, it has been frequently associated with both clinical and subclinical bovine mastitis (Rato *et al.*, 2011). Several virulence factors have been identified, associated with surface proteins that specifically interact with the host's extracellular matrix, or plasma proteins (Mamo *et al.*, 1987; Leigh *et al.*, 1998). Other virulence-associated genes coding for mastitis-causing proteins, such as alpha-2-macroglobulin, immunoglobulin G- or immunoglobulin A-binding proteins, have also been described (Valentinweigand *et al.*, 1990).

In *S. uberis*, mostly an environmental agent, a number of virulence factors related to the pathogenesis of bovine mastitis have been identified (Reinoso *et al.*, 2011). Some of the most prevalent genes are responsible for promoting invasion of the host tissue, survival in the host environment, evasion of the host immune response and internalization in the mammary gland cells, suggesting a particular importance of this virulence pattern. Furthermore, recent studies have identified, within *S. uberis* strains, a nisin U operon with close similarity to *S. agalactiae*, which has been suggested to provide these bacteria with a competitive advantage during mastitis infection (Wirawan *et al.*, 2006; Richards *et al.*, 2011).

## 2.2. Diagnostic techniques

Etiological agents and microbiological profiles vary greatly between different geographic regions and also between types of mastitis (clinical and subclinical), therefore requiring fast and efficient detection methods (Petrovski *et al.*, 2011; Sharma *et al.*, 2012)

Bovine mastitis' monitoring and diagnosis can be based solely on clinical signs, by visual observation of abnormal changes, or indirect measurements, like somatic cell count using the California Mastitis Test (CMT) (Dohoo and Leslie, 1991) and electrical conductivity (EC) measurement of the milk using a hand-held meter (Hillerton and Semmens, 1999). However, these techniques do not identify the causative agent, which is essential for the prevention, treatment and control of this disease.

To circumvent this limitation, bacteriological culturing methods are implemented in the routine identification of mastitis agents, using selective growth medium for prevalent known pathogens (Sears and McCarthy, 2003) or automated bacterial identification systems like VITEK 2. These phenotypic tests are based on particular differences in bacteria metabolism. Nevertheless, these methods also carry some limitations, namely the fact that they can only detect culturable organisms and are inherently biased towards those that grow more rapidly (Amann *et al.*, 1995). The accuracy of these tests is also a major disadvantage, ranging from 50% to 70% in some cases (Ieven *et al.*, 1995; Bal *et al.*, 2010). In fact, some bacteria cannot be efficiently differentiated by biochemical tests, including the mastitis-causing *S. uberis*, which cannot be distinguished from *Streptococcus parauberis* (Facklam, 2002).

Culture-independent molecular methods are now becoming increasingly important as mastitis detection techniques, providing a more accurate and reliable approach. A number of studies have already been published using PCR-based techniques for identification of mastitis pathogens (Lee *et al.*, 1998; Hassan *et al.*, 2001). A multiplex PCR study has been developed for the simultaneous detection of 10 prevalent mastitis pathogens (Shome *et al.*, 2011), while another was specifically developed towards streptococcal species using the *cpn60* gene (Dmitriev *et al.*, 2006).

Hybridisation-based techniques have also been used in the detection of bovine mastitis-causing pathogens. Microarray technology and dot bot hybridisation studies

have been published for detection of *S. uberis*, allowing to infer genome diversity and plasticity (Lang *et al.*, 2009).

## 2.3. Epidemiological studies

Diagnostic techniques play a significant role in determining mastitis etiology and a number of detection platforms have already been developed for well-known pathogens. Furthermore, genotyping techniques are also an essential element to decide on treatment and control. This is especially significant considering that bovine mastitis can either be caused by contagious or environmental pathogens. Epidemiological research can help understand the population structure, diversity and behaviour of important species such as *S. agalactiae* and *S. uberis*, the most prevalent contagious and environmental agents, respectively (Keefe, 1997; Rato *et al.*, 2008).

In regards to mastitis epidemiology, DGGE and RFLP studies of 16S rRNA genes have been described (McDonald *et al.*, 2005; Kuang *et al.*, 2009). However, these techniques have not been successfully used for strain identification, showing limited discriminatory resolution. On the other hand, PFGE and MLST studies have been successfully used, describing infrasubspecific diversity (Mork *et al.*, 2005). MLST platforms, based on polymorphisms of housekeeping genes, have been published for important species identified in the bovine environment, including *S. agalactiae* and *S. uberis* (Jones *et al.*, 2003; Zadoks *et al.*, 2005; Rato *et al.*, 2008). More recently, MLVA has also been used as a reliable typing technique for the epidemiological characterisation of *S. agalactiae* (Radtke *et al.*, 2012).

# 3. Objectives

Streptococcaceae is one of the major taxa responsible for bovine mastitis. In the present work the general goal was to develop a fast and effective detection and typing platform for mastitis-causing bacteria within this family. Therefore, the specific objectives of this work were the following:

1) To find molecular markers with specificity to the family, genus and species of mastitis-causing pathogens from the Streptococcaceae family, using bioinformatics tools.

2) To find additional markers of strain-specific genes, capable of discriminating particular traits and the overall diversity of the most relevant bovine mastitis-causing pathogens.

3) To validate the selected markers using PCR-based techniques and dot blot hybridisation assays.

# Materials and methods

## 1. Selection of DNA signatures and *in silico* analyses

Identification of taxa-specific DNA signatures was carried out using the Insignia online application (Phillippy *et al.*, 2007), taking into account only chromosomal data. For each target taxon, the 10 largest signatures found were analysed for their specificity using the BLAST (blastn) utility (Altschul *et al.*, 1990), and the most promising regions were selected for further analyses. One broad range signature for the Streptococcaceae family was selected (Ins1), in addition to specific regions for *Lactococcus* (Ins2), *Streptococcus* (Ins3), *S. agalactiae* (Ins4) and *S. uberis* (Ins5).

For the selection of the Streptococcaceae specific region (Ins1), a total of 65 sequenced genomes were compared and analysed in Insignia (Table 1). To retrieve a *Lactococcus*-specific region (Ins2), the two available genomes were used: *Lactococcus lactis* subsp. *cremoris* SK11 and *Lactococcus lactis* subsp. *cremoris* MG1363. For the *Streptococcus* signature (Ins3), several genus members' full genome sequences were analysed (Table 2). The *S. agalactiae* signature (Ins4) was chosen based on the eight specific available genomes (Table 3). The *S. uberis*-specific region (Ins5) was selected using *S. uberis* 0140J, the only representative with a sequenced genome.

One additional signature (Pf1) specific to the *Streptococcus* genus was selected using the Protein Family Database (Pfam) (Finn *et al.*, 2006). Using the "Taxonomy search" function of the database, the protein domains exclusive to *Streptococcus* were filtered. The corresponding DNA sequences were obtained and specificity was verified using the BLAST (blastn) utility (Altschul *et al.*, 1990).

Overall, the bioinformatics analyses carried out, using Insignia and Pfam, allowed the selection of six DNA signatures specific to the target pathogens (Table 4).

**Table 1**. Sequenced genomes used in Insignia for the selection of a Streptococcaceae-specific region (Ins1).

| **Sequenced genomes** | |
|---|---|
| *Streptococcus gordonii* str. *Challis* substr. CH1 *Challis* | *Streptococcus pneumoniae* CDC3059-06 |
| *Streptococcus salivarius* SK126 | *Streptococcus pneumoniae* MLV-016 |
| *Streptococcus sanguinis* SK36 | *Streptococcus pneumoniae* Taiwan19F-14 |
| *Streptococcus suis* 89/1591 | *Streptococcus pneumoniae* Hungary19A-6 |
| *Streptococcus suis* 05ZYH33 | *Streptococcus pneumoniae* 70585 |
| *Streptococcus suis* 98HAH33 | *Streptococcus pneumoniae* JJA |
| *Streptococcus thermophilus* LMG 18311 | *Streptococcus pneumoniae* P1031 |
| *Streptococcus thermophilus* CNRZ1066 | *Streptococcus pneumoniae* G54 |
| *Streptococcus thermophilus* LMD-9 | *Streptococcus pneumoniae* CGSP14 |
| *Streptococcus mutans* UA159 | *Streptococcus pneumoniae* TCH8431/19A |
| *Streptococcus agalactiae* 2603V/R | *Streptococcus pneumoniae* ATCC 700669 |
| *Streptococcus agalactiae* NEM316 | *Streptococcus pneumoniae* CCRI 1974M2 |
| *Streptococcus agalactiae* 18RS21 | *Streptococcus pyogenes* M1 GAS SF370 |
| *Streptococcus agalactiae* 515 | *Streptococcus pyogenes* MGAS5005 |
| *Streptococcus agalactiae* H36B | *Streptococcus pyogenes* SSI-1 |
| *Streptococcus agalactiae* COH1 | *Streptococcus pyogenes* MGAS315 |
| *Streptococcus agalactiae* CJB111 | *Streptococcus pyogenes* str. Manfredo |
| *Streptococcus agalactiae* A909 | *Streptococcus pyogenes* MGAS10394 |
| *Streptococcus pneumoniae* TIGR4 | *Streptococcus pyogenes* MGAS8232 |
| *Streptococcus pneumoniae* R6 | *Streptococcus pyogenes* M49 591 |
| *Streptococcus pneumoniae* D39 | *Streptococcus pyogenes* ATCC BAA-1633 |
| *Streptococcus pneumoniae* SP3-BS71 | *Streptococcus pyogenes* MGAS6180 |
| *Streptococcus pneumoniae* SP6-BS73 | *Streptococcus pyogenes* MGAS9429 |
| *Streptococcus pneumoniae* SP9-BS68 | *Streptococcus pyogenes* MGAS2096 |
| *Streptococcus pneumoniae* SP11-BS70 | *Streptococcus pyogenes* MGAS10270 |
| *Streptococcus pneumoniae* SP14-BS69 | *Streptococcus pyogenes* MGAS10750 |
| *Streptococcus pneumoniae* SP18-BS74 | *Streptococcus uberis* 0140J |
| *Streptococcus pneumoniae* SP19-BS75 | *Streptococcus infantarius* subsp. *infantarius* ATCC BAA-102 |
| *Streptococcus pneumoniae* SP23-BS72 | *Streptococcus equi* subsp. *zooepidemicus* ATCC BAA1716 |
| *Streptococcus pneumoniae* CDC1087-00 | *Streptococcus equi* subsp. *equi* 4047 |
| *Streptococcus pneumoniae* CDC1873-00 | *Lactococcus lactis* subsp. *cremoris* SK11 |
| *Streptococcus pneumoniae* SP195 | *Lactococcus lactis* subsp. *cremoris* MG1363 |
| *Streptococcus pneumoniae* CDC0288-04 | |

**Table 2.** Sequenced genomes used in Insignia for the selection of a *Streptococcus*-specific signature (Ins3).

| **Sequenced genomes** | |
|---|---|
| *Streptococcus gordonii* str. *Challis* substr. CH1 *Challis* | *Streptococcus pneumoniae* CDC0288-04 |
| *Streptococcus salivarius* SK126 | *Streptococcus pneumoniae* CDC3059-06 |
| *Streptococcus sanguinis* SK36 | *Streptococcus pneumoniae* MLV-016 |
| *Streptococcus suis* 89/1591 | *Streptococcus pneumoniae* Taiwan19F-14 |
| *Streptococcus suis* 05ZYH33 | *Streptococcus pneumoniae* Hungary19A-6 |
| *Streptococcus suis* 98HAH33 | *Streptococcus pneumoniae* 70585 |
| *Streptococcus thermophilus* LMG 18311 | *Streptococcus pneumoniae* JJA |
| *Streptococcus thermophilus* CNRZ1066 | *Streptococcus pneumoniae* P1031 |
| *Streptococcus thermophilus* LMD-9 | *Streptococcus pneumoniae* G54 |
| *Streptococcus mutans* UA159 | *Streptococcus pneumoniae* CGSP14 |
| *Streptococcus agalactiae* 2603V/R | *Streptococcus pneumoniae* TCH8431/19A |
| *Streptococcus agalactiae* NEM316 | *Streptococcus pneumoniae* ATCC 700669 |
| *Streptococcus agalactiae* 18RS21 | *Streptococcus pneumoniae* CCRI 1974M2 |
| *Streptococcus agalactiae* 515 | *Streptococcus pyogenes* M1 GAS SF370 |
| *Streptococcus agalactiae* H36B | *Streptococcus pyogenes* MGAS5005 |
| *Streptococcus agalactiae* COH1 | *Streptococcus pyogenes* SSI-1 |
| *Streptococcus agalactiae* CJB111 | *Streptococcus pyogenes* MGAS315 |
| *Streptococcus agalactiae* A909 | *Streptococcus pyogenes* str. Manfredo |
| *Streptococcus pneumoniae* TIGR4 | *Streptococcus pyogenes* MGAS10394 |
| *Streptococcus pneumoniae* R6 | *Streptococcus pyogenes* MGAS8232 |
| *Streptococcus pneumoniae* D39 | *Streptococcus pyogenes* M49 591 |
| *Streptococcus pneumoniae* SP3-BS71 | *Streptococcus pyogenes* ATCC BAA-1633 |
| *Streptococcus pneumoniae* SP6-BS73 | *Streptococcus pyogenes* MGAS6180 |
| *Streptococcus pneumoniae* SP9-BS68 | *Streptococcus pyogenes* MGAS9429 |
| *Streptococcus pneumoniae* SP11-BS70 | *Streptococcus pyogenes* MGAS2096 |
| *Streptococcus pneumoniae* SP14-BS69 | *Streptococcus pyogenes* MGAS10270 |
| *Streptococcus pneumoniae* SP18-BS74 | *Streptococcus pyogenes* MGAS10750 |
| *Streptococcus pneumoniae* SP19-BS75 | *Streptococcus uberis* 0140J |
| *Streptococcus pneumoniae* SP23-BS72 | *Streptococcus infantarius* subsp. *infantarius* ATCC BAA-102 |
| *Streptococcus pneumoniae* CDC1087-00 | *Streptococcus equi* subsp. *zooepidemicus* ATCC BAA1716 |
| *Streptococcus pneumoniae* CDC1873-00 | *Streptococcus equi* subsp. *equi* 4047 |
| *Streptococcus pneumoniae* SP195 | |

**Table 3**. Sequenced genomes used in the Insignia bioinformatics tool for the selection of a *S. agalactiae* specific signature (Ins4).

| **Sequenced genomes** | |
|---|---|
| *Streptococcus agalactiae* 2603V/R | *Streptococcus agalactiae* H36B |
| *Streptococcus agalactiae* NEM316 | *Streptococcus agalactiae* COH1 |
| *Streptococcus agalactiae* 18RS21 | *Streptococcus agalactiae* CJB111 |
| *Streptococcus agalactiae* 515 | *Streptococcus agalactiae* A909 |

Further *in silico* analyses were performed on the obtained DNA signatures (Table 4). The circular chromosome map was visualised using Geneious Pro (Drummond *et al.*, 2012), and the position of each marker was determined in *Streptococcus uberis* 0140J for signatures Ins1, Ins3 and Ins5; in *Lactococcus lactis* subsp. *lactis* CV56 for Ins2; and in *Streptococcus agalactiae* 2603V/R for the Ins4 region. The Codon Adaptation Index (CAI), the expected CAI and GC content were calculated using the CAIcal server (Puigbo *et al.*, 2008).

**Table 4**. Selected DNA signatures, using Insignia and Pfam.

| **Signature** | **Specificity** | **Source** |
|---|---|---|
| Ins1 | Streptococcaceae | Insignia |
| Ins2 | *Lactococcus* | Insignia |
| Ins3 | *Streptococcus* | Insignia |
| Ins4 | *S. agalactiae* | Insignia |
| Ins5 | *S. uberis* | Insignia |
| Pf1 | *Streptococcus* | Pfam |

## 2. DNA markers design

The selection of DNA markers was based on the previously obtained signatures and carried out using the AlignX utility from the Vector NTI software (Invitrogen, Carlsbad, CA). Primer pairs (Table 5) were designed for each of the six selected regions, using the Vector NTI software (Invitrogen, Carlsbad, CA). All primer pairs were chosen with a predicted amplicon size of 150 to 500 bp and a calculated optimal annealing temperature greater than 50 ºC.

One marker specific to the Streptococcaceae family was selected taking into account the alignment of 15 target sequences (F1). Two markers from the *Streptococcus*-specific signature (Ins3) were obtained, based on nine available genomes (ST1 and ST2). One *Lactococcus*-specific marker (LC2) was selected using the alignment of six available sequences. One marker exclusive to *S. uberis* (SU) was chosen based solely on the Ins3 signature sequence of *S. uberis* 0140J. Two markers were obtained from a

*S. agalactiae*-specific signature (Ins4), using five target genomes (A1 and A2). One additional marker (ST3) was selected based on the *Streptococcus* signature obtained with Pfam (Pf1), using seven target sequences.

In addition, three primer pairs were designed for genes associated with the fructose operon of *S. agalactiae*: one for a transcriptional regulator gene (FO1), one for a fructose-1-phosphate kinase (FO2) and another for a phosphotransferase system (PTS) fructose-specific component (FO3). Two primer pairs were also designed for the nisin U operon of *S. uberis*: one for the gene responsible for the operon regulation (NU1), and another for the gene responsible for immunity to nisin U (NU3).

Primer pairs were designed having as template the sequence of *Streptococcus uberis* 0140J for primers F1 FWD/REV, ST1 FWD/REV, ST2 FWD/REV and SU FWD/REV; of *Lactococcus lactis* subsp. *lactis* CV56 for primers LC2 FWD/REV; of *Streptococcus agalactiae* 2603V/R for primers ST3 FWD/REV, A1 FWD/REV, A2 FWD/REV; of *Streptococcus agalactiae* FSL S3-026 for FO1 FWD/REV, FO2 FWD/REV and FO3 FWD/REV; and of *Streptococcus uberis* strain 42 for primers NU1 FWD/REV and NU3 FWD/REV. Primers were designed to anneal to the sites of each marker that showed higher specificity to the selected targets.

Three additional primer pairs were selected based on virulence-associated genes described in *S. uberis* (Ward *et al.*, 2001; Smith *et al.*, 2002; Reinoso *et al.*, 2011): the hyaluronic acid operon gene *hasC* (V1), the glyceraldehyde 3-phosphate dehydrogenase gene *gapC* (V2) and the oligopeptide permease gene *oppF* (V3).

A total of 16 taxa-specific and functional markers were obtained for experimental validation (Tables 5 and 6).

**Table 5.** Selected taxa-specific markers with primer sequence, annealing temperature (Ta), amplicon length and marker specificity.

| Signature | Marker | | Primer Sequence (5' - 3') | $T_a$ | Amplicon length | Specificity | Source |
|---|---|---|---|---|---|---|---|
| **Ins1** | **F1** | FWD | TTATGCTCGTCTTGCTCTTTACGG | 54.6 ºC | 281 bp | Streptococcaceae | Insignia |
| | | REV | GCACACGTCCAAGTGATGTAGCTG | | | | |
| **Ins2** | **LC2** | FWD | TTTATGATTCAAAATTTAACCGCT | 51.8 ºC | 251 bp | *Lactococcus* | Insignia |
| | | REV | TGAATGCCGTATGCTCTTCC | | | | |
| **Ins3** | **ST1** | FWD | TCCAGTTATGGTGACGCAATATGAT | 53.4 ºC | 333 bp | *Streptococcus* | Insignia |
| | | REV | GCTAAACTAGTATTCGGATGGGCTG | | | | |
| | **ST2** | FWD | CATTGGGAAAAGAGTCAGTGTTAG | 51.1 ºC | 194 bp | *Streptococcus* | Insignia |
| | | REV | TGATTCTGGCAATTTCTGTATAAG | | | | |
| **Pf1** | **ST3** | FWD | GTTATGGATGGCTCCTGGAT | 50.7 ºC | 265 bp | *Streptococcus* | Pfam |
| | | REV | TCCCTAGTCTTAGATAGAACCGTTA | | | | |
| **Ins4** | **A1** | FWD | ATGTAGCTGCTGATTCTGTCATAA | 52.6 ºC | 314 bp | *S. agalactiae* | Insignia |
| | | REV | AATAGCTGGTGTAGATTTGACTGC | | | | |
| | **A2** | FWD | ATGAACACAAAACAGCGTTTTTCA | 50.8 ºC | 192 bp | *S. agalactiae* and *S. dysgalactiae* | Insignia |
| | | REV | AGTAGGTGTCTCATTTGCTATGCT | | | | |
| **Ins5** | **SU** | FWD | TCGTTTGTATACGCTTGATGCT | 50.6 ºC | 229 bp | *S. uberis* | Insignia |
| | | REV | CACGTCTCTATAAAAGGAATTCCC | | | | |

**Table 6.** Selected functional markers with primer sequence, annealing temperature (Ta), amplicon length and target gene.

| Functional traits | Marker | | Primer Sequence (5' - 3') | $T_a$ | Amplicon length | Target gene | Source |
|---|---|---|---|---|---|---|---|
| **Fructose Operon** | **FO1** | FWD | CGACATCAAAAAAACAACTAACAC | 50.6 ºC | 331 bp | Regulation gene (*fruR*) | Richards. *et al*. (2011) |
| | | REV | TCCACCACGTTATTGAGAGTTT | | | | |
| | **FO2** | FWD | CCGAGTCACTTATGAGTAAACAGCC | 51.3 ºC | 279 bp | Kinase gene (*fruP*) | Richards. *et al*. (2011) |
| | | REV | GGGGGATCTCCACAGAAATTTTTT | | | | |
| | **FO3** | FWD | TCTCAATTTCTTCGATCTCATGTGC | 52.6 ºC | 348 bp | PTS component gene (*fruD*) | Richards. *et al*. (2011) |
| | | REV | CAGGTCTTGTTGTCGAAAACGATTA | | | | |
| **Nisin Operon** | **NU1** | FWD | CCAAGGTTGCAGCGCATTT | 51.5 ºC | 331 bp | Regulation gene (*nsuR*) | Richards. *et al*. (2011) |
| | | REV | CCCCTTATTGTCTTGATGGGATT | | | | |
| | **NU3** | FWD | AATCAAATCGTTGATGAAAATGACC | 50.6 ºC | 502 bp | Nisin immunity gene (*nsuI*) | Richards. *et al*. (2011) |
| | | REV | AAACTTCTCCGTAATCCCAAACTTC | | | | |
| **Virulence-associated genes** | **V1** | FWD | TGCTTGGTGACGATTTGATG | 58.0 ºC | 300 bp | Hyaluronic acid operon gene (*hasC*) | Ward. *et al.* (2001) |
| | | REV | GTCCAATGATAGCAAGGTACAC | | | | |
| | **V2** | FWD | GCTCCTGGTGGAGATGATGT | 55.0 ºC | 189 bp | Glyceraldehyde 3-phosphate dehydrogenase gene (*gapC*) | Reinoso. *et al*. (2011) |
| | | REV | GTCACCAGTGTAAGCGTGGA | | | | |
| | **V3** | FWD | GGCCTAACCAAAACGAAACA | 54.0 ºC | 419 bp | Oligopeptide permease gene (*oppF*) | Smith. *et al*. (2002) |
| | | REV | GGCTCTGGAATTGCTGAAAG | | | | |

## 3. Bacterial strains, culture conditions and DNA extraction

A total of 50 bacterial strains were used in this work, corresponding to 15 reference strains and 35 isolates representative of the Streptococcaceae family, of closely related species and of organisms with common habitats (Table 7). The bacterial isolates, obtained from different mastitic milk samples within Portugal, were provided by SEGALAB (Laboratório de Sanidade Animal e Segurança Alimentar, S.A.). Species identification was done using the VITEK 2 system (bioMérieux, Durham, NC).

All strains were cultured in Brain Heart Infusion (BHI) medium (Oxoid, Hampshire, England) at 37 $^{o}$C, with the exception of *Lactovum miscens* DSM 14925, which was cultured in MRS broth medium (Oxoid, Hampshire, England) pre-reduced with cysteine 0.05% and supplemented with N-acetylglucosamine 2 mM at 25 $^{o}$C, in anaerobic conditions, using the CampyGen Atmosphere Generation System (Oxoid, Hampshire, England).

DNA was extracted from pure bacterial cultures using the EaZy Nucleic Acid (E.Z.N.A.) bacterial DNA purification kit (Omega Bio-Tek, Norcross, GA), following the manufacturer's instructions. The quality of the extracted DNA was assessed by electrophoresis in 1% agarose gel and each DNA sample was quantified using the Qubit 2.0 Fluorometer HS Assay (Invitrogen, Carlsbad, CA), according to the manufacturer's instructions.

## 4. PCR

The PCR mastermix was prepared with 1x DreamTaq buffer, containing 1.5 mM of $MgCl_2$ (Fermentas, Ontario, Canada), 0.2 mM of each deoxynucleoside triphosphate (dNTP; Fermentas), 0.2 µM of each primer, 1 U of DreamTaq DNA polymerase (Fermentas) and ≈ 25 ng of genomic DNA as template for a total reaction volume of 20 µl. The PCR conditions were as follows: an initial denaturation step of 5 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 45 s at 55 °C, and 45 s at 72 °C, with a final extension step of 10 min at 72 °C.

Amplification of the 16S rRNA gene (primers listed in Table 8) was performed as mentioned above with 35 cycles of 30 s denaturing at 95 ºC, 30 s annealing at 55 ºC and 90 s extension at 72 ºC .

**Table 7.** Bacterial strains used in this study.

| Species | Abbreviation | Location | Source |
|---|---|---|---|
| *Enterobacter aerogenes* LMG 2094 | *Eae* 2094 | - | LMG |
| *Enterococcus faecalis* LMG 7937 | *Efc* 7937 | - | LMG |
| *Enterococcus faecium* | *Efa* EF7 | Trofa | SEGALAB |
| *Enterococcus faecium* LMG 11397 | *Efa* 11397 | - | LMG |
| *Enterococcus faecium* LMG 11423 | *Efa* 11423 | - | LMG |
| *Klebisella pneumoniae* subsp. *pneumoniae* LMG 2095 | *Kpp* 2095 | - | LMG |
| *Lactococcus lactis* subsp. *lactis* LMG 6890 | *Lll* 6890 | - | LMG |
| *Lactovum miscens* DSM 14925 | *Lmi* 14925 | - | DSMZ |
| *Proteus mirabilis* LMG 3257 | *Pmi* 3257 | - | LMG |
| *Staphylococcus aureus* | *Sau* SA1 | - | SEGALAB |
| *Staphylococcus aureus* | *Sau* SA2 | - | SEGALAB |
| *Staphylococcus aureus* LMG 8224 | *Sau* 8224 | - | LMG |
| *Staphylococcus haemolyticus* LMG 13349 | *Sha* 13349 | - | LMG |
| *Staphylococcus pasteuri* | *Spa* SP1 | Vila do Conde | SEGALAB |
| *Staphylococcus pasteuri* | *Spa* SP2 | Vila do Conde | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA2 | Vila do Conde | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA3 | Vila do Conde | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA4 | Vila do Conde | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA6 | Vila do Conde | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA7 | Vila do Conde | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA8 | Vila do Conde | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA9 | Vila do Conde | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA10 | Vila do Conde | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA11 | Trofa | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA21 | Barcelos | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA25 | Póvoa de Varzim | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA28 | Póvoa de Varzim | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA29 | Póvoa de Varzim | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA30 | Póvoa de Varzim | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA31 | Póvoa de Varzim | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA32 | Barcelos | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA33 | Barcelos | SEGALAB |
| *Streptococcus agalactiae* | *Sag* SA34 | Barcelos | SEGALAB |
| *Streptococcus agalactiae* LMG 15083 | *Sag* 15083 | - | LMG |
| *Streptococcus bovis* LMG 8518 | *Sbo* 8518 | - | LMG |
| *Streptococcus dysgalactiae* | *Sdy* SD1 | - | SEGALAB |
| *Streptococcus dysgalactiae* | *Sdy* SS4 | Póvoa de Varzim | SEGALAB |
| *Streptococcus parauberis* LMG 12174 | *Spu* 12174 | - | LMG |
| *Streptococcus salivarius* | *Ssa* 112 | - | SEGALAB |
| *Streptococcus uberis* | *Sub* SU1 | Barcelos | SEGALAB |
| *Streptococcus uberis* | *Sub* SU2 | Barcelos | SEGALAB |
| *Streptococcus uberis* | *Sub* SU3 | Barcelos | SEGALAB |
| *Streptococcus uberis* | *Sub* SU4 | Paços de Ferreira | SEGALAB |
| *Streptococcus uberis* | *Sub* SU5 | Matosinhos | SEGALAB |
| *Streptococcus uberis* | *Sub* SU6 | Matosinhos | SEGALAB |
| *Streptococcus uberis* | *Sub* SU7 | Matosinhos | SEGALAB |
| *Streptococcus uberis* | *Sub* SU8 | Matosinhos | SEGALAB |
| *Streptococcus uberis* | *Sub* SU9 | Matosinhos | SEGALAB |
| *Streptococcus uberis* LMG 9465 | *Sub* 9465 | - | LMG |
| *Vagococcus fluvialis* LMG 12318 | *Vfl* 12318 | - | LMG |

LMG - Belgian Co-ordinated collections of microorganisms, Gent, Belgium.
DSMZ - German collection of microorganisms and cell cultures, Braunschweig, Germany.

**Table 8.** Primers used for the amplification of the 16S rRNA gene. M = A or C; Y = C or T.

| Gene | | Primer Sequence (5' - 3') | Source |
|---|---|---|---|
| 16S rRNA | FWD | AGAGTTTGATCMTGGCTCAG | Weisburg, *et al*. (1991) |
| | REV | TACGGYTACCTTGTTACGACTT | |

## 5. Dot blot validation and automatic image analysis

DNA probes were obtained from purified PCR amplicons, using the GFX PCR DNA and gel band purification kit (GE Healthcare, Buckinghamshire, United Kingdom). The identity of these amplicons was confirmed by sequencing (STAB Vida, Portugal).

Markers F1, ST1, ST2, SU, V1, V2 and V3 were obtained from *Streptococcus uberis* LMG 9465 (*Sub* 9465); marker LC2 was obtained from *Lactococcus lactis* subsp. *lactis* LMG 6890 (*Lll* 6890); markers ST3, A1 and A2 were obtained from *Streptococcus agalactiae* LMG 15083 (*Sag* 15083); markers FO1, FO2 and FO3 were obtained from *Streptococcus agalactiae* (*Sag* SA11) and markers NU1 and NU3 were obtained from *Streptococcus uberis* (*Sub* SU3).

Identification of *Efa* EF7, *Spa* SP1, *Spa* SP2 and of the isolates used as marker template (*Sag* SA11 and *Sub* SU3) was confirmed by sequencing of the 16S rRNA gene (Fig. S1-S4 of Supplementary data). The sequences obtained of each isolate were analysed by BLAST (Altschul *et al.*, 1990) and on the Ribosomal Database Project (Cole *et al.*, 2009).

DNA probes were labelled with digoxigenin (DIG), using the DIG-High Prime labelling kit (Roche, Basel, Switzerland) according to the manufacturer's instructions. For dot blot assays, 100 ng of heat-denatured DNA from each bacterial strain were spotted into a nylon membrane using a Bio-Dot apparatus (Bio-Rad). Hybridisation was carried out overnight at 68 °C, with a final probe concentration of 100 ng/ml. Washing and detection steps were performed according to the DIG system recommendations (Roche). DIG-labelled nucleic acids were detected by chemiluminescence using X-ray films (GE Healthcare) and a Molecular Imager Chemi-Doc system (Bio-Rad).

The analysis of the hybridisation data was complemented with an algorithm developed to automatically process and interpret the dot blot images. This software adjusts each image to a user-defined grid and outputs a probability value of each dot being a

positive signal, using as references the positive and negative controls present in each membrane (Marçal *et al.*, 2009; Caridade *et al.*, 2010).

# Results

## 1. Selection of DNA signatures and *in silico* analyses

Numerous bacterial species belonging to Streptococcaceae have been associated to bovine mastitis (Wyder *et al.*, 2011). In this work, six DNA signatures specific to different genera and species within this family were obtained.

Five taxonomic regions were selected using Insignia (Phillippy *et al.*, 2007). This utility calculates 20 mer DNA signatures specific to the selected target organisms. The 10 largest signatures obtained for each taxon were analysed for specificity using BLAST (Altschul *et al.*, 1990) and one region was selected for further experimental validation, taking into account the specificity results (i.e. the signature with the lowest E value).

For the selection of one Streptococcaceae-specific signature, a total of 3711 signatures were obtained and ordered by sequence length. The largest signature specific to Streptococcaceae was selected, corresponding to a 300 bp region (Ins1). Other regions exclusive to *Lactococcus* and *Streptococcus* were retrieved: a 282 bp signature was selected for *Lactococcus* (Ins2), the largest and most specific region out of 20288 computed sequences, while a specific 840 bp signature (Ins3) was filtered from 2775 sequences, found particularly unique to mastitis-causing streptococci. Furthermore, two additional signatures were obtained for prevalent mastitis pathogens. A 444 bp signature specific to *S. agalactiae* (Ins4) and a 400 bp *S. uberis*-specific sequence (Ins5), the largest out of 71208 and out of 108723 signatures obtained, respectively.

The Protein Family Database was used to obtain one additional *Streptococcus*-specific signature (Finn *et al.*, 2006). Three protein domains exclusive to *Streptococcus* were obtained. The primary structure of the proteins was analysed and two multiple copy domains were filtered out. The remaining region of 195 bp (Pf1) was subjected to a specificity analysis using BLAST, which confirmed its specificity towards the *Streptococcus* genus, particularly *S. agalactiae*.

Therefore, six DNA signatures specific to taxonomic groups within Streptococcaceae were selected (Table 9).

**Table 9**: Selected DNA signatures, using Insignia and Pfam.

| Signature | Source | Specificity | Signature length |
|---|---|---|---|
| Ins1 | Insignia | Streptococcaceae | 300 bp |
| Ins2 | Insignia | *Lactococcus* | 282 bp |
| Ins3 | Insignia | *Streptococcus* | 840 bp |
| Ins4 | Insignia | *S. agalactiae* | 444 bp |
| Ins5 | Insignia | *S. uberis* | 400 bp |
| Pf1 | Pfam | *Streptococcus* | 195 bp |

To further assess the genomic properties of the DNA regions obtained, their chromosomal location, GC content and Codon Adaptation Index (CAI) with expected values (eCAI) were calculated (Fig. 3-5). Signature Ins2 was obtained from an intergenic region, so the Codon Adaptation Index was not calculated. Signatures Ins1 and Ins4 appeared to have a codon usage adapted to the genome, with a normalised Codon Adaptation Index ratio (CAI/eCAI) closest to 1 (i.e. close to the value usually obtained for housekeeping genes). Concerning Ins3, Ins5 and Pf1, the data obtained suggested a biased codon usage. The GC content calculated presented significant differences to the overall GC content of the reference genomes used, suggesting the possible occurrence of horizontal gene transfer events, although some GC content variation is to be expected along the chromosome.

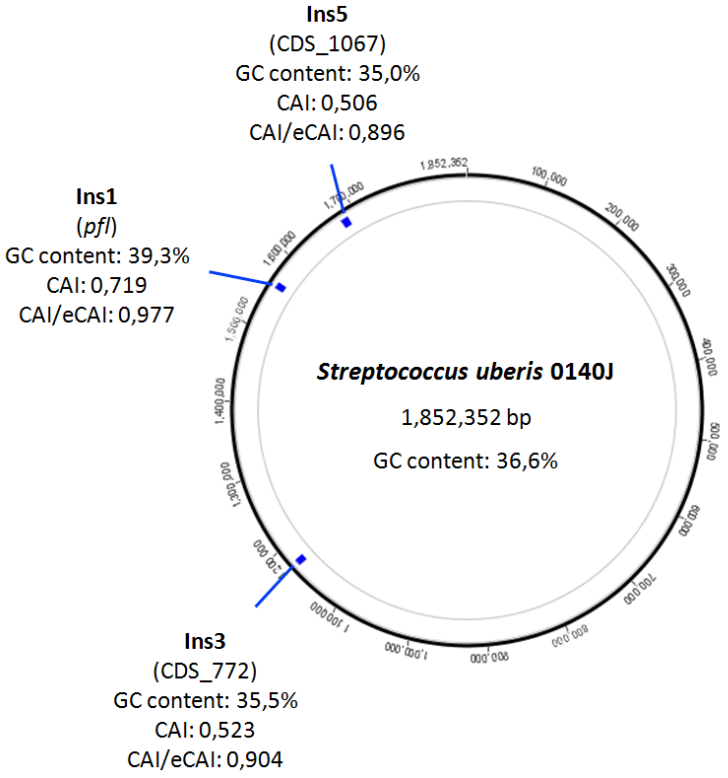


**Figure 3.** Genome map of *Streptococcus uberis* 0140J with genome coordinates (bp) and selected DNA signatures (blue). The Coding DNA Sequence (CDS) annotation, GC content, Codon Adaptation Index (CAI) and normalised CAI (CAI/eCAI) values are shown for each region.

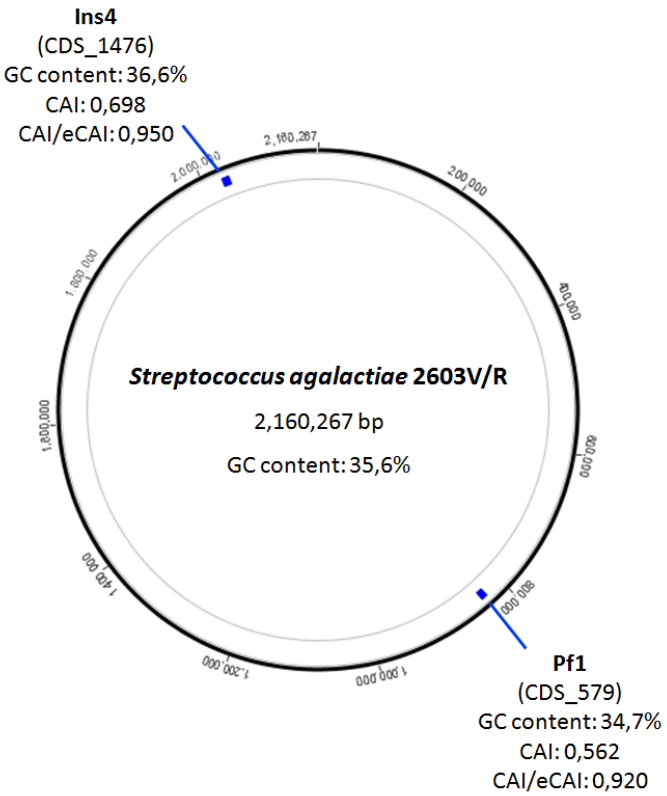


**Figure 4.** Genome map of *Streptococcus agalactiae* 2603V/R with genome coordinates (bp) and selected DNA signatures (blue). The Coding DNA Sequence (CDS) annotation, GC content, Codon Adaptation Index (CAI) and normalised CAI (CAI/eCAI) values are shown for each region.

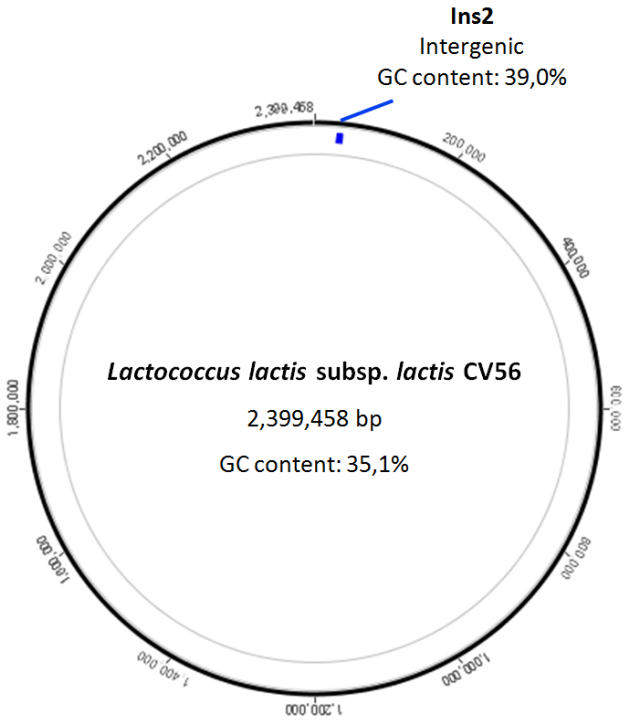


**Figure 5.** Genome map of *Lactococcus lactis* subsp. *lactis* CV56 with genome coordinates (bp) and the selected DNA signature (blue). The Coding DNA Sequence (CDS) annotation, GC content, Codon Adaptation Index (CAI) and normalised CAI (CAI/eCAI) values are shown for this region.

## 2. DNA markers design

Based on the genomic regions obtained using Insignia and Pfam, taxa-specific DNA markers were retrieved. Target sequences were analysed with AlignX and primer pairs were designed with the Vector NTI software (Invitrogen, Carlsbad, CA).

From a 300 bp signature (Ins1), a 281 bp marker was obtained, which showed to be conserved among Streptococcaceae organisms (F1 - Fig. S21 of Supplementary data). A 251 bp marker (LC2 - Fig. S22 of Supplementary data) was obtained from a 282 bp *Lactococcus*-specific signature (Ins2). From a 840 bp *Streptococcus*-specific signature (Ins3), two markers were selected (Fig. S23 of Supplementary data): one 333 bp marker particularly specific to important mastitis pathogens (ST1), namely *S. agalactiae*, *S. dysgalactiae* and *S. uberis*, and one 194 bp marker transversal across the *Streptococcus* genus (ST2). From a 444 bp *S. agalactiae*-specific signature (Ins4) two DNA markers were obtained (Fig. S24 of Supplementary data): a 314 bp marker exclusive to *S. agalactiae* (A1) and a 192 bp marker which also showed to be specific to *S. dysgalactiae* (A2). A 229 bp *S. uberis*-specific marker was obtained from a 400 bp signature (Ins5).

Due to the small size of the *Streptococcus*-specific signature (195 bp) obtained by Pfam (Pf1), the sequence length was increased to the flanking regions, using *S. agalactiae* 2603V/R as a reference organism, until specific primer pairs were able to be designed. A final 265 bp *Streptococcus*-specific marker was selected (Fig. S25 of Supplementary data).

To increase the discriminatory resolution and gain insight into strain-specific patterns and the pathogenic potential of Streptococcaceae organisms, additional functional markers were selected from phenotypic traits associated with virulent strains responsible for bovine mastitis.

Recent studies have identified a particular evolution and adaptation of *S. agalactiae* strains to bovine hosts, due to horizontal gene transfer events (Richards *et al.*, 2011). A number of genomic islands have been described exclusively or more significantly present in bovine isolates of *S. agalactiae* that present a competitive advantage during infection of bovine hosts. One of these regions is located in a four-gene operon responsible for fructose utilisation, with high sequence similarity with *S. dysgalactiae*. Therefore, markers were designed for three different genes of this operon: a 331 bp

marker (FO1) from a 756 bp transcriptional regulator gene (*fruR*), a 279 bp marker (FO2) from a 933 bp gene coding for a fructose-1-phosphate kinase (*fruC*) and a 348 bp marker (FO3) from a 447 bp gene coding for a phosphotransferase system (PTS) fructose-specific component (*fruD*) (Fig. 6).

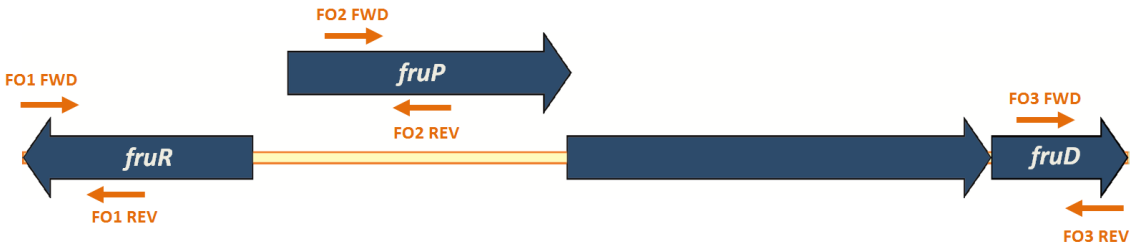


**Figure 6**. Primer design of three markers from the fructose operon, based on the *fruR* gene coding for an operon regulator (FO1), the *fruP* gene coding for a kinase (FO2) and the *fruD* gene coding for a PTS system component of this operon (FO3).

Furthermore, an 11-gene operon involved in the production of nisin, a lantibiotic previously described in *L. lactis*, has been identified in *S. uberis* strains predominating in intramammary infections (Richards *et al.*, 2011). This suggests that this bacteriocin confers a competitive advantage and resistance to these particular strains, although no further knowledge exists on its potential role during infection. A number of genes from *Streptococcus* and *Lactococcus* species, responsible for bovine mastitis, have been found with high sequence similarity to this operon, which suggests that a shared environment could have resulted in lateral gene transfer between these species. Two markers were selected from the nisin U operon of *S. uberis*: a 331 bp marker (NU1) from a 699 bp signature of the operon regulator gene (*nsuR*) and a 502 bp marker (NU3) from a 717 bp gene responsible for strain resistance to nisin U (*nsuI*) (Fig. 7).

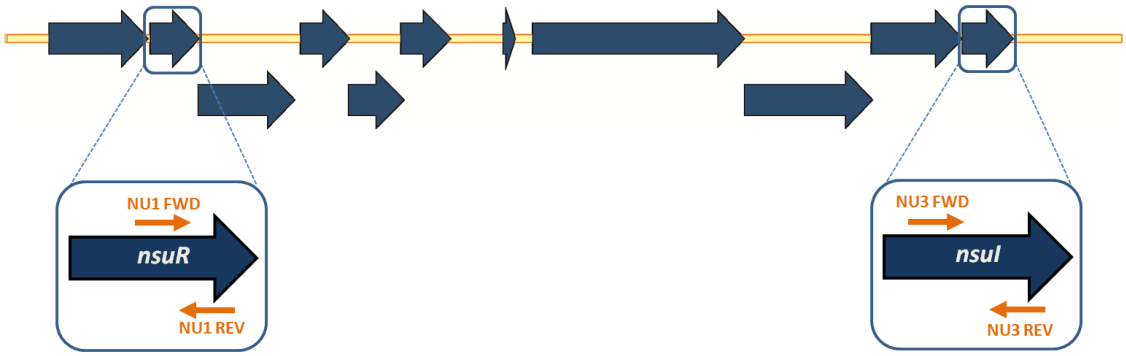


**Figure 7**. Primer design of two markers from the nisin operon, obtained from the regulation gene *nsuR* (NU1) and the *nsuI* gene responsible for immunity to nisin (NU3).

From previous studies (Ward *et al.*, 2001; Smith *et al.*, 2002; Reinoso *et al.*, 2011), three virulence-associated markers linked to the presence of an especially prevalent virulent phenotype of *S. uberis* were selected: a 300 bp marker based on the *hasC* gene of the hyaluronic acid capsule operon (V1), a 189 bp marker from the *gapC* gene, coding for a glyceraldehyde 3-phosphate dehydrogenase responsible for oxidative stress resistance and binding to several host proteins (V2), and an additional marker of the oligopeptide permease gene *oppF*, involved in amino acid utilisation and transport (V3).

A total of 16 DNA markers were designed and selected for experimental validation (Tables 10 and 11).

**Table 10.** Selected taxa-specific markers with target taxon (specificity), probe size (amplicon length) and the BLAST specificity of the first unrelated organism (amplicon best BLAST hit).

| Signature | Marker | Specificity | Amplicon length | Amplicon best BLAST hit (E value / Query coverage) | Source |
|---|---|---|---|---|---|
| **Ins1** | **F1** | Streptococcaceae | 281 bp | *Staphylococcus aureus* subsp. *aureus* 11819-97<br>2e-10 / 81% | Insignia |
| **Ins2** | **LC2** | *Lactococcus* | 251 bp | *Homo sapien*s FOSMID clone ABC12-46373500B5<br>0.094 / 13% | Insignia |
| **Ins3** | **ST1** | *Streptococcus* | 333 bp | *Streptococcus pneumoniae* 670-6B<br>0,00007 / 34% | Insignia |
| | **ST2** | *Streptococcus* | 194 bp | *Lactococcus garvieae* Lg2<br>0,0000009 / 52% | Insignia |
| **Pf1** | **ST3** | *Streptococcus* | 265 bp | *Lactococcus lactis* subsp. *lactis* IO-1<br>0,36 / 15% | Pfam |
| **Ins4** | **A1** | *S. agalactiae* | 314 bp | *Trichophyton verrucosum* HKI 0517<br>0,12 / 9% | Insignia |
| | **A2** | *S. agalactiae* and *S. dysgalactiae* | 192 bp | *Trichophyton verrucosum* HKI 0517<br>0,068 / 15% | Insignia |
| **Ins5** | **SU** | *S. uberis* | 229 bp | *Populus trichocarpa* clone POP037-L01<br>0,084 / 17% | Insignia |

**Table 11.** Selected functional markers with target gene and probe size (amplicon length).

| Functional trait | Marker | Target gene | Amplicon length | Source |
|---|---|---|---|---|
| **Fructose Operon** | **FO1** | Regulation gene (*fruR*) | 331 bp | Richards, *et al*. (2011) |
| | **FO2** | Kinase gene (*fruP*) | 279 bp | Richards, *et al*. (2011) |
| | **FO3** | PTS component gene (*fruD*) | 348 bp | Richards, *et al*. (2011) |
| **Nisin Operon** | **NU1** | Regulation gene (*nsuR*) | 331 bp | Richards, *et al*. (2011) |
| | **NU3** | Nisin immunity gene (*nsuI*) | 502 bp | Richards, *et al*. (2011) |
| **Virulence-associated genes** | **V1** | Hyaluronic acid operon gene (*hasC*) | 300 bp | Ward, *et al.* (2001) |
| | **V2** | Glyceraldehyde 3-phosphate dehydrogenase gene (*gapC*) | 189 bp | Reinoso, *et al*. (2011) |
| | **V3** | Oligopeptide permease gene (*oppF*) | 419 bp | Smith, *et al*. (2002) |

## 3. Dot blot validation

Preliminary dot blot results with the selected taxa-specific markers presented unexpected patterns for some *S. agalactiae* and *S. dysgalactiae* isolates. The identity of these strains, previously assessed using the VITEK 2 automated identification system, was confirmed by sequencing of the 16S rRNA gene, a more reliable approach for the identification of bovine mastitis pathogens (Bal *et al.*, 2010; Ajitkumar *et al.*, 2012). Comparative sequence analyses revealed that these isolates had been misidentified (Figure S1-S3 of Supplementary data). One isolate identified as *S. agalactiae* showed to be in fact an *Enterococcus*, most likely of the *E. faecium* species, whereas another isolate of *S. agalactiae*, in addition to one isolate of *S. dysgalactiae*, showed to belong to the *Staphylococcus* genus, probably to *S. pasteuris* species. Furthermore, a *S. uberis* isolate was incorrectly identified as *S. agalactiae.* These experimental data further emphasises the potential unreliability of culture-based approaches. In the frame of this new information, the annotation of these strains was rectified for the dot blot validation (Table 12).

**Table 12.** Four isolates provided by SEGALAB previously identified by the VITEK 2 system and their identification by sequence analysis of the 16S rRNA gene.

| Abbreviation (Table 7) | Biochemical identification (VITEK 2) | Genomic identification (16S rRNA gene sequencing) |
|---|---|---|
| *Ef* EF7 | *Streptococcus agalactiae* | *Enterococcus faecium* |
| *Spa* SP1 | *Streptococcus agalactiae* | *Staphylococcus pasteuris* |
| *Spa* SP2 | *Streptococcus dysgalactiae* | *Staphylococcus pasteuris* |
| *Sub* SU3 | *Streptococcus agalactiae* | *Streptococcus uberis* |

In order to validate the selected markers by dot blot hybridisation, a set of 44 reference strains and isolates was used, representative of the Streptococcaceae family, of closely related species and of organisms with matching hosts (Table 13): 30 strains belonging to the Streptococcaceae family (28 strains of *Streptococcus*, one strain of *Lactococcus* and one strain of *Lactovum*) and 14 strains from other taxa responsible for bovine mastitis or isolated from the bovine environment. Taxa-specific probes (F1, LC2, ST1, ST2, ST3, A1, A2 and SU) and functional markers (FO1, FO2, FO3, NU1, NU3, V1, V2 and V3) were tested with the entire set of bacterial strains used. Sequencing of the DNA probes confirmed the identity of the selected markers (Fig. S5-S20 of Supplementary data). For each membrane, duplicate samples were used to account for hybridisation inconsistencies.

**Table 13.** Layout of the membranes used in the dot blot hybridisation assays. Abbreviations are listed in Table 7. Species are highlighted according to the following colour scheme: green indicates strains belonging to *L. lactis*, brown to *L. miscens*, yellow to *S. agalactiae*, purple to *S. dysgalactiae,* blue to *S. uberis* and the red outline indicates organisms representative of the *Streptococcus* genus.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **A** | C+ | *Lll* 6890 | *Lmi* 14925 | *Sag* 15083 | *Sag* SA2 | *Sag* SA3 | *Sag* SA4 | *Sag* SA6 | *Sag* SA7 | *Sag* SA8 | *Sag* SA9 | C+ |
| **B** | *Sag* SA10 | *Sag* SA11 | *Sag* SA21 | *Sag* SA25 | *Sag* SA28 | *Sag* SA29 | *Sag* SA30 | *Sag* SA31 | *Sag* SA32 | *Sag* SA33 | *Sag* SA34 | *Sbo* 8518 |
| **C** | *Sdy* SD1 | *Sdy* SS4 | *Spu* 12174 | *Ssa* 112 | *Sub* 9465 | *Sub* SU1 | *Sub* SU2 | *Sub* SU3 | *Efc* 7937 | *Efa* 11397 | *Efa* 11423 | *Efa* EF7 |
| **D** | C- | *Vfl* 12318 | *Eae* 2094 | *Kpp* 2095 | *Pmi* 3257 | *Sau* 8224 | *Sau* SA1 | *Sau* SA2 | *Spa* SP1 | *Spa* SP2 | *Sha* 13349 | C- |
| **E** | C- | *Lll* 6890 | *Lmi* 14925 | *Sag* 15083 | *Sag* SA2 | *Sag* SA3 | *Sag* SA4 | *Sag* SA6 | *Sag* SA7 | *Sag* SA8 | *Sag* SA9 | C- |
| **F** | *Sag* SA10 | *Sag* SA11 | *Sag* SA21 | *Sag* SA25 | *Sag* SA28 | *Sag* SA29 | *Sag* SA30 | *Sag* SA31 | *Sag* SA32 | *Sag* SA33 | *Sag* SA34 | *Sbo* 8518 |
| **G** | *Sdy* SD1 | *Sdy* SS4 | *Spu* 12174 | *Ssa* 112 | *Sub* 9465 | *Sub* SU1 | *Sub* SU2 | *Sub* SU3 | *Efc* 7937 | *Efa* 11397 | *Efa* 11423 | *Efa* EF7 |
| **H** | C+ | *Vfl* 12318 | *Eae* 2094 | *Kpp* 2095 | *Pmi* 3257 | *Sau* 8224 | *Sau* SA1 | *Sau* SA2 | *Spa* SP1 | *Spa* SP2 | *Sha* 13349 | C+ |

## 3.1. Taxa-specific markers

Taxa-specific markers of prevalent mastitis groups were tested (Fig. 8). Concerning the Streptococcaceae-specific marker (F1), experimental validation revealed positive results for all species belonging to the *Streptococcus* genus, but no signal was present for strains representing the other genera (*Lll* 6890 and *Lmi* 14925). On the other hand, the marker selected for *Lactococcus* (LC2) was detected exclusively in the representative strain (*Lll* 6890). The two markers transversal to the *Streptococcus* genus (ST1 and ST2) only presented positive hybridisation in *S. uberis*, while ST3 was only specific to *S. agalactiae* (17 out of 19 strains). Marker A1, selected for *S. agalactiae*, presented positive results in all 19 *S. agalactiae* strains, whereas marker A2, expected to be also specific to *S. dysgalactiae*, was only detected in *S. agalactiae* strains and showed a significantly lower hybridisation signal in *Sag* SA2 and *Sag* SA3. The *S. uberis*-specific marker (SU) showed complete specificity to all four strains tested.

## 3.2. Functional markers

Eight functional markers, based on phenotypic traits of bovine mastitis-causing pathogens, were validated by dot blot hybridisation (Fig. 9). Results obtained with markers from the fructose operon revealed that FO1 and FO3 were present in all 19 strains of *S. agalactiae*, although *S. agalactiae* LMG 15083 (*Sag* 15083) of human origin presented a lower signal. These markers were also detected in one strain of *S. dysgalactiae* (*Sdy* SS4) and *S. parauberis* (*Spu* 12174), whereas FO2 showed an

overall unspecific hybridisation. The nisin U operon markers (NU1 and NU3) were present in one strain of *S. uberis* (*Sub* SU3) and 10 strains of *S. agalactiae*, sampled from the same location. The virulence-associated marker for *hasC* (V1) was specific to *S. uberis* with positive signals for all four strains. The *gapC* marker (V2) was present in all species of *Streptococcus* tested (except *Sbo* 8518) and other taxa closely related to Streptococaceae (e.g. *Efc* 7937 and *Vfl* 12318). Dot blots using the *oppF* gene marker (V3), detected positive hybridisation in all four *S. uberis* strains tested, in addition to *S. bovis* (*Sbo* 8518) and *S. parauberis* (*Spu* 12174).

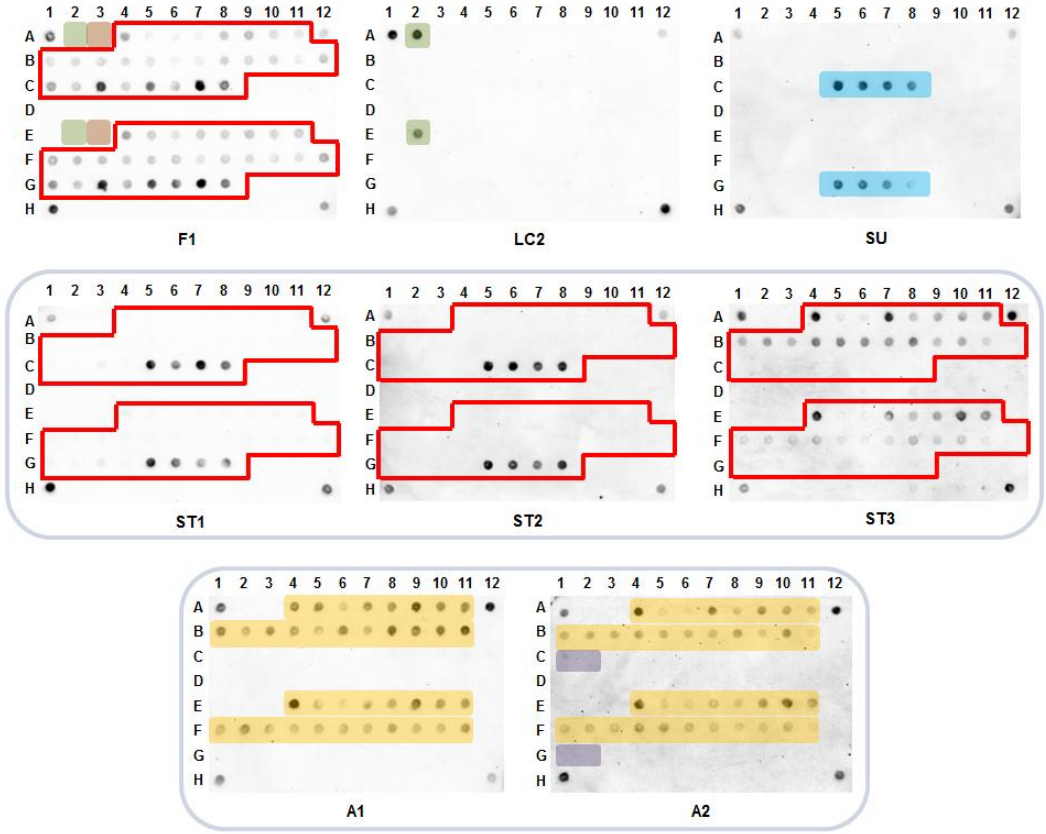


**Figure 8**. Dot blots of taxa-specific markers for Streptococcaceae (F1), *Lactococcus* (LC2), *S. uberis* (SU), *Streptococcus* (ST1, ST2 and ST3) and *S. agalactiae* (A1 and A2). The target taxon for each marker is highlighted according to the following colour scheme: green indicates strains belonging to *L. lactis*, brown to *L. miscens*, yellow to *S. agalactiae*, purple to *S. dysgalactiae*, blue to *S. uberis* and the red outline indicates organisms representative of the *Streptococcus* genus.

## 3.3. Automatic image analysis

To complement the qualitative validation of the results obtained from the dot blot assays, an image analysis algorithm was used. This software avoids operator-dependent interpretation of the results which is important for the implementation of these techniques in routine diagnostic laboratories (Marçal *et al.*, 2009; Caridade *et al.*, 2010). The algorithm identifies variation in dot intensities, outputting a probability value

based on the measured signal of the pixels in relation to the positive and negative experimental controls. The average calculated values obtained confirmed most of the results validated by qualitative interpretation (Tables 14 and 15). Results obtained with markers F1, LC2, ST1, ST2, A1, SU, V1, V2, V3, FO1, FO2 and FO3 were supported by the image analysis algorithm. Marker F1 presented high probability values for all streptococci (although a slightly lower value for *Sag* SA3). Analysis of genus-specific markers LC2, ST1 and ST2 revealed a probability value of 1.00 for all results visually interpreted as positive signals, with strain *Sdy* SD1 presenting a moderate probability value of 0.47 for marker ST2. For markers A1, SU, V1, V2, V3, FO1, FO2 and FO3 probability values between 0.50 and 1.00 were computed for all expected strains. However, validation of markers A2, ST3, NU1 and NU3 presented negligible probability estimates for strains *Sag* SA2 and *Sag* SA3.

The data complemented by the automatic image analysis algorithm highlights the role of this tool as an essential utility for dot blot validation, but some discrepancies indicate that a thorough experimental validation is required for a reliable analysis.

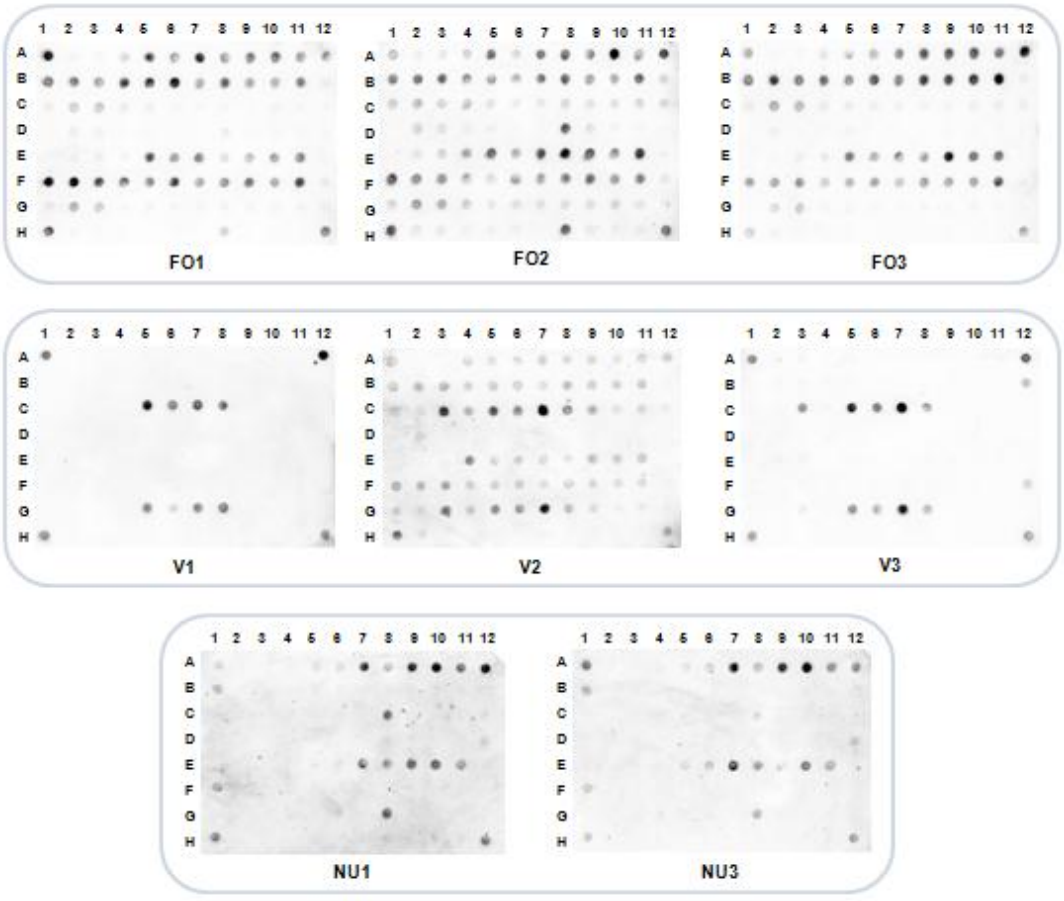


**Figure 9**. Dot blots of functional markers from the fructose operon (FO1, FO2 and FO3), the virulence-associated genes *hasC* (V1), *gapC* (V2) and *oppF* (V3) and from the nisin U operon (NU1 and NU3).

**Table 14.** Average probability values of the results obtained from the dot blot assays, using the taxa-specific markers selected. Values below 0.25 are indicated as red, values between 0.25 and 0.5 as yellow and values above 0.5 are shown as light green. Species are highlighted according to the following colour scheme: green indicates strains belonging to *L. lactis*, brown to *L. miscens*, yellow to *S. agalactiae*, purple to *S. dysgalactiae*, blue to *S. uberis* and the red outline indicates organisms representative of the *Streptococcus* genus.

| Species | Abbreviation | Calculated probability (Taxa-specific markers) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | F1 | LC2 | ST1 | ST2 | ST3 | A1 | A2 | SU |
| *Enterobacter aerogenes* LMG 2094 | *Eae* 2094 | 0.01 | 0.01 | 0.00 | 0.00 | 0.11 | 0.02 | 0.00 | 0.00 |
| *Enterococcus faecalis* LMG 7937 | *Efc* 7937 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Enterococcus faecium* | *Efa* EF7 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.05 | 0.00 | 0.00 |
| *Enterococcus faecium* LMG 11397 | *Efa* 11397 | 0.01 | 0.01 | 0.00 | 0.00 | 0.02 | 0.02 | 0.00 | 0.00 |
| *Enterococcus faecium* LMG 11423 | *Efa* 11423 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 |
| *Klebisella pneumoniae* subsp. *pneumoniae* LMG 2095 | *Kpp* 2095 | 0.01 | 0.01 | 0.01 | 0.00 | 0.09 | 0.00 | 0.00 | 0.00 |
| *Lactococcus lactis* subsp. *lactis* LMG 6890 | *Lll* 6890 | 0.03 | **1.00** | 0.01 | 0.00 | 0.02 | 0.01 | 0.00 | 0.01 |
| *Lactovum miscens* DSM 14925 | *Lmi* 14925 | 0.04 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.01 | 0.01 |
| *Proteus mirabilis* LMG 3257 | *Pmi* 3257 | 0.01 | 0.01 | 0.01 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus aureus* | *Sau* SA1 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus aureus* | *Sau* SA2 | 0.00 | 0.01 | 0.00 | 0.00 | 0.18 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus aureus* LMG 8224 | *Sau* 8224 | 0.01 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus haemolyticus* LMG 13349 | *Sha* 13349 | 0.01 | 0.01 | 0.00 | 0.00 | 0.07 | 0.01 | 0.00 | 0.00 |
| *Staphylococcus pasteuri* | *Spa* SP1 | 0.01 | 0.01 | 0.00 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus pasteuri* | *Spa* SP2 | 0.01 | 0.02 | 0.00 | 0.00 | 0.13 | 0.00 | 0.00 | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA10 | **0.88** | 0.01 | 0.04 | 0.01 | **0.59** | **0.99** | **0.93** | 0.02 |
| *Streptococcus agalactiae* | *Sag* SA11 | **0.95** | 0.01 | 0.03 | 0.00 | **0.84** | **0.99** | **0.92** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA2 | **0.56** | 0.00 | 0.03 | 0.00 | 0.02 | **0.96** | 0.07 | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA21 | **0.91** | 0.01 | 0.04 | 0.00 | **0.84** | **0.97** | **0.92** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA25 | **0.92** | 0.00 | 0.03 | 0.00 | **0.95** | **0.97** | **0.99** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA28 | **0.83** | 0.00 | 0.04 | 0.00 | **0.57** | **0.94** | **0.97** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA29 | **0.56** | 0.01 | 0.01 | 0.00 | 0.50 | **0.99** | **0.91** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA3 | 0.44 | 0.00 | 0.01 | 0.00 | 0.01 | **0.55** | 0.09 | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA30 | 0.50 | 0.01 | 0.01 | 0.00 | **0.57** | **0.98** | **0.83** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA31 | **0.93** | 0.01 | 0.02 | 0.00 | **0.96** | **1.00** | **0.65** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA32 | **0.90** | 0.01 | 0.02 | 0.00 | **0.79** | **0.99** | 0.48 | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA33 | **0.82** | 0.01 | 0.02 | 0.00 | **0.85** | **0.98** | **0.96** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA34 | **0.55** | 0.00 | 0.02 | 0.00 | 0.28 | **1.00** | 0.04 | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA4 | **0.53** | 0.00 | 0.01 | 0.00 | **0.99** | **1.00** | **0.92** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA6 | **0.96** | 0.00 | 0.02 | 0.00 | **0.86** | **1.00** | **0.74** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA7 | **0.98** | 0.01 | 0.02 | 0.00 | **0.97** | **1.00** | **1.00** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA8 | **0.90** | 0.00 | 0.02 | 0.00 | **0.98** | **1.00** | **0.99** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA9 | **0.84** | 0.00 | 0.05 | 0.00 | **0.99** | **1.00** | **0.93** | 0.02 |
| *Streptococcus agalactiae* LMG 15083 | *Sag* 15083 | **1.00** | 0.01 | 0.07 | 0.00 | **1.00** | **1.00** | **1.00** | 0.00 |
| *Streptococcus bovis* LMG 8518 | *Sbo* 8518 | **0.98** | 0.00 | 0.07 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 |
| *Streptococcus dysgalactiae* | *Sdy* SD1 | **1.00** | 0.01 | 0.06 | 0.47 | 0.02 | 0.01 | 0.00 | 0.00 |
| *Streptococcus dysgalactiae* | *Sdy* SS4 | **0.95** | 0.00 | 0.02 | 0.00 | 0.14 | 0.04 | 0.00 | 0.00 |
| *Streptococcus parauberis* LMG 12174 | *Spu* 12174 | **1.00** | 0.01 | 0.19 | 0.00 | 0.11 | 0.00 | 0.00 | 0.00 |
| *Streptococcus salivarius* | *Ssa* 112 | **0.96** | 0.01 | 0.04 | 0.00 | 0.15 | 0.00 | 0.00 | 0.00 |
| *Streptococcus uberis* | *Sub* SU1 | **0.97** | 0.01 | **1.00** | **1.00** | 0.00 | 0.00 | 0.00 | **1.00** |
| *Streptococcus uberis* | *Sub* SU2 | **1.00** | 0.03 | **1.00** | **1.00** | 0.00 | 0.00 | 0.00 | **1.00** |
| *Streptococcus uberis* | *Sub* SU3 | **1.00** | 0.02 | **1.00** | **1.00** | 0.04 | 0.02 | 0.00 | **0.59** |
| *Streptococcus uberis* LMG 9465 | *Sub* 9465 | **1.00** | 0.02 | **1.00** | **1.00** | 0.12 | 0.00 | 0.03 | **1.00** |
| *Vagococcus fluvialis* LMG 12318 | *Vfl* 12318 | 0.02 | 0.03 | 0.00 | 0.00 | 0.12 | 0.02 | 0.00 | 0.01 |

**Table 15.** Average probability values of the results obtained from the dot blot assays, using the functional markers selected. Values below 0.25 are indicated as red, values between 0.25 and 0.5 as yellow and values above 0.5 are shown as light green.

| Species | Abbreviation | Calculated probability (Functional markers) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | FO1 | FO2 | FO3 | NU1 | NU3 | V1 | V2 | V3 |
| *Enterobacter aerogenes* LMG 2094 | *Eae* 2094 | 0.02 | 0.41 | 0.02 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 |
| *Enterococcus faecalis* LMG 7937 | *Efc* 7937 | 0.01 | **0.59** | 0.04 | 0.04 | 0.00 | 0.00 | **0.89** | 0.00 |
| *Enterococcus faecium* | *Efa* EF7 | 0.08 | 0.49 | 0.16 | 0.00 | 0.00 | 0.07 | 0.04 | 0.00 |
| *Enterococcus faecium* LMG 11397 | *Efa* 11397 | 0.02 | **0.75** | 0.12 | 0.00 | 0.00 | 0.00 | 0.22 | 0.00 |
| *Enterococcus faecium* LMG 11423 | *Efa* 11423 | 0.03 | **0.57** | 0.18 | 0.00 | 0.00 | 0.03 | 0.39 | 0.00 |
| *Klebisella pneumoniae* subsp. *pneumoniae* LMG 2095 | *Kpp* 2095 | 0.00 | 0.08 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Lactococcus lactis* subsp. *lactis* LMG 6890 | *Lll* 6890 | 0.06 | 0.23 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 |
| *Lactovum miscens* DSM 14925 | *Lmi* 14925 | 0.06 | **0.55** | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.06 |
| *Proteus mirabilis* LMG 3257 | *Pmi* 3257 | 0.00 | 0.08 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus aureus* | *Sau* SA1 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.01 | 0.00 |
| *Staphylococcus aureus* | *Sau* SA2 | 0.49 | **1.00** | 0.00 | 0.46 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus aureus* LMG 8224 | *Sau* 8224 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| *Staphylococcus haemolyticus* LMG 13349 | *Sha* 13349 | 0.00 | 0.10 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus pasteuri* | *Spa* SP1 | 0.00 | 0.40 | 0.01 | 0.23 | 0.00 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus pasteuri* | *Spa* SP2 | 0.00 | 0.08 | 0.01 | 0.26 | 0.00 | 0.01 | 0.00 | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA10 | **1.00** | **1.00** | **1.00** | **0.95** | **0.92** | 0.00 | **1.00** | 0.01 |
| *Streptococcus agalactiae* | *Sag* SA11 | **1.00** | **1.00** | **1.00** | 0.20 | 0.00 | 0.00 | **1.00** | 0.01 |
| *Streptococcus agalactiae* | *Sag* SA2 | **1.00** | **1.00** | **0.96** | 0.00 | 0.02 | 0.00 | **0.79** | 0.01 |
| *Streptococcus agalactiae* | *Sag* SA21 | **1.00** | **1.00** | **1.00** | 0.00 | 0.00 | 0.00 | **1.00** | 0.01 |
| *Streptococcus agalactiae* | *Sag* SA25 | **1.00** | **0.97** | **0.97** | 0.00 | 0.00 | 0.00 | **0.99** | 0.01 |
| *Streptococcus agalactiae* | *Sag* SA28 | **1.00** | **0.80** | **1.00** | 0.00 | 0.00 | 0.00 | **0.99** | 0.01 |
| *Streptococcus agalactiae* | *Sag* SA29 | **1.00** | **0.97** | **1.00** | 0.00 | 0.00 | 0.00 | **0.99** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA3 | **0.99** | **0.87** | **1.00** | 0.00 | **0.61** | 0.00 | **0.89** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA30 | **0.98** | **1.00** | **0.99** | 0.00 | 0.00 | 0.00 | **0.86** | 0.01 |
| *Streptococcus agalactiae* | *Sag* SA31 | **1.00** | **1.00** | **1.00** | 0.00 | 0.00 | 0.00 | **0.98** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA32 | **1.00** | **1.00** | **1.00** | 0.00 | 0.00 | 0.00 | **0.92** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA33 | **0.99** | **1.00** | **1.00** | 0.00 | 0.00 | 0.00 | **0.74** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA34 | **1.00** | **1.00** | **1.00** | 0.00 | 0.00 | 0.01 | 0.48 | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA4 | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | 0.00 | **0.95** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA6 | **0.80** | **1.00** | **1.00** | **0.91** | **0.98** | 0.00 | **0.53** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA7 | **0.97** | **1.00** | **1.00** | **1.00** | **0.50** | 0.00 | **0.96** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA8 | **1.00** | **1.00** | **1.00** | **1.00** | **1.00** | 0.02 | **0.75** | 0.00 |
| *Streptococcus agalactiae* | *Sag* SA9 | **1.00** | **0.99** | **1.00** | **1.00** | **1.00** | 0.01 | **0.93** | 0.02 |
| *Streptococcus agalactiae* LMG 15083 | *Sag* 15083 | 0.37 | **0.93** | **0.60** | 0.00 | 0.00 | 0.00 | **0.95** | 0.00 |
| *Streptococcus bovis* LMG 8518 | *Sbo* 8518 | 0.19 | **0.62** | 0.33 | 0.00 | 0.00 | 0.03 | 0.00 | **0.95** |
| *Streptococcus dysgalactiae* | *Sdy* SD1 | 0.21 | **0.87** | 0.20 | 0.02 | 0.00 | 0.00 | **1.00** | 0.02 |
| *Streptococcus dysgalactiae* | *Sdy* SS4 | **0.94** | **0.99** | **0.93** | 0.00 | 0.00 | 0.00 | **0.97** | 0.00 |
| *Streptococcus parauberis* LMG 12174 | *Spu* 12174 | **0.91** | **0.91** | **0.98** | 0.00 | 0.00 | 0.00 | **1.00** | **0.60** |
| *Streptococcus salivarius* | *Ssa* 112 | 0.07 | **0.93** | 0.25 | 0.00 | 0.00 | 0.00 | **1.00** | 0.02 |
| *Streptococcus uberis* | *Sub* SU1 | 0.01 | 0.40 | 0.13 | 0.23 | 0.00 | **0.77** | **1.00** | **0.99** |
| *Streptococcus uberis* | *Sub* SU2 | 0.01 | 0.45 | 0.07 | 0.00 | 0.00 | **1.00** | **1.00** | **1.00** |
| *Streptococcus uberis* | *Sub* SU3 | 0.02 | **0.75** | 0.07 | **1.00** | 0.50 | **1.00** | **1.00** | **0.98** |
| *Streptococcus uberis* LMG 9465 | *Sub* 9465 | 0.04 | 0.32 | 0.25 | 0.00 | 0.00 | **1.00** | **1.00** | **1.00** |
| *Vagococcus fluvialis* LMG 12318 | *Vfl* 12318 | 0.06 | **0.56** | 0.11 | 0.00 | 0.00 | 0.00 | **0.77** | 0.03 |

## 3.4. Further experimental validation with *S. uberis*

In order to assess the consistency and stability of the taxa-specific marker SU and the functional markers NU1 and NU3, further experimental analyses were done in six additional *S. uberis* isolates provided by SEGALAB (Table 16). Unrelated organisms were added to the assay as negative controls.

**Table 16.** Layout of the membranes used for the dot blot validation with a set of 10 *S. uberis* strains (highlighted in blue). Abbreviations are listed in Table 7.

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **A** | C+ | *Sub* 9465 | *Sub* SU1 | *Sub* SU2 | C+ |
| **B** | *Sub* SU3 | *Sub* SU4 | *Sub* SU5 | *Sub* SU6 | *Sub* SU7 |
| **C** | *Sub* SU8 | *Sub* SU9 | *Efa* 11423 | *Kpp* 2095 | *Pmi* 3257 |
| **D** | C- | *Spa* SP1 | *Spa* SP2 | *Vfl* 12318 | C- |
| **E** | C- | *Sub* 9465 | *Sub* SU1 | *Sub* SU2 | C- |
| **F** | *Sub* SU3 | *Sub* SU4 | *Sub* SU5 | *Sub* SU6 | *Sub* SU7 |
| **G** | *Sub* SU8 | *Sub* SU9 | *Efa* 11423 | *Kpp* 2095 | *Pmi* 3257 |
| **H** | C+ | *Spa* SP1 | *Spa* SP2 | *Vfl* 12318 | C+ |

The taxonomic marker specific to *S. uberis* (SU) showed a positive hybridisation signal specific to all 10 *S. uberis* strains tested, whereas NU1 and NU3 markers from the nisin operon were only detected in *Sub* SU3, as previously observed (Fig. 10).

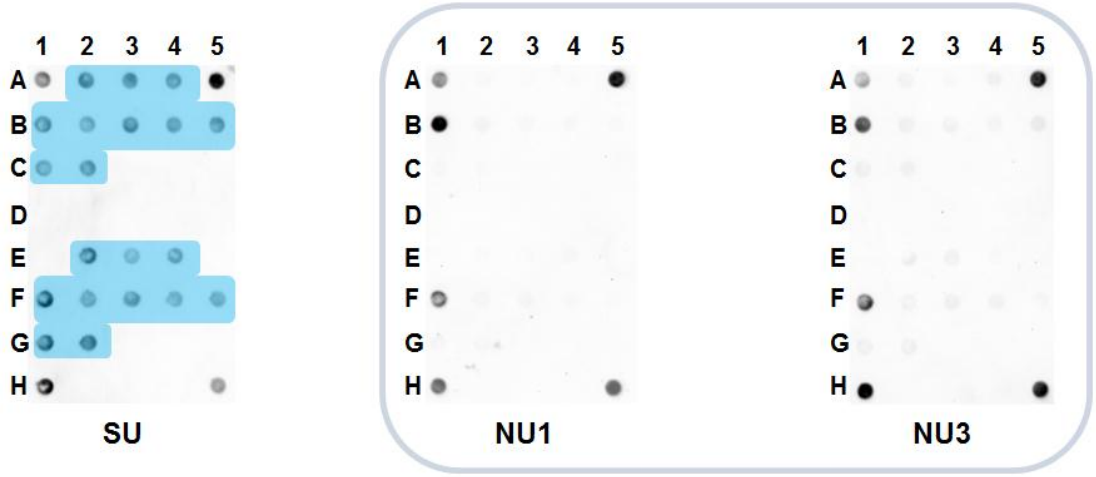


**Figure 10**. Dot blots of the taxa-specific marker for *S. uberis* (highlighted in blue - SU) and functional markers based on the nisin operon (NU1 and NU3) with a set of 10 *S. uberis* strains and 6 unrelated organisms.

The complementary analysis with the image algorithm confirmed the qualitative validation of the dot blots (Table 17). All *S. uberis* strains detected presented a probability value above 0.97.

**Table 17.** Average probability values of the results obtained from the dot blot assays with *S. uberis* (highlighted in blue), using the taxa-specific marker SU and the functional markers from the nisin operon (NU1 and NU3). Values below 0.25 are indicated as red, values between 0.25 and 0.5 as yellow and values above 0.5 are shown as light green.

| Species | Abbreviation | Calculated probability (Taxa-specific markers) | | |
|---|---|---|---|---|
| | | SU | NU1 | NU3 |
| *Enterococcus faecium* LMG 11423 | *Efa* 11423 | 0.01 | 0.00 | 0.00 |
| *Klebisella pneumoniae* subsp. *pneumoniae* LMG 2095 | *Kpp* 2095 | 0.01 | 0.00 | 0.00 |
| *Proteus mirabilis* LMG 3257 | *Pmi* 3257 | 0.00 | 0.00 | 0.00 |
| *Staphylococcus pasteuri* | *Spa* SP1 | 0.05 | 0.00 | 0.00 |
| *Staphylococcus pasteuri* | *Spa* SP2 | 0.00 | 0.00 | 0.01 |
| *Streptococcus uberis* | *Sub* SU1 | **0.97** | 0.01 | 0.04 |
| *Streptococcus uberis* | *Sub* SU2 | **1.00** | 0.02 | 0.05 |
| *Streptococcus uberis* | *Sub* SU3 | **1.00** | **1.00** | **1.00** |
| *Streptococcus uberis* | *Sub* SU4 | **0.99** | 0.07 | 0.08 |
| *Streptococcus uberis* | *Sub* SU5 | **1.00** | 0.05 | 0.08 |
| *Streptococcus uberis* | *Sub* SU6 | **0.99** | 0.03 | 0.06 |
| *Streptococcus uberis* | *Sub* SU7 | **0.99** | 0.02 | 0.12 |
| *Streptococcus uberis* | *Sub* SU8 | **1.00** | 0.04 | 0.09 |
| *Streptococcus uberis* | *Sub* SU9 | **1.00** | 0.03 | 0.08 |
| *Streptococcus uberis* LMG 9465 | *Sub* 9465 | **1.00** | 0.01 | 0.06 |
| *Vagococcus fluvialis* LMG 12318 | *Vfl* 12318 | 0.00 | 0.00 | 0.01 |

The combined results obtained confirm the specificity and stability of the taxa-specific marker for *S. uberis* (SU). On the other hand, the nisin operon markers (NU1 and NU3) were only detected in one *S. uberis* strain, suggesting that a significantly larger sample base is required to assess the prevalence and relation of the ability to produce nisin with the virulence potential of this species during bovine mastitis infection.

# Discussion

## 1. Current diagnostic methodologies

The characterisation of microbial communities is one of the main focuses of microbial ecology. Nevertheless, routine diagnostic laboratories still heavily rely on culture-based methods for the identification and detection of microorganisms. Moreover, traditional biochemical assays frequently lack the reliability and accuracy of modern, DNA-based approaches (Ieven *et al.*, 1995; Facklam, 2002). Thus, detection techniques, based on PCR and hybridisation technologies, have shown to be more successful, reliable and efficient methods for identification and detection of bacterial species in the fields of medicine, agriculture, public health and forensics (Trevors and Masson, 2010).

Regarding bovine mastitis, PCR-based methods have been developed for the detection of *Streptococcus*, *Staphylococcus* and other organisms associated to bovine infections (Kuang *et al.*, 2009; Shome *et al.*, 2011; Ajitkumar *et al.*, 2012). However, detection techniques lack the discriminatory resolution of traditional genotyping methods, which is essential for epidemiological characterisation, especially considering that bovine mastitis can be caused by a number of contagious pathogens. Detection methods are designed to determine the presence or absence of a particular organism, with no additional insight into the infrasubspecific diversity, whereas the most effective epidemiological studies are based on extremely conserved regions (e.g. housekeeping genes) that cannot be used as DNA markers for the detection of pathogenic organisms at a species level (Zadoks and Schukken, 2006; Zadoks *et al.*, 2011). This results in a more laborious and time-consuming effort, limiting the potential implementation of these methods in routine diagnostic laboratories. Therefore, currently one of the major concerns is the development of a platform able to shorten the gap between detection and typing methodologies, for a more efficient response during a bovine infection outbreak.

## 2. Selection of discriminatory DNA markers

Bioinformatics tools allow a simple and accurate selection of taxa-specific markers, and with the advent of culture-independent techniques, an increased number of databases and genomes are available for efficient *in silico* predictions. Most diagnostic studies of streptococci have been based on the comparative analysis of the 16S rRNA gene sequence (Kuang *et al.*, 2009; Shome *et al.*, 2011; Ajitkumar *et al.*, 2012). However,

this genomic region lacks the reliability and discriminatory potential of taxa-specific signatures obtained by current bioinformatics utilities. Insignia (Phillippy *et al.*, 2007) was used to identify five taxonomic markers. This utility has proven to be successful for the *in silico* selection of discriminatory markers, revealing reliable results (Albuquerque *et al.*, 2012). Given that plasmid-based markers could present high variability across different strains, the selection of DNA signatures only included chromosomal data. In this work, special emphasis was given to Streptococcaceae because a high number of prevalent mastitis-causing pathogens have been associated to this taxon, which include *L. lactis*, *S. agalactiae* and *S. uberis* (McDonald *et al.*, 2005; Kuang *et al.*, 2009). Higher taxonomic ranks would significantly increase the sample size for the *in silico* analyses, reducing the chance of finding specific and exclusive markers. Therefore, genera (*Streptococcus* and *Lactococcus*) and species (*S. agalactiae* and *S. uberis*) of important bovine mastitis agents were also selected to increase the reliability and discriminatory potential of this approach.

Bioinformatics tools rely on their own databases and algorithms for the selection of specific DNA signatures, so the Protein Family Database (Finn *et al.*, 2006) was also used to increase the accuracy of the *in silico* predictions. Since the database allows to search for protein domains conserved across all target sequences available, this approach potentially enables the selection of a more specific and transversal DNA region. This tool was used to obtain a DNA marker specific to *Streptococcus*, considered one of the main genus responsible for bovine mastitis.

To confirm the specificity of the DNA signatures obtained by *in silico* analyses, a BLAST specificity test was carried out within the NCBI genome databases (Table 10). Due to the limited database of each bioinformatics tools and the constant addition of new DNA sequences to NCBI, this analysis ensures an up-to-date validation of the selected signatures.

Taxa-specific markers identify the target taxonomic rank, with no additional insight into the organisms' metabolism and virulence capabilities. In regards to bovine mastitis, *S. agalactiae* stands out as one of the most important and prevalent contagious species (Keefe, 1997), which requires efficient and reliable genotyping techniques for strain identification and source tracking. In fact, *S. agalactiae* isolates of human and bovine origin have been grouped in distinct genomic clusters (Martinez *et al.*, 2000). Furthermore, due to lateral gene transfer, more resistant strains of *S. agalactiae* and *S.*

*uberis* are emerging as a more frequent source of infection of bovine mastitis (Richards *et al.*, 2011).

Therefore, to further increase the discriminatory resolution of the analysis, eight functional markers were selected. Strain-specific genes were chosen to target isolates of *S. agalactiae* and *S. uberis* with a bovine origin (Richards *et al.*, 2011). Epidemiological studies of *S. agalactiae* have identified a number of genomic islands associated to bovine isolates, including a four-gene operon responsible for fructose utilisation. Three different markers were designed for genes with important functions for this operon (Fig. 6). Moreover, an 11-gene operon of nisin has been identified in *S. uberis* strains with a competitive advantage during mastitis infection. Two additional markers were designed for a regulation gene and a nisin immunity gene of this operon (Fig. 7). Furthermore, three broad range virulence-associated genes, corresponding to a virulence pattern frequently identified in mastitis isolates of *S. uberis*, were selected (Reinoso *et al.*, 2011).

## 3. Experimental validation by dot blot hybridisation

Hybridisation-based techniques are reliable molecular methods to complement PCR-based approaches. The dot blot throughput, coupled with high-stringency conditions results in a more specific analysis in relation to regular PCR techniques, with a larger sample base for validation.

Preliminary data obtained suggested reliable results for all the taxa-specific markers tested, but some isolates presented unexpected patterns. These strains had been previously identified by the VITEK 2 system, so species identity was further verified by sequencing of the 16S rRNA gene. This revealed an incorrect identification of these isolates (Figure S1-S3 of Supplementary data). Samples previously identified as *S. agalactiae* and *S. dysgalactiae* were, in fact, enterocci and staphylococci, whereas a particular isolate belonging to *S. uberis* was misidentified as *S. agalactiae*. These experimental data further emphasises the increased reliability and efficiency of culture-independent tools over traditional biochemical and phenotypic assays.

After this initial analysis, validation of all taxa-specific and functional markers was carried out by dot blot hybridisation. A set of 44 strains, consisting of reference strains and isolates obtained from different mastitic milk samples and representing different locations within Portugal, was validated by experimental analysis. These included

organisms from Streptococcaceae, *Enterococcus*, *Staphylococcus* and other taxa involved in the bovine mastitis disease.

Results obtained with the taxa-specific markers presented a good overall specificity to the intended targets (Fig. 8). The Streptococcaceae broad-spectrum marker (F1) was only specific to the *Streptococcus* genus, whereas the more specific markers towards *Lactococcus* (LC2), *S. agalactiae* (A1, A2) and *S. uberis* (SU) presented the expected results. In regards to the *S. agalactiae* markers, A1 was the most specific and transversal, with A2 presenting more inconsistent results. Genus-specific markers of *Streptococcus*, obtained by Insignia and Pfam revealed limited specificity: ST1 and ST2 markers, obtained by Insignia, were only detected in *S. uberis* strains, whereas ST3, selected using Pfam, was present exclusively in *S. agalactiae* (17 out of 19 strains). These discrepancies indicate that the differences between the target sequences of most of the streptococci selected for the alignment and design of these markers (Fig. S23 and S25 of Supplementary data) was sufficient to prevent hybridisation with the marked probes. This confirms that dot blot hybridisation assays are highly specific methods and, on the other hand, that more informative and appropriate bioinformatics tools and databases should be used for the selection of broad-spectrum markers.

Functional traits responsible for pathogenicity are usually regulated within unstable or dynamic regions, which are less reliable for detection purposes. Nevertheless they complement the data obtained by taxa-specific markers and can provide additional information on the virulence potential of pathogenic species. Results obtained with the functional markers ranged from the broad-spectrum pattern of V2 and FO2 to a more specific and narrow detection of V1 and NU1 (Fig. 9). In fact, the NU1 and NU3 markers were specific to particular lineages of *S. agalactiae* from Vila do Conde and one strain of *S. uberis*, which, interestingly, was previously identified as *S. agalactiae* by culture-based methods.

Validation of SU, NU1 and NU3 markers in six additional *S. uberis* isolates confirmed previous analyses, but a larger sample base of *S. uberis* strains is required to gain further insight into the diversity of this species and the virulence potential of particular strains with the ability to produce the bacteriocin nisin U.

The collective results obtained by dot blot were additionally validated using an automatic image analysis algorithm (Tables 14, 15 and 17). Due to signal variations,

operator-independent methods are important for the validation of experimental results. Most of the results interpreted by visual observation were confirmed by the program algorithm. However, the particular discrepancies between the qualitative results observed and the data computed through automatic image analysis could be explained, on one hand, because of a low hybridisation signal of these samples in relation to the background noise or, on the other hand, because of an extremely low background signal that could lead to the positive detection of unspecific hybridisation signals. Therefore, dot blot assays should involve the use of a high number of DNA markers and replicates to attenuate experimental inconsistencies.

## 4. Developing a detection and typing platform

Data obtained after experimental validation of all taxa-specific and functional markers reveals a promising set of DNA markers for the simultaneous detection and typing of etiological agents of bovine mastitis within the Streptococcaceae family. Marker F1, shown to be specific to *Streptococcus*, coupled with LC2, A1 and SU is able to discriminate species from *Streptococcus*, *Lactococcus*, *S. agalactiae* and *S. uberis*. Marker A2, specific to *S. agalactiae* can be used to increase the reliability and consistency of the results obtained with the A1 marker.

Functional markers revealed particular patterns and traits of some of the sample strains tested. The fructose operon, especially the FO1 marker of the regulation gene, showed a promising potential to detect bovine isolates of *S. agalactiae*, a particularly useful aspect since diverging niches of human and bovine isolates of *S. agalactiae*, due to horizontal gene transfer events, have been described. However, a larger sample base is required to further validate these results. Furthermore, dot blots with markers from the nisin U operon revealed specificity to *S. agalactiae* strains with a common origin, and one strain of *S. uberis*. This indicates that these markers, associated with other taxa-specific markers, would be able to effectively distinguish these particular lineages of *S. agalactiae* and *S. uberis* strains. Additional epidemiological research would clarify if these specific strains share a common genetic background, or if our selected markers were not able to assess the complete diversity within this group. Furthermore, the virulence-associated marker V2 revealed potential application as a broad range marker for the detection of streptococci and closely related organisms responsible for mastitis infections. The virulence-associated gene *oppF* (V3), coupled with other discriminatory markers, would also able to distinguish *S. bovis* and *S. parauberis* strains involved in

bovine mastitis, which has been acknowledged as a limitation of traditional biochemical assays (Facklam, 2002).

Consequently, the simultaneous use of the most specific taxonomic markers (F1, LC2, A1 and SU) together with the most promising functional markers (V2, V3, FO1 and NU3) can help gain an initial insight into species population and infrasubspecific diversity of bovine mastitis samples for further characterisation and treatment of this disease (Fig. 11). The concepts and methodologies developed in this study can also be applied for the discrimination of bacterial species in other areas such as forensics, biotechnology and agriculture, which reveals the general importance and application of this work.

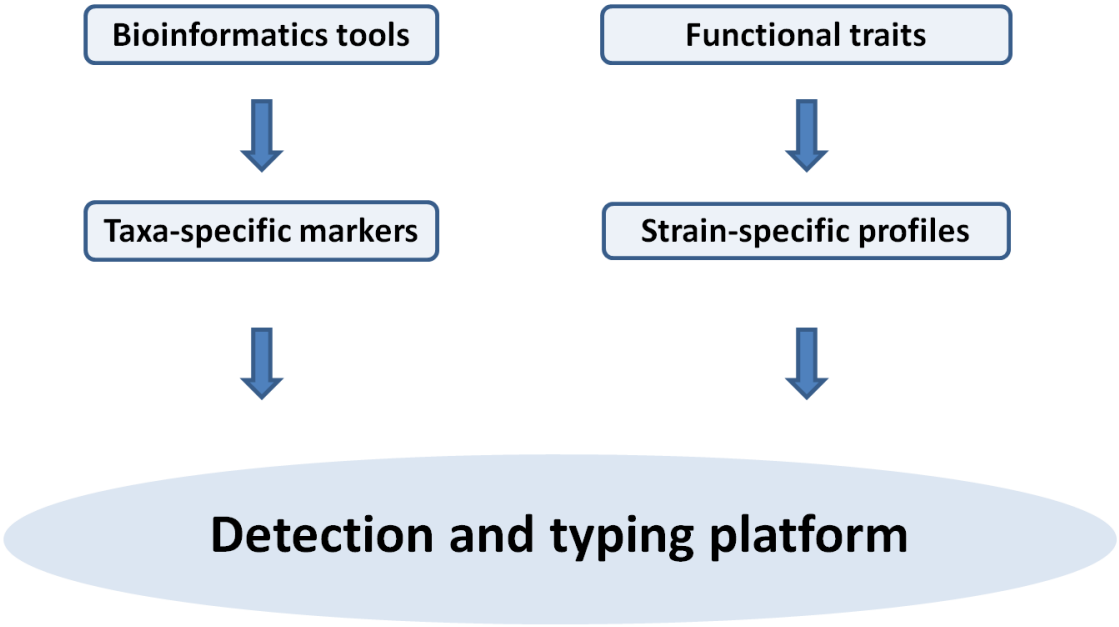


**Figure 11.** Outline of the methodology developed in this study.

# Future perspectives

The results obtained in this work provide a promising methodology that may contribute to an efficient control, treatment and prevention of bovine mastitis, one of the major causes of economic losses in the dairy industry. The simultaneous use of taxa-specific and functional markers can help understand the complexity and diversity of mastitis-causing pathogens that are increasingly providing numerous challenges for the treatment of this illness. Nevertheless, additional research is required to validate the approach described in this study for future implementation in routine diagnostic laboratories:

(1) The addition of new DNA markers of specific functional traits (e.g. responsible for decreased susceptibility to antibiotic treatment) can increase the discriminatory resolution of traditional detection assays, providing information on the infrasubspecific diversity and contributing to the understanding of the evolution and adaptation to the bovine environment of prevalent pathogens. However, the discriminatory potential of these markers will need to be compared with established and reliable genotyping signatures such as housekeeping genes to assess their applicability for strain identification and source tracking.

(2) With a higher number of functional markers, it will be possible to establish a relationship between strain-specific traits and patterns of pathogenic organisms with the clinical signs and severity of bovine mastitis, in order to assess the virulence potential of these strains during infection.

(3) For a faster detection and typing approach, the development of an inverted dot blot platform with the most discriminatory markers will be able to simultaneously evaluate the selected probes with a single hybridisation assay.

(4) For routine implementation, these techniques will need to be validated directly with contaminated milk samples (i.e. without previous isolation in pure culture), comparing their sensitivity and specificity to more controlled laboratory assays.

Therefore, future research in this field must continue to stride towards the development of simple, reliable and cost-effective methods of prevention, treatment and control of bovine mastitis.

# References

Aguilar, J., V. Urday-Cornejo, S. Donabedian, M. Perri, R. Tibbetts and M. Zervos (2010). "*Staphylococcus aureus* meningitis case series and literature review." *Medicine* **89**(2): 117-125.

Ajitkumar, P., H. W. Barkema and J. De Buck (2012). "Rapid identification of bovine mastitis pathogens by high-resolution melt analysis of 16S rDNA sequences." *Veterinary Microbiology* **155**(2-4): 332-340.

Albuquerque, P., C. M. R. Caridade, A. R. S. Marcal, J. Cruz, L. Cruz, C. L. Santos, M. V. Mendes and F. Tavares (2011). "Identification of *Xanthomonas fragariae*, *Xanthomonas axonopodis pv. phaseoli*, and *Xanthomonas fuscans subsp. fuscans* with novel markers and using a dot blot platform coupled with automatic data analysis." *Applied and Environmental Microbiology* **77**(16): 5619-5628.

Albuquerque, P., C. M. R. Caridade, A. S. Rodrigues, A. R. S. Marcal, J. Cruz, L. Cruz, C. L. Santos, M. V. Mendes and F. Tavares (2012). "Evolutionary and experimental assessment of novel markers for detection of *Xanthomonas euvesicatoria* in plant samples." *PLoS ONE* **7**(5): e37836.

Albuquerque, P., M. V. Mendes, C. L. Santos, P. Moradas-Ferreira and F. Tavares (2009). "DNA signature-based approaches for bacterial detection and identification." *Science of the Total Environment* **407**(12): 3641-3651.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." *Journal of Molecular Biology* **215**(3): 403-410.

Amann, R., B. M. Fuchs and S. Behrens (2001). "The identification of microorganisms by fluorescence in situ hybridisation." *Current Opinion in Biotechnology* **12**(3): 231-236.

Amann, R. I., W. Ludwig and K. H. Schleifer (1995). "Phylogenetic identification and in situ detection of individual microbial cells without cultivation." *Microbiological Reviews* **59**(1): 143-169.

Bal, E. B. B., M. A. Bal, T. Isevi and E. Yula (2010). "Application of PCR-RFLP of gap gene method as a molecular typing tool for coagulase negative Staphylococci from bovine and human origin identified with VITEK 2." *African Journal of Microbiology Research* **4**(9): 775-782.

Caridade, C., A. Marcal, T. Mendonca, P. Albuquerque, M. Mendes and F. Tavares (2010). "Automatic analysis of macroarrays images." *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, Aug. 31 2010–Sept. 4 2010*: 6122–6125.

Cleary, P. and Q. Cheng (2006). Medically Important Beta-Hemolytic Streptococci. Prokaryotes. **4:** 108-148.

Cole, J. R., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. J. Farris, A. S. Kulam-Syed-Mohideen, D. M. McGarrell, T. Marsh, G. M. Garrity and J. M. Tiedje (2009). "The Ribosomal Database Project: improved alignments and new tools for rRNA analysis." *Nucleic Acids Research* **37**: D141-D145.

Daleine, G. and P. H. Lagrange (1995). "Preliminary evaluation of a *Mycobacterium tuberculosis* lipooligosaccharide (LOS) antigen in the serological diagnosis of tuberculosis in HIV-seropositive and seronegative patients." *Tubercle and Lung Disease* **76**(3): 234-239.

Dance, A. (2008). "Anthrax case ignites new forensics field." *Nature* **454**(7206): 813-813.

Dmitriev, A., M. Bhide and I. Mikula (2006). "cpn60 gene based multiplex-PCR assay for simultaneous identification of streptococcal species." *Acta Veterinaria Brno* **75**(2): 235-240.

Dohoo, I. R. and K. E. Leslie (1991). "Evaluation of changes in somatic cell counts as indicators of new intramammary infections." *Preventive Veterinary Medicine* **10**(3): 225-237.

Drummond, A. J., B. Ashton, S. Buxton, M. Cheung, A. Cooper, C. Duran, M. Field, J. Heled, M. Kearse, S. Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer and A. Wilson (2012). "Geneious v5.6, Available from http://www.geneious.com."

Endo, A., Y. Futagawa-Endo and L. M. T. Dicks (2011). "Influence of carbohydrates on the isolation of lactic acid bacteria." *Journal of Applied Microbiology* **110**(4): 1085-1092.

Engvall, E. and P. Perlmann (1972). "Enzyme-Linked Immunosorbet Assay, ELISA. 3. Quantitation of specific antibodies by enzyme-labeled anti-imunoglobulin in antigen-coated tubes." *Journal of Immunology* **109**(1): 129-&.

Ercolini, D. (2004). "PCR-DGGE fingerprinting: novel strategies for detection of microbes in food." *Journal of Microbiological Methods* **56**(3): 297-314.

Facklam, R. (2002). "What happened to the streptococci: Overview of taxonomic and nomenclature changes." *Clinical Microbiology Reviews* **15**(4): 613-+.

Fierer, N., C. L. Lauber, N. Zhou, D. McDonald, E. K. Costello and R. Knight (2010). "Forensic identification using skin bacterial communities." *Proceedings of the National Academy of Sciences of the United States of America* **107**(14): 6477-6481.

Finn, R. D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. L. Sonnhammer and A. Bateman (2006). "Pfam: clans, web tools and services." *Nucleic Acids Research* **34**: D247-D251.

Forsman, P., A. TilsalaTimisjarvi and T. Alatossava (1997). "Identification of staphylococcal and streptococcal causes of bovine mastitis using 16S-23S rRNA spacer regions." *Microbiology-Uk* **143**: 3491-3500.

Fortin, M., S. Messier, J. Pare and R. Higgins (2003). "Identification of catalase-negative, non-beta-hemolytic, gram-positive cocci isolated from milk samples." *Journal of Clinical Microbiology* **41**(1): 106-109.

Gunzburg, S. T., N. G. Tornieporth and L. W. Riley (1995). "Identification of enteropathogenic *Escherichia coli* by PCR-based detection of the bundle-formng pilus gene." *Journal of Clinical Microbiology* **33**(5): 1375-1377.

Hale, H. H., E. F. Stula, W. N. Plastridge and C. F. Helmboldt (1962). "Bovine mastitis caused by a *Mycoplasma* species." *Cornell Veterinarian* **52**(4): 582-&.

Harmon, R. J. (1994). "Physiology of mastitis and factors affecting somatic cell counts." *Journal of Dairy Science* **77**(7): 2103-2112.

Hassan, A. A., I. U. Khan, A. Abdulmawjood and C. Lammler (2001). "Evaluation of PCR methods for rapid identification and differentiation of *Streptococcus uberis* and *Streptococcus parauberis*." *Journal of Clinical Microbiology* **39**(4): 1618-1621.

Hernandez-Pando, R., M. Jeyanathan, G. Mengistu, D. Aguilar, H. Orozco, M. Harboe, G. A. W. Rook and G. Bjune (2000). "Persistence of DNA from Mycobacterium tuberculosis in superficially normal lung tissue during latent infection." *Lancet* **356**(9248): 2133-2138.

Hillerton, J. E. and J. E. Semmens (1999). "Comparison of treatment of mastitis by oxytocin or antibiotics following detection according to changes in milk electrical conductivity prior to visible signs." *Journal of Dairy Science* **82**(1): 93-98.

Ieven, M., J. Verhoeven, S. R. Pattyn and H. Goossens (1995). "Rapid and economical method for species identification of clinically significant coagulase-negative staphylococci." *Journal of Clinical Microbiology* **33**(5): 1060-1063.

Jacobs, E. (1993). "Serological diagnosis of *Mycoplasma pneumoniae* infections - A critical review of current procedures." *Clinical Infectious Diseases* **17**: S79-S82.

Jones, G. M. and T. Bailey (2009). "Understanding the basics of mastitis." *Virginia Cooperative Extension*(404-233).

Jones, N., J. F. Bohnsack, S. Takahashi, K. A. Oliver, M. S. Chan, F. Kunst, P. Glaser, C. Rusniok, D. W. M. Crook, R. M. Harding, N. Bisharat and B. G. Spratt (2003). "Multilocus sequence typing system for group B streptococcus." *Journal of Clinical Microbiology* **41**(6): 2530-2536.

Keefe, G. P. (1997). "*Streptococcus agalactiae* mastitis: a review." *Canadian Veterinary Journal-Revue Veterinaire Canadienne* **38**(7): 429-437.

Kitchen, B. J. (1981). "Review of the progress of dairy science: Bovine mastitis: Milk compositional changes and related diagnostic tests." *Journal of Dairy Research* **48**(1): 167-188.

Kuang, Y., K. Tani, A. J. Synnott, K. Ohshima, H. Higuchi, H. Nagahata and Y. Tanji (2009). "Characterization of bacterial population of raw milk from bovine mastitis by culture-independent PCR-DGGE method." *Biochemical Engineering Journal* **45**(1): 76-81.

Lang, P., T. Lefebure, W. Wang, R. N. Zadoks, Y. Schukken and M. J. Stanhope (2009). "Gene content differences across strains of *Streptococcus uberis* identified using oligonucleotide microarray comparative genomic hybridization." *Infection Genetics and Evolution* **9**(2): 179-188.

Lee, N. K., I. A. Chal, Y. L. Park, Y. H. Rark, J. M. Kim, H. M. Nam, S. C. Jung and H. D. Paik (2011). "Medium optimization for lacticin NK34 production by *Lactococcus lactis* NK34 against bovine mastitis pathogen." *Milchwissenschaft-Milk Science International* **66**(1): 68-71.

Lee, S. U., M. Quesnell, L. K. Fox, J. W. Yoon, Y. H. Park, W. C. Davis, D. Falk, C. F. Deobald and G. A. Bohach (1998). "Characterization of staphylococcal bovine mastitis isolates using the polymerase chain reaction." *Journal of Food Protection* **61**(10): 1384-1386.

Leigh, J. A., S. M. Hodgkinson and R. A. Lincoln (1998). "The interaction of *Streptococcus dysgalactiae* with plasmin and plasminogen." *Veterinary Microbiology* **61**(1-2): 121-135.

Lister, J. (1873). "A further contribution to the natural history of bacteria and the germ theory of fermentative changes." *Quart. Microbiol. Sci* **13**: 380-408.

Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman and B. G. Spratt (1998). "Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms." *Proceedings of the National Academy of Sciences of the United States of America* **95**(6): 3140-3145.

Mamo, W., G. Froman, A. Sundas and T. Wadstrom (1987). "Binding of fibronectin, fibrinogen and type-II collagen to streptococci isolated from bovine mastitis." *Microbial Pathogenesis* **2**(6): 417-424.

Marçal, A. R. S., C. M. R. Caridade, P. Albuquerque, M. V. Mendes and F. Tavares (2009). "Automatic detection of molecular markers in digital images." *Engineering in Medicine and Biology Society (EMBC), 2009 Annual International Conference of the IEEE, 3–6 Sept. 2009*: 6710–6713.

Martinez, G., J. Harel, R. Higgins, S. Lacouture, D. Daignault and M. Gottschalk (2000). "Characterization of *Streptococcus agalactiae* isolates of bovine and human origin by randomly amplified polymorphic DNA analysis." *Journal of Clinical Microbiology* **38**(1): 71-78.

Mazumder, R., D. A. Natale, S. Murthy, R. Thiagarajan and C. H. Wu (2005). "Computational identification of strain-, species- and genus-specific proteins." *Bmc Bioinformatics* **6**.

McDonald, J. S. (1979). "Bovine mastitis: Introductory remarks." *Journal of Dairy Science* **62**(1): 117-118.

McDonald, W. L., B. N. Fry and M. A. Deighton (2005). "Identification of *Streptococcus spp*. causing bovine mastitis by PCR-RFLP of 16S-23S ribosomal DNA." *Veterinary Microbiology* **111**(3-4): 241-246.

Michon, A.-L., F. Aujoulat, L. Roudiere, O. Soulier, I. Zorgniotti, E. Jumas-Bilak and H. Marchandin (2010). "Intragenomic and intraspecific heterogeneity in *rrs* may surpass interspecific variability in a natural population of *Veillonella*." *Microbiology-Sgm* **156**: 2080-2091.

Mork, T., T. Tollersrud, B. Kvitle, H. J. Jorgensen and S. Waage (2005). "Comparison of *Staphylococcus aureus* genotypes recovered from cases of bovine, ovine, and caprine mastitis." *Journal of Clinical Microbiology* **43**(8): 3979-3984.

Munoz, M. A., F. L. Welcome, Y. H. Schukken and R. N. Zadoks (2007). "Molecular epidemiology of two *Klebsiella pneumoniae* mastitis outbreaks on a dairy farm in New York State." *Journal of Clinical Microbiology* **45**(12): 3964-3971.

Nakanishi, H., A. Kido, T. Ohmori, A. Takada, M. Hara, N. Adachi and K. Saito (2009). "A novel method for the identification of saliva by detecting oral streptococci using PCR." *Forensic Science International* **183**(1-3): 20-23.

Nordberg, E. K. (2005). "YODA: selecting signature oligonucleotides." *Bioinformatics* **21**(8): 1365-1370.

O'Hara, C. M., F. W. Brenner and J. M. Miller (2000). "Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*." *Clinical Microbiology Reviews* **13**(4): 534-+.

Ote, I., B. Taminiau, J.-N. Duprez, I. Dizier and J. G. Mainil (2011). "Genotypic characterization by polymerase chain reaction of *Staphylococcus aureus* isolates associated with bovine mastitis." *Veterinary Microbiology* **153**(3-4): 285-292.

Pavlidou, S., D. Bozoudi, M. Hatzikamari, N. Tzanetakis and E. Litopoulou-Tzanetaki (2011). "Differentiation of Lactococci from 2 greek cheeses with protected designation of origin by phenotypic criteria and RAPD-PCR." *Journal of Food Science* **76**(3): M175-M183.

Perez, E. A., L. Pusztai and M. van De Vijver (2004). "Improving patient care through molecular diagnostics." *Seminars in Oncology* **31**(5): 14-20.

Petrovski, K. R., N. B. Williamson, N. Lopez-Villalobos, T. J. Parkinson and I. G. Tucker (2011). "Culture results from milk samples submitted to veterinary diagnostic laboratories from August 2003 to December 2006 in New Zealand." *New Zealand Veterinary Journal* **59**(6): 317-322.

Phillippy, A. M., J. A. Mason, K. Ayanbule, D. D. Sommer, E. Taviani, A. Huq, R. R. Colwell, I. T. Knight and S. L. Salzberg (2007). "Comprehensive DNA signature discovery and validation." *Plos Computational Biology* **3**(5): 887-894.

Puigbo, P., I. G. Bravo and S. Garcia-Vallve (2008). "CAIcal: A combined set of tools to assess codon usage adaptation." *Biology Direct* **3**.

Radtke, A., T. Bruheim, J. E. Afset and K. Bergh (2012). "Multiple-locus variant-repeat assay (MLVA) is a useful tool for molecular epidemiologic analysis of *Streptococcus agalactiae* strains causing bovine mastitis." *Veterinary Microbiology* **157**(3-4): 398-404.

Rato, M. G., R. Bexiga, S. F. Nunes, L. M. Cavaco, C. L. Vilela and I. Santos-Sanches (2008). "Molecular epidemiology and population structure of bovine *Streptococcus uberis*." *Journal of Dairy Science* **91**(12): 4542-4551.

Rato, M. G., A. Nerlich, R. Bergmann, R. Bexiga, S. F. Nunes, C. L. Vilela, I. Santos-Sanches and G. S. Chhatwal (2011). "Virulence gene pool detected in bovine group C streptococcus dysgalactiae subsp dysgalactiae Isolates by use of a group A s. pyogenes virulence microarray." *Journal of Clinical Microbiology* **49**(7): 2470-2479.

Reinoso, E. B., M. C. Lasagno, S. A. Dieser and L. M. Odierno (2011). "Distribution of virulence-associated genes in *Streptococcus uberis* isolated from bovine mastitis." *Fems Microbiology Letters* **318**(2): 183-188.

Richards, V. P., P. Lang, P. D. P. Bitar, T. Lefebure, Y. H. Schukken, R. N. Zadoks and M. J. Stanhope (2011). "Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*." *Infection Genetics and Evolution* **11**(6): 1263-1275.

Saini, A., H. Kaur, S. Purwar, S. D. Kholkute and S. Roy (2012). "Discrepancies in identification of *Vibrio cholerae* strains as members of Aeromonadaceae and Enterobacteriaceae by automated microbial identification system." *Letters in Applied Microbiology* **55**(1): 22-26.

Schleifer, K. H., J. Kraus, C. Dvorak, R. Kilpperbalz, M. D. Collins and W. Fischer (1985). "Transfer of *Streptococcus lactis* and related streptococci to the genus *Lactococcus gen-nov*." *Systematic and Applied Microbiology* **6**(2): 183-195.

Sears, P. M. and K. K. McCarthy (2003). "Diagnosis of mastitis for therapy decisions." *Veterinary Clinics of North America-Food Animal Practice* **19**(1): 93-+.

Sharma, N., G. J. Rho, Y. H. Hong, T. Y. Kang, H. K. Lee, T. Y. Hur and D. K. Jeong (2012). "Bovine mastitis: an Asian perspective." *Asian Journal of Animal and Veterinary Advances* **7**(6): 454-476.

Shome, B. R., S. Das Mitra, M. Bhuvana, N. Krithiga, D. Velu, R. Shome, S. Isloor, S. B. Barbuddhe and H. Rahman (2011). "Multiplex PCR assay for species identification of bovine mastitis pathogens." *Journal of Applied Microbiology* **111**(6): 1349-1356.

Smith, A. J., A. J. Kitt, P. N. Ward and J. A. Leigh (2002). "Isolation and characterization of a mutant strain of *Streptococcus uberis*, which fails to utilize a plasmin derived beta-casein peptide for the acquisition of methionine." *Journal of Applied Microbiology* **93**(4): 631-639.

Taponen, S., L. Salmikivi, H. Simojoki, M. T. Koskinen and S. Pyorala (2009). "Real-time polymerase chain reaction-based identification of bacteria in milk samples from bovine clinical mastitis with no growth in conventional culturing." *Journal of Dairy Science* **92**(6): 2610-2617.

Teuber, M. and A. Geis (2006). The Genus *Lactococcus*. Prokaryotes. **4:** 205-228.

Torsvik, V., K. Salte, R. Sorheim and J. Goksoyr (1990). "Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria." *Applied and Environmental Microbiology* **56**(3): 776-781.

Trevors, J. T. and L. Masson (2010). "DNA technologies: what's next applied to microbiology research?" *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **98**(3): 249-262.

Unnerstad, E. H., A. Lindberg, K. Persson Waller, T. Ekman, K. Artursson, M. Nilsson-Ost and B. Bengtsson (2009). "Microbial aetiology of acute clinical mastitis and agent-specific risk factors." *Veterinary Microbiology* **137**(1-2): 90-97.

Valentinweigand, P., M. Y. Traore, H. Blobel and G. S. Chhatwal (1990). "Role of alpha-2-macroglobulin in phagocytosis of group A and group C streptococci." *Fems Microbiology Letters* **70**(3): 321-324.

Vieira, J., M. V. Mendes, P. Albuquerque, P. Moradas-Ferreira and F. Tavares (2007). "A novel approach for the identification of bacterial taxa-specific molecular markers." *Letters in Applied Microbiology* **44**(5): 506-512.

Volokhov, D. V., J. George, S. X. Liu, P. Ikonomi, C. Anderson and V. Chizhikov (2006). "Sequencing of the intergenic 16S-23S rRNA spacer (ITS) region of Mollicutes species and their identification using microarray-based assay and DNA sequencing." *Applied Microbiology and Biotechnology* **71**(5): 680-698.

Ward, P. N., T. R. Field, W. G. F. Ditcham, E. Maguin and J. A. Leigh (2001). "Identification and disruption of two discrete loci encoding hyaluronic acid capsule biosynthesis genes hasA, hasB, and hasC in *Streptococcus uberis*." *Infection and Immunity* **69**(1): 392-399.

Watts, J. L. (1988). "Etiological agents of bovine mastitis." *Veterinary Microbiology* **16**(1): 41-66.

Wilson, I. G. (1997). "Inhibition and facilitation of nucleic acid amplification." *Applied and Environmental Microbiology* **63**(10): 3741-3751.

Wirawan, R. E., N. A. Kleese, R. W. Jack and J. R. Tagg (2006). "Molecular and genetic characterization of a novel nisin variant produced by *Streptococcus uberis*." *Applied and Environmental Microbiology* **72**(2): 1148-1156.

Wyckoff, H. A., J. Chow, T. R. Whitehead and M. A. Cotta (1997). "Cloning, sequence, and expression of the L-(+) lactate dehydrogenase of *Streptococcus bovis*." *Current Microbiology* **34**(6): 367-373.

Wyder, A. B., R. Boss, J. Naskova, T. Kaufmann, A. Steiner and H. U. Graber (2011). "*Streptococcus* spp. and related bacteria: Their identification and their pathogenic potential for chronic mastitis - a molecular approach." *Research in veterinary science* **91**(3): 349-357.

Yu, J., W. H. Wang, B. L. G. Menghe, M. T. Jiri, H. M. Wang, W. J. Liu, Q. H. Bao, Q. Lu, J. C. Zhang, F. Wang, H. Y. Xu, T. S. Sun and H. P. Zhang (2011). "Diversity of lactic acid bacteria associated with traditional fermented dairy products in Mongolia." *Journal of Dairy Science* **94**(7): 3229-3241.

Zadoks, R. N., J. R. Middleton, S. McDougall, J. Katholm and Y. H. Schukken (2011). "Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans." *Journal of Mammary Gland Biology and Neoplasia* **16**(4): 357-372.

Zadoks, R. N. and Y. H. Schukken (2006). "Use of molecular epidemiology in veterinary practice." *Veterinary Clinics of North America-Food Animal Practice* **22**(1): 229-+.

Zadoks, R. N., Y. H. Schukken and M. Wiedmann (2005). "Multilocus sequence typing of *Streptococcus uberis* provides sensitive and epidemiologically relevant subtype information and reveals positive selection in the virulence gene pauA." *Journal of Clinical Microbiology* **43**(5): 2407-2417.

Zancolli, M., P. Canepa, A. Ceravolo, V. Parodi and F. Ansaldi (2011). "Determinants of invasiveness and ability to cause invasive pneumococcal disease, pneumonia and acute otitis media of different serotypes of *Streptococcus pneumoniae*." *Journal of preventive medicine and hygiene* **52**(2): 47-54.

Zaror, L., K. Valenzuela and J. Kruze (2011). "Bovine mastitis caused by *Prototheca zopfii*: first isolation in Chile." *Archivos De Medicina Veterinaria* **43**(2): 173-176.

Zhang, S.-L., J.-G. Shen, G.-H. Shen, Z.-Q. Sun, P.-H. Xu, Y.-L. Peng, Z.-R. Yang and Q. Sun (2007). "Use of a novel multiplex probe array for rapid identification of *Mycobacterium* species from clinical isolates." *World Journal of Microbiology & Biotechnology* **23**(12): 1779-1788.

# Supplementary data

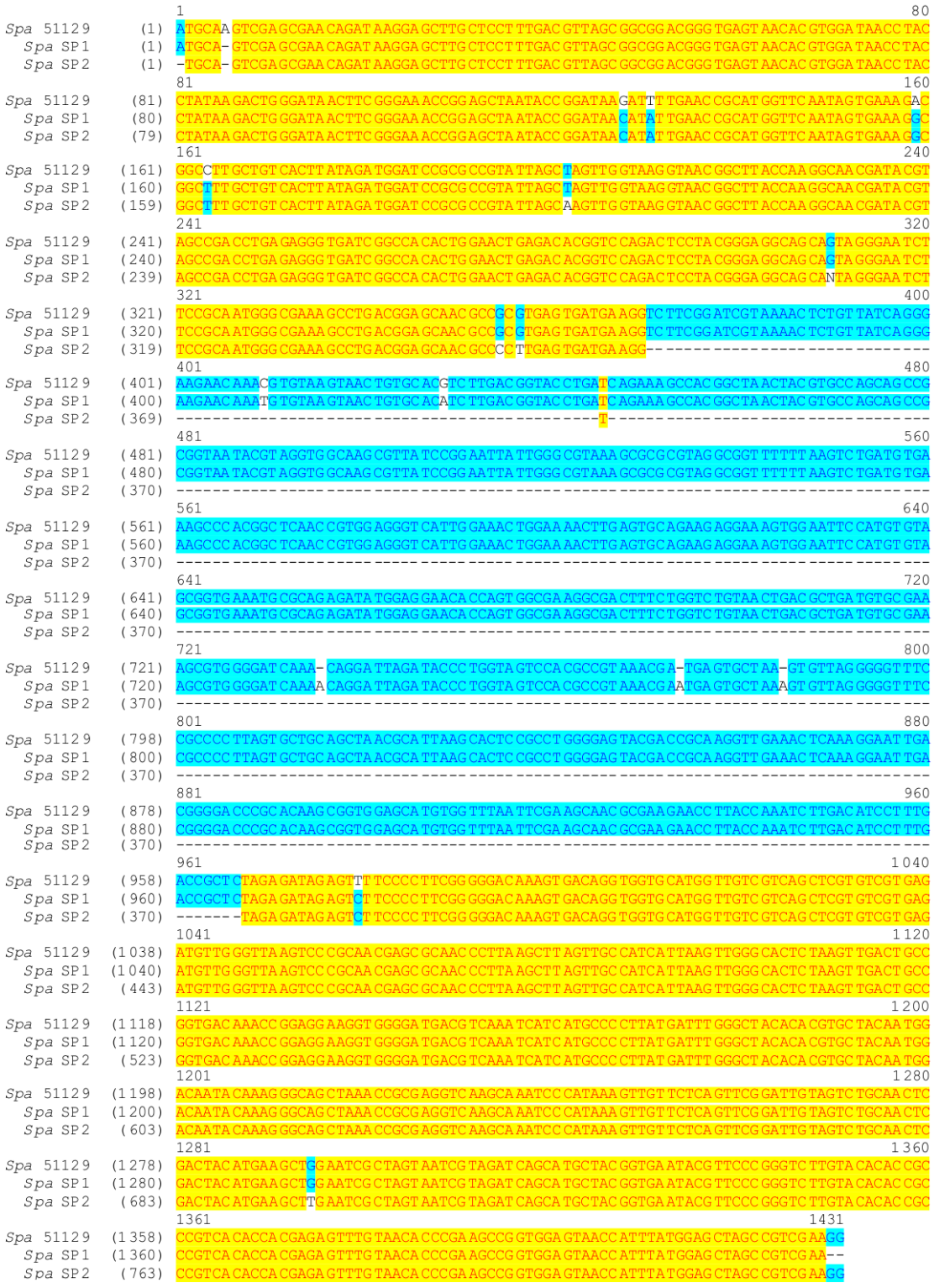

**Figure S1**. Sequence alignment of the 16S rRNA gene of the type strain *Staphylococcus pasteuris* (*Spa* 51129) and the two isolates previously identified as *Streptococcus agalactiae* (*Spa* SP1) and *Streptococcus dysgalactiae* (*Spa* SP2). Yellow indicates base identity between all tested strains, while blue is indicative of the presence of at least one difference in the alignment.

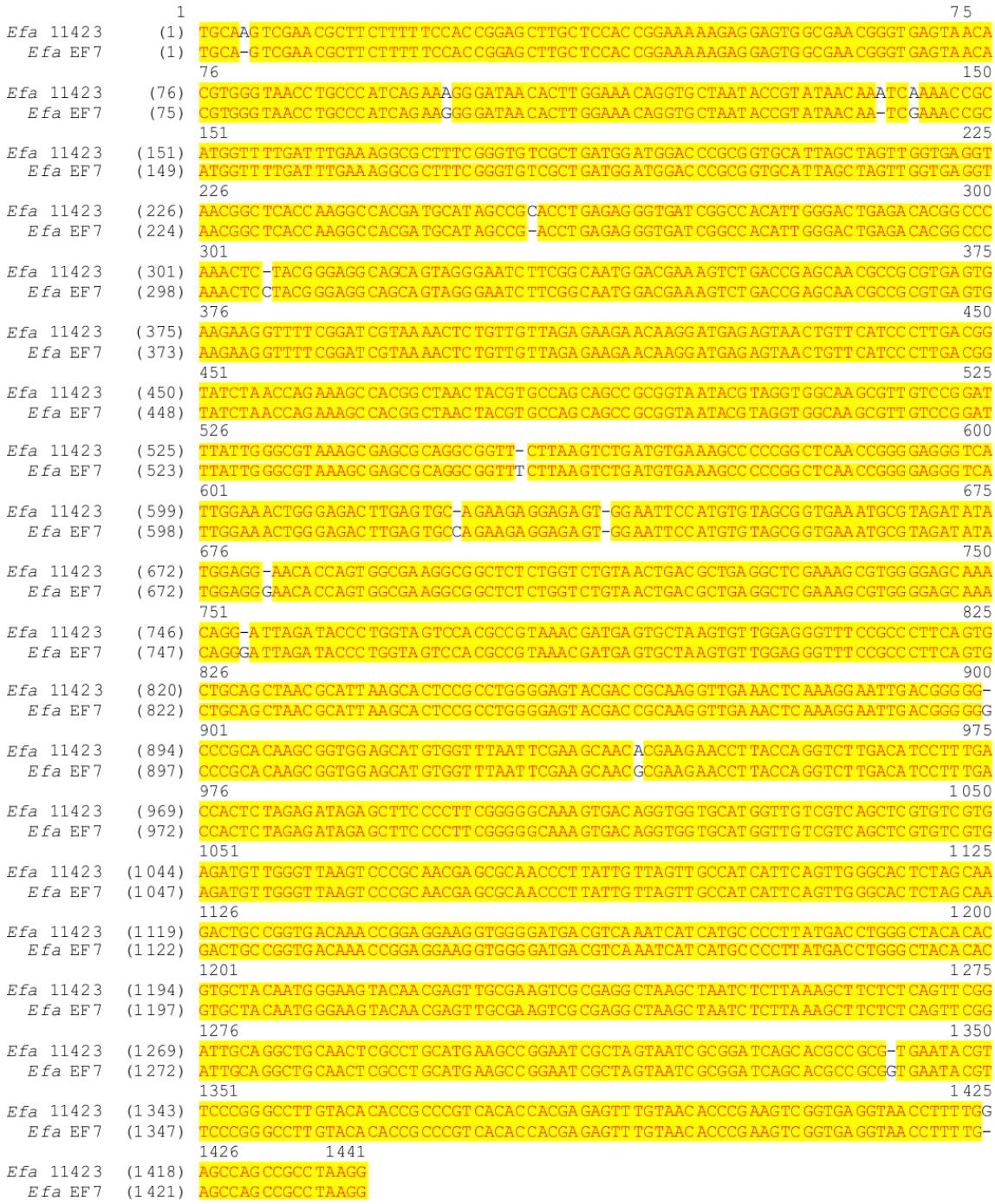


**Figure S2**. Sequence alignment of the 16S rRNA gene of the type strain *Enterococcus faecium* (*Efa* 11423) and the isolate previously identified as *Streptococcus agalactiae* (*Efa* EF7). Yellow indicates base identity between both strains.

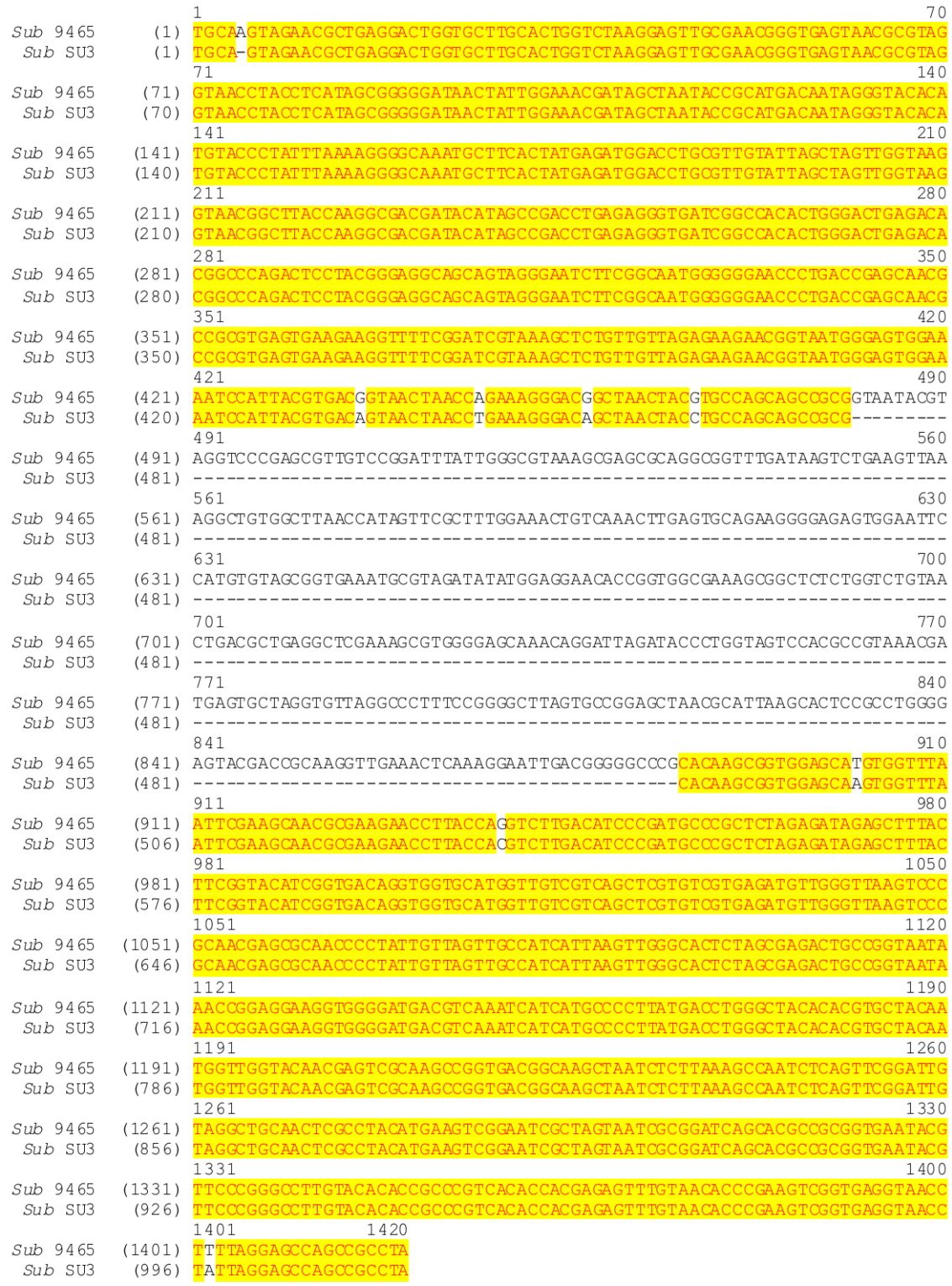


**Figure S3**. Sequence alignment of the 16S rRNA gene of the type strain *Streptococcus uberis* (*Sub 9465*) and the isolate used in this study for the design of the NU1 and NU3 markers, previously identified as *Streptococcus agalactiae* (*Sub* SU3). Yellow indicates base identity between both strains.

```
                    1                                                                                                   70
Sag 14694     (1)   TGCAAGTAGAACGCTGAGGTTTGGTGTTTACACTAGACTGATGAGTTGCGAACGGGTGAGTAACGCGTAG
 Sag SA11     (1)   TGCA-GTAGAACGCTGAGGTTTGGTGTTTACACTAGACTGATGAGTTGCGAACGGGTGAGTAACGCGTAG
                    71                                                                                                  140
Sag 14694    (71)   GTAACCTGCCTCATAGCGGGGGATAACTATTGGAAACGATAGCTAATACCGCATAAGAGTAATTAACACA
 Sag SA11    (70)   GTAACCTGCCTCATAGCGGGGGATAACTATTGGAAACGATAGCTAATACCGCATAAGAGTAATTAACACA
                    141                                                                                                 210
Sag 14694   (141)   TGTTAGTTATTTAAAAGGAGCAATTGCTTCACTGTGAGATGGACCTGCGTTGTATTAGCTAGTTGGTGAG
 Sag SA11   (140)   TGTTAGTTATTTAAAAGGAGCAATTGCTTCACTGTGAGATGGACCTGCGTTGTATTAGCTAGTTGGTGAG
                    211                                                                                                 280
Sag 14694   (211)   GTAAAGGCTCACCAAGGCGACGATACATAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACA
 Sag SA11   (210)   GTAAAGGCTCACCAAGGCGATGATACATAGCCGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACA
                    281                                                                                                 350
Sag 14694   (281)   CGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCGGCAATGGACGGAAGTCTGACCGAGCAACG
 Sag SA11   (280)   CGGCCCAGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCGGCAATGGACGGAAGTCTGACCGAGCAACG
                    351                                                                                                 420
Sag 14694   (351)   CCGCGTG-AGTGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTAGAGAAGAACGTTGGTAGGAGTGGA
 Sag SA11   (350)   CCGCGTGGAGTGAAGAAGGTTTTCGGATCGTAAAGCTCTGTTGTTAGAGAAGAACGTTGGTAGGAGTGGA
                    421                                                                                                 490
Sag 14694   (420)   AAATCTACC-AAGTGACGGTAACTAACCAGAAA-GGGACGGCTAACTACGTGCCAGCAGCCGCGGTAA-T
 Sag SA11   (420)   AAATCTACCCAAGTGACGGTAACTAACCAGAAAAGGGACGGCTAACTACGTGCCAGCAGCCGCGGTAAAT
                    491                                                                                                 560
Sag 14694   (487)   AC-GTAGGTCCCGAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTTCTTTAAGTCTGAA
 Sag SA11   (490)   AC-GTAGGTCCCGAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTTCTTTAAGTCTGAA
                    561                                                                                                 630
Sag 14694   (556)   GTTAAAGGCAGTGGCTTAACCATTGTACGCTTTGGAAACTGGAGGACTTGAGTGCAGAAGGGGAGAGTGG
 Sag SA11   (559)   GTTAAAGGCAGTGGCTTAACCATTGTACGCTTTGGAAACTGGAGGACTTGAGTGCAGAAGGGGAGAGTGG
                    631                                                                                                 700
Sag 14694   (626)   AATTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAGGAACACCGGTGGCGAAAGCGGCTCTCTGGTC
 Sag SA11   (629)   AATTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAGGAACACCGGTGGCGAAAGCGGCTCTCTGGTC
                    701                                                                                                 770
Sag 14694   (696)   TGTAACTGACGCTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTA
 Sag SA11   (699)   TGTAACTGACGCTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTA
                    771                                                                                                 840
Sag 14694   (766)   AACGATGAGTGCTAGGTGTTAGGCCCTTTCCGGGGCTTAGTGCCGCAGCTAACGCATTAAGCACTCCGCC
 Sag SA11   (769)   AACGATGAGTGCTAGGTGTTAGGCCCTTTCCGGGGCTTAGTGCCGCAGCTAACGCATTAAGCACTCCGCC
                    841                                                                                                 910
Sag 14694   (836)   TGGGG-AGTACGACCGCAAGGTTGAAA-CTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATG
 Sag SA11   (839)   TGGGGGAGTACGACCGCAAGGTTGAAA-CTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATG
                    911                                                                                                 980
Sag 14694   (904)   TGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTTCTGACCGGCCTAGAGATAGG
 Sag SA11   (908)   TGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCTTCTGACCGGCCTAGAGATAGG
                    981                                                                                                1050
Sag 14694   (974)   CTTTCTCTTCGGAGCAGAAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTT
 Sag SA11   (978)   CTTTCTCTTCGGAGCAGAAGTGACAGGTGGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTT
                    1051                                                                                               1120
Sag 14694  (1044)   AAGTCCCGCAACGAGCGCAACCCCTATTGTTAGTTGCCATCATTAAGTTGGGCACTCTAGCGAGACTGCC
 Sag SA11  (1048)   AAGTCCCGCAACGAGCGCAACCCCTATTGTTAGTTGCCATCATTAAGTTGGGCACTCTAGCGAGACTGCC
                    1121                                                                                               1190
Sag 14694  (1114)   GGTAATAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGT
 Sag SA11  (1118)   GGTAATAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGT
                    1191                                                                                               1260
Sag 14694  (1184)   GCTACAATGGTTGGTACAACGAGTCGCAAGCCGGTGACGGCAAGCTAATCTCTTAAAGCCAATCTCAGTT
 Sag SA11  (1188)   GCTACAATGGTTGGTACAACGAGTCGCAAGCCGGTGACGGCAAGCTAATCTCTTAAAGCCAATCTCAGTT
                    1261                                                                                               1330
Sag 14694  (1254)   CGGATTGTAGGCTGCAACTCGCCTACATGAAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGT
 Sag SA11  (1258)   CGGATTGTAGGCTGCAACTCGCCTACATGAAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGT
                    1331                                                                                               1400
Sag 14694  (1324)   GAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGA
 Sag SA11  (1328)   GAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAACACCCGAAGTCGGTGA
                    1401                1428
Sag 14694  (1394)   GGTAACCTTTTAGGAGCCAGCCGCCTAA
 Sag SA11  (1398)   GGTAACCTTTTAGGAGCCAGCCGCCTAA
```

**Figure S4**. Sequence alignment of the 16S rRNA gene of the type strain *Streptococcus agalactiae* (*Sag* 14694) and the isolate used in this study for the design of the FO1, FO2 and FO3 markers (*Sag* SA11). Yellow indicates base identity between both strains.


**Figure S5**. Sequence alignment of the F1 marker designed with the published genome of *Streptococcus uberis* 0140J (*Sub* 0140J) and the amplicon sequence obtained in this study with *S. uberis* LMG 9465 (*Sub* 9465). Yellow indicates base identity between the two strains tested.


**Figure S6**. Sequence alignment of the LC2 marker designed with the published genome of *Lactococcus lactis* subsp. *lactis* CV56 (*Lll* CV56) and the amplicon sequence obtained in this study with *Lactococcus lactis* subsp. *lactis* LMG 6890 (*Lll* 6890). Yellow indicates base identity between the two strains tested.

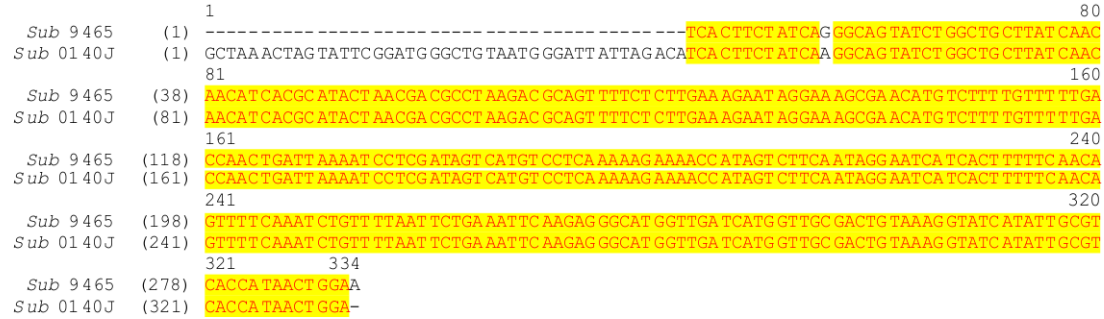


**Figure S7**. Sequence alignment of the ST1 marker designed with the published genome of *Streptococcus uberis* 0140J (*Sub* 0140J) and the amplicon sequence obtained in this study with *S. uberis* LMG 9465 (*Sub* 9465). Yellow indicates base identity between the two strains tested.


**Figure S8**. Sequence alignment of the ST2 marker designed with the published genome of *Streptococcus uberis* 0140J (*Sub* 0140J) and the amplicon sequence obtained in this study with *S. uberis* LMG 9465 (*Sub* 9465). Yellow indicates base identity between the two strains tested.


**Figure S9**. Sequence alignment of the ST3 marker designed with the published genome of *Streptococcus agalactiae* 2603V/R (*Sag* 2603V/R) and the amplicon sequence obtained in this study with *S. agalactiae* LMG 15083 (*Sag* 15083). Yellow indicates base identity between the two strains tested.


**Figure S10**. Sequence alignment of the A1 marker designed with the published genome of *Streptococcus agalactiae* 2603V/R (*Sag* 2603V/R) and the amplicon sequence obtained in this study with *S. agalactiae* LMG 15083 (*Sag* 15083). Yellow indicates base identity between the two strains tested.

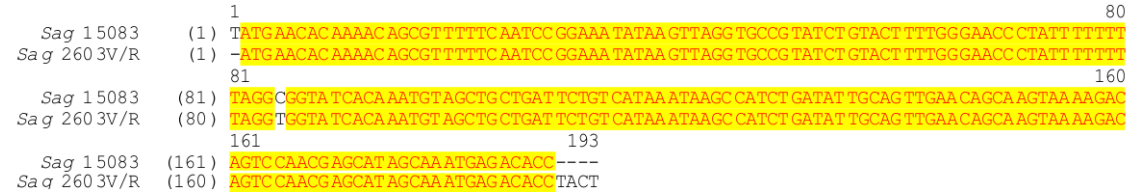


**Figure S11**. Sequence alignment of the A2 marker designed with the published genome of *Streptococcus agalactiae* 2603V/R (*Sag* 2603V/R) and the amplicon sequence obtained in this study with *S. agalactiae* LMG 15083 (*Sag* 15083). Yellow indicates base identity between the two strains tested.


**Figure S12**. Sequence alignment of the SU marker designed with the published genome of *Streptococcus uberis* 0140J (*Sub* 0140J) and the amplicon sequence obtained in this study with *S. uberis* LMG 9465 (*Sub* 9465). Yellow indicates base identity between the two strains tested.

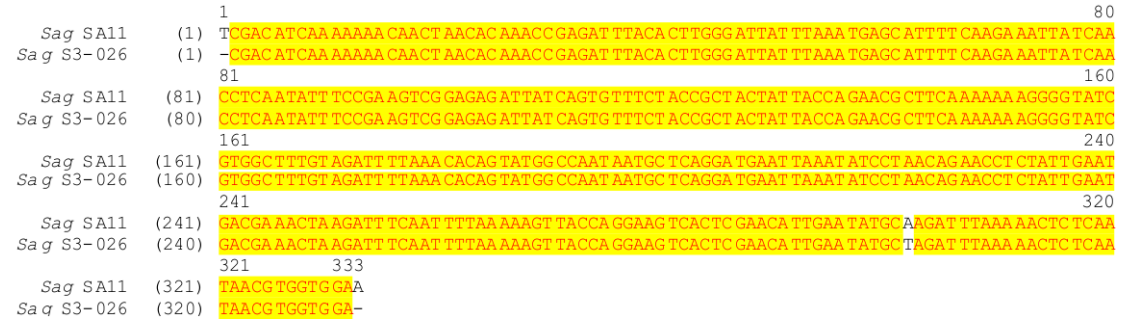


**Figure S13**. Sequence alignment of the FO1 marker designed with *Streptococcus agalactiae* FSL S3-026 (*Sag* S3-026) and the amplicon sequence obtained in this study with *S. agalactiae* SA11 (*Sag* SA11). Yellow indicates base identity between the two strains tested.

**Figure S14**. Sequence alignment of the FO2 marker designed with *Streptococcus agalactiae* FSL S3-026 (*Sag* S3-026) and the amplicon sequence obtained in this study with *S. agalactiae* SA11 (*Sag* SA11). Yellow indicates base identity between the two strains tested.

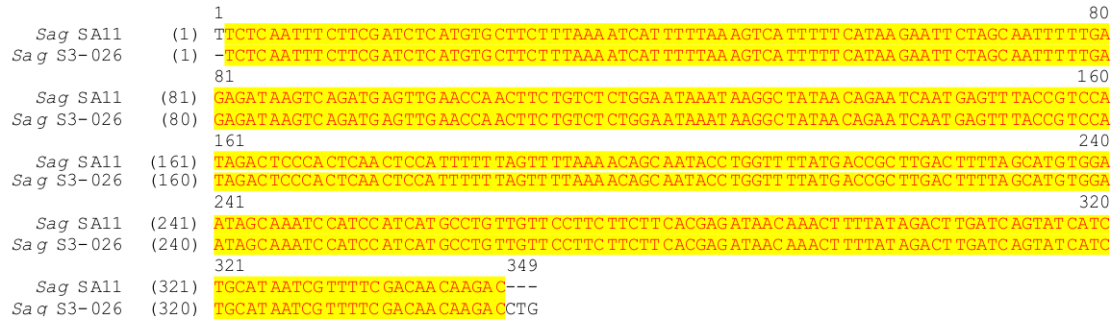

**Figure S15**. Sequence alignment of the FO3 marker designed with *Streptococcus agalactiae* FSL S3-026 (*Sag* S3-026) and the amplicon sequence obtained in this study with *S. agalactiae* SA11 (*Sag* SA11). Yellow indicates base identity between the two strains tested.

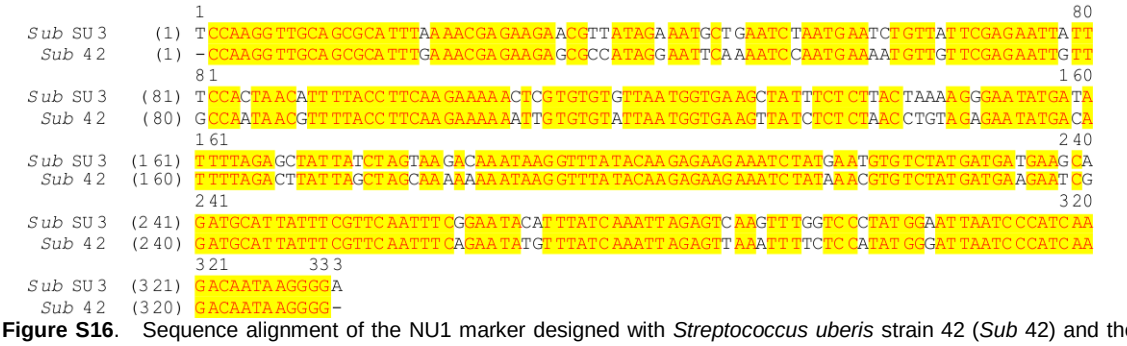

**Figure S16**. Sequence alignment of the NU1 marker designed with *Streptococcus uberis* strain 42 (*Sub* 42) and the amplicon sequence obtained in this study with *S. uberis* SU3 (*Sub* SU3). Yellow indicates base identity between the two strains tested.

**Figure S17**. Sequence alignment of the NU3 marker designed with *Streptococcus uberis* strain 42 (*Sub* 42) and the amplicon sequence obtained in this study with *S. uberis* SU3 (*Sub* SU3). Yellow indicates base identity between the two strains tested.


**Figure S18**. Sequence alignment of the V1 marker designed with the published genome of *Streptococcus uberis* 0140J (*Sub* 0140J) and the amplicon sequence obtained in this study with *S. uberis* LMG 9465 (*Sub* 9465). Yellow indicates base identity between the two strains tested.


**Figure S19**. Sequence alignment of the V2 marker designed with the published genome of *Streptococcus uberis* 0140J (*Sub* 0140J) and the amplicon sequence obtained in this study with *S. uberis* LMG 9465 (*Sub* 9465). Yellow indicates base identity between the two strains tested.

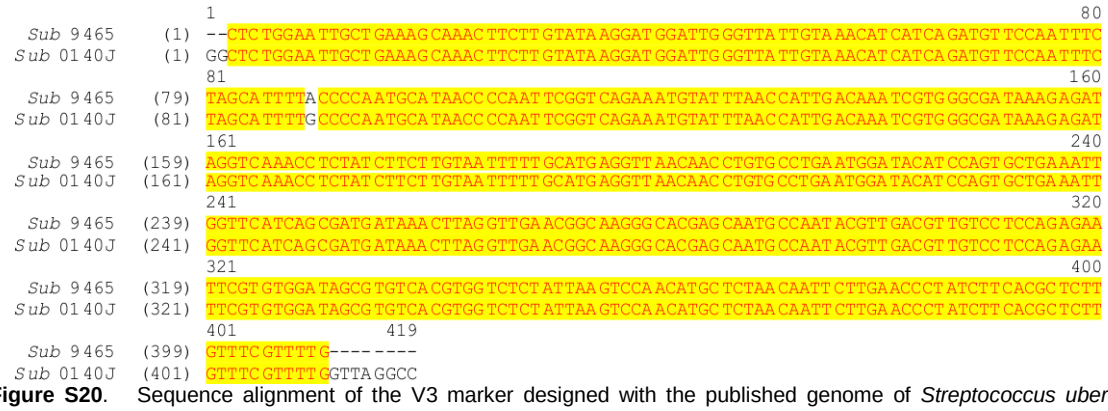


**Figure S20**. Sequence alignment of the V3 marker designed with the published genome of *Streptococcus uberis* 0140J (*Sub* 0140J) and the amplicon sequence obtained in this study with *S. uberis* LMG 9465 (*Sub* 9465). Yellow indicates base identity between the two strains tested.

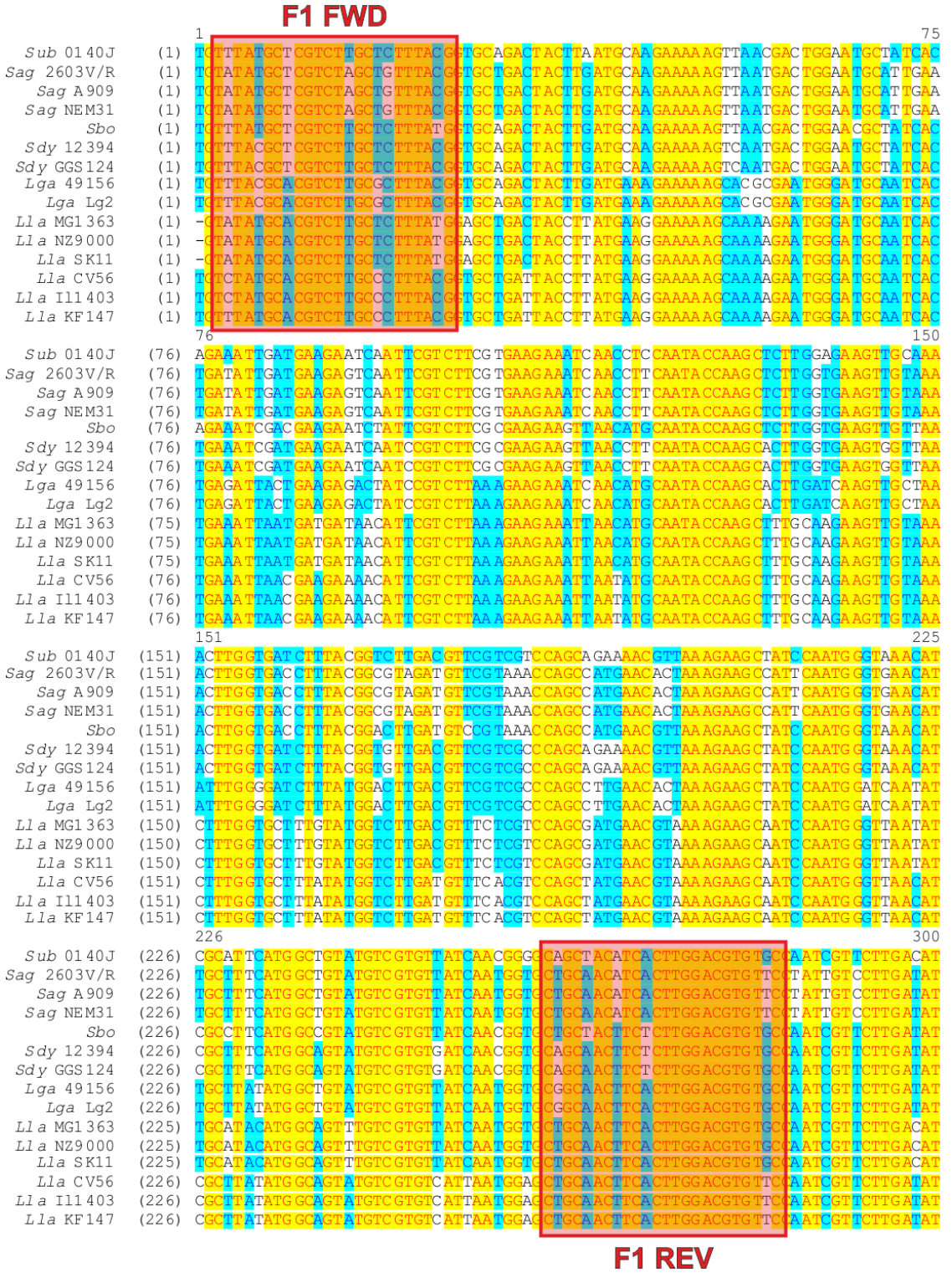


**Figure S21**. Primer design and alignment of the Streptococcaceae-specific marker (F1) with available genomes from the NCBI database. *Sub* = *S. uberis*; *Sag* = *S. agalactiae*; *Sbo* = *S. bovins*; *Sdy* = *S. dysgalactiae*; *Lga* = *L. garviae* and *Lla* = *L. lactis.* Yellow indicates base identity between all tested strains, while blue is indicative of the presence of at least one difference in the alignment.


**Figure S22**. Primer design and alignment of the *Lactococcus*-specific marker (LC2) with available genomes from the NCBI database.. *Lla* = *L. lactis* and *Lga* = *L. garviae*. Yellow indicates base identity between all tested strains, while blue is indicative of the presence of at least one difference in the alignment.


**Figure S23**. Primer design and alignment of two *Streptococcus*-specific markers (ST1 and ST2) with available genomes from the NCBI database. *Sub* = *S. uberis*; *Sag* = *S. agalactiae*; *Sdy* = *S. dysgalactiae*; *Spn* = *Streptococcus pneumoniae* and *Ste* = *Streptococcus termophilus.* Yellow indicates base identity between all tested strains, while blue is indicative of the presence of at least one difference in the alignment.

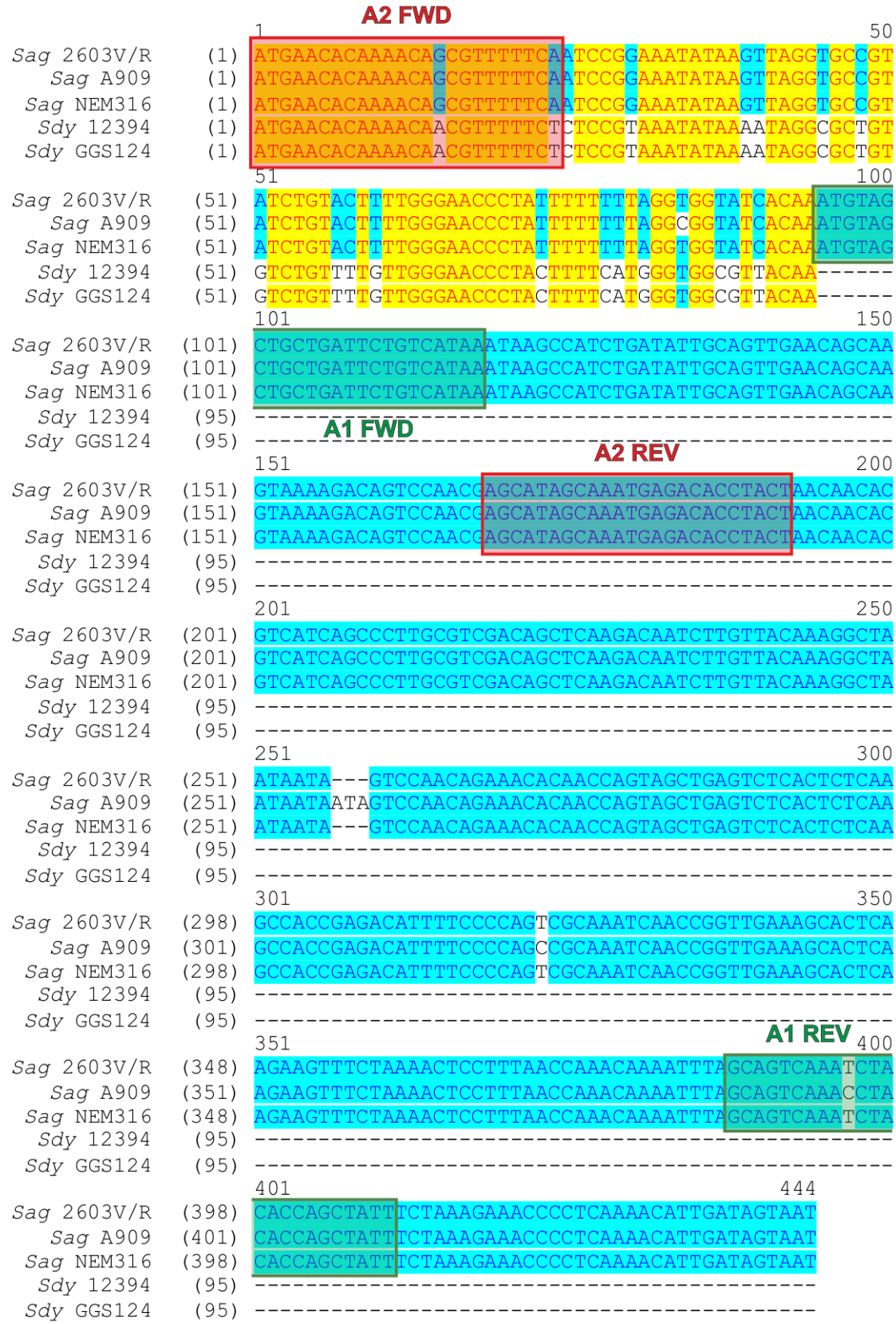


**Figure S24**. Primer design and alignment of two *Streptococcus agalactiae*-specific markers (A1 and A2) with available genomes from the NCBI database. *Sag* = *S. agalactiae* and *Sdy* = *S. dysgalactiae*. Yellow indicates base identity between all tested strains, while blue is indicative of the presence of at least one difference in the alignment.

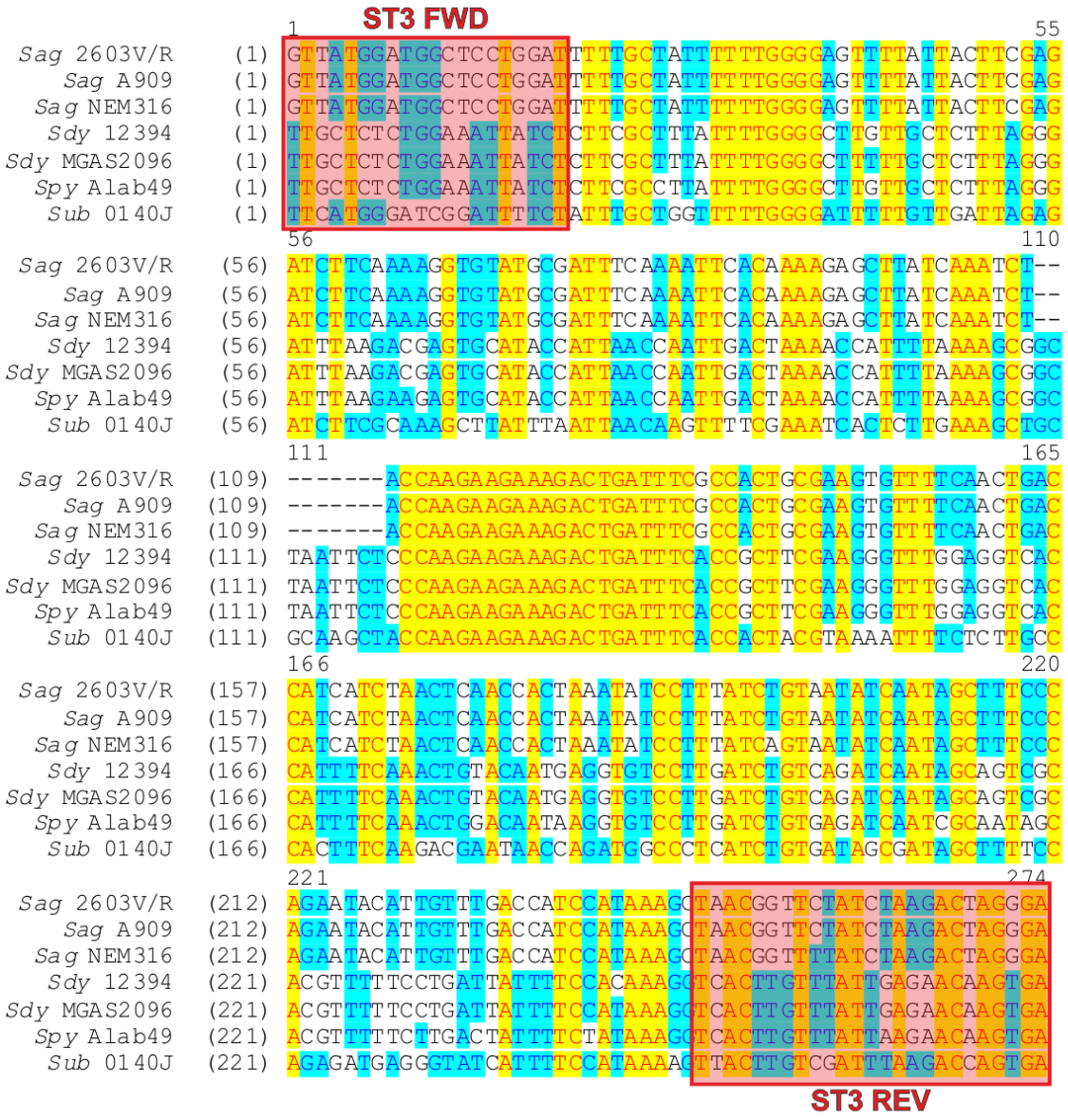


**Figure S25**. Primer design and alignment of a *Streptococcus*-specific marker (ST3) with available genomes from the NCBI database. *Sag* = *S. agalactiae*; *Sdy* = *S. dysgalactiae*; *Spyo* = *Streptococcus pyogenes* and *Sub* = *S. uberis*. Yellow indicates base identity between all tested strains, while blue is indicative of the presence of at least one difference in the alignment.