Institute of Medical Microbiology and Hygiene
Ulm University Medical Center
Prof. Dr. med. Steffen Stenger

Design of a molecular method to detect *Streptococcus gallolyticus* in human stools: prevalence of the bacteria in healthy and colorectal conditions

Dissertation
to obtain the Doctoral Degree of Medicine (Dr. med.)
at the Faculty of Medicine, Ulm University

Julian Lichtl-Häfele
Kirchheim unter Teck
2020
Acting dean: Prof. Dr. rer. nat. Thomas Wirth

1. examiner: Prof. Dr. med Barbara Spellerberg
2. examiner: Prof. Dr. rer. med. Alexander Kleger

Doctorate awarded: December 17th, 2021
Parts of this dissertation have already been published in the following journal article:

# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABBREVIATIONS</td>
<td>V</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 The gastrointestinal tract and the intestinal microbiome</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Colorectal cancer and microbes</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Colorectal cancer</td>
<td>2</td>
</tr>
<tr>
<td>1.2.2 Microbes and colorectal cancer</td>
<td>5</td>
</tr>
<tr>
<td>1.3 <em>Streptococcus bovis/Streptococcus equinus</em> complex (SBSEC)</td>
<td>7</td>
</tr>
<tr>
<td>1.3.1 Characteristics of SBSEC</td>
<td>7</td>
</tr>
<tr>
<td>1.3.2 The pathogenic potential of SBSEC</td>
<td>9</td>
</tr>
<tr>
<td>1.3.3 Methods used to identify <em>S. galolyticus</em> subsp. <em>galolyticus</em> (SGG)</td>
<td>10</td>
</tr>
<tr>
<td>1.3.4 <em>SGG</em> versus colorectal cancer</td>
<td>11</td>
</tr>
<tr>
<td>1.3.5 <em>SGG</em> as an opportunist passenger in colorectal cancer</td>
<td>11</td>
</tr>
<tr>
<td>1.3.6 <em>SGG</em> as a promoting agent for colorectal cancer</td>
<td>12</td>
</tr>
<tr>
<td>1.3.7 <em>SGG</em> plays a dual role in colorectal cancer</td>
<td>12</td>
</tr>
<tr>
<td>1.4 Aim of the present study</td>
<td>13</td>
</tr>
<tr>
<td>2 MATERIAL AND METHODS</td>
<td>15</td>
</tr>
<tr>
<td>2.1 Material</td>
<td>15</td>
</tr>
<tr>
<td>2.1.1 Cohort and Patient material</td>
<td>15</td>
</tr>
<tr>
<td>2.1.2 Oligonucleotides</td>
<td>15</td>
</tr>
<tr>
<td>2.2 Methods</td>
<td>17</td>
</tr>
<tr>
<td>2.2.1 DNA isolation methods</td>
<td>17</td>
</tr>
<tr>
<td>2.2.2 Amplification and analysis of DNA</td>
<td>17</td>
</tr>
<tr>
<td>3 RESULTS</td>
<td>19</td>
</tr>
<tr>
<td>3.1 Methodology</td>
<td>19</td>
</tr>
<tr>
<td>3.1.1 Selection of oligonucleotides to identify <em>S. galolyticus</em> subsp. (SG)</td>
<td>19</td>
</tr>
</tbody>
</table>
**TABLE OF CONTENTS**

3.1.2 DNA isolation of clinical fecal samples.................................................. 29  
3.1.3 PCR in combination with a RFLP assay .............................................. 30  
3.1.4 Real-time PCR / PCR product nucleotide sequence assay ................. 34  
3.2 The prevalence of *S. gallolyticus* among clinical samples.................. 38  
3.2.1 General characteristics of our cohorts............................................ 38  
3.2.2 Detection of *S. gallolyticus* member bacteria in fecal DNA............ 40  

4 DISCUSSION........................................................................................................ 46  
4.1 The prevalence of *S. gallolyticus* infections ........................................ 47  
4.1.1 *S. gallolyticus* subsp. *gallolyticus* ................................................. 47  
4.1.2 *S. gallolyticus* subsp. *macedonicus* and subsp. *pasteurianus* ...... 49  
4.1.3 Simultaneous presence of *S. gallolyticus* subsp......................... 50  
4.2 Review about Series A and Series B ..................................................... 50  
4.3 Outlook......................................................................................................... 51  

5 SUMMARY.......................................................................................................... 53  

6 REFERENCES..................................................................................................... 55  

7 APPENDIX.......................................................................................................... 66
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>°C</td>
<td>degree Celsius</td>
</tr>
<tr>
<td>AOM</td>
<td>azoxymethane</td>
</tr>
<tr>
<td>ASRi</td>
<td>age-standardized incidence rate</td>
</tr>
<tr>
<td>BMI</td>
<td>body mass index</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>CFU</td>
<td>colony forming unit</td>
</tr>
<tr>
<td>CIN</td>
<td>chromosomal instability</td>
</tr>
<tr>
<td>CN</td>
<td>control</td>
</tr>
<tr>
<td>COX-2</td>
<td>cyclooxygenase 2</td>
</tr>
<tr>
<td>CRA</td>
<td>colorectal adenoma</td>
</tr>
<tr>
<td>CRC</td>
<td>colorectal cancer</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double stranded deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>eg.</td>
<td>exempli gratia</td>
</tr>
<tr>
<td>et al.</td>
<td>et alii</td>
</tr>
<tr>
<td>ETBF</td>
<td>Enterotoxigenic <em>Bacteroides fragilis</em></td>
</tr>
<tr>
<td>FAP</td>
<td>familial adenomatous polyposis</td>
</tr>
<tr>
<td>gDNA</td>
<td>genomic deoxyribonucleic acid</td>
</tr>
<tr>
<td>HNPCC</td>
<td>hereditary nonpolyposis colorectal cancer</td>
</tr>
<tr>
<td>IBD</td>
<td>inflammatory bowel disease</td>
</tr>
<tr>
<td>IGFβ2R</td>
<td>insulin like growth factor beta 2 receptor</td>
</tr>
<tr>
<td>IL-1/8</td>
<td>interleukin-1/8</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>matrix-assisted laser desorption/ionization time of flight</td>
</tr>
<tr>
<td>min</td>
<td>minutes</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MSI</td>
<td>microsatellite instability</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>NSAID</td>
<td>nonsteroidal anti-inflammatory drug</td>
</tr>
<tr>
<td>PCNA</td>
<td>proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PKS</td>
<td>polyketide synthase</td>
</tr>
<tr>
<td>RFLP</td>
<td>restriction fragment length polymorphism</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>real-time polymerase chain reaction</td>
</tr>
<tr>
<td>SBSEC</td>
<td><em>Streptococcus bovis/Streptococcus equinus</em> complex</td>
</tr>
<tr>
<td>SGMB</td>
<td><em>Streptococcus gallolyticus</em> member bacteria</td>
</tr>
<tr>
<td>SMAD2/4</td>
<td>signal proteins for the tumor growth factor beta</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphisms</td>
</tr>
<tr>
<td>ssDNA</td>
<td>single stranded deoxyribonucleic acid</td>
</tr>
<tr>
<td>TGFβ2R</td>
<td>tumor growth factor beta 2 receptor</td>
</tr>
<tr>
<td>Tis</td>
<td>carcinoma in situ</td>
</tr>
<tr>
<td>TP53</td>
<td>tumor suppressor protein 53</td>
</tr>
<tr>
<td>UICC</td>
<td>Union of International Cancer Control</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet light</td>
</tr>
<tr>
<td>V</td>
<td>volt</td>
</tr>
<tr>
<td>VIS</td>
<td>visual light</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>WNT</td>
<td>signaling transduction pathway</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 The gastrointestinal tract and the intestinal microbiome

The central role of the gastrointestinal tract is the transformation of ingested nourishment into absorbable components. This process is divided in several steps taking place in different segments of the alimentary system and requires the usage of many different physical, chemical and biological components as well as the supply of associated organs.

One mandatory element for a functional digestive system is the commensal intestinal microbiota. This microbiological community varies heavily in bacterial composition and numbers along the intestinal axis, thereby mediating several key functions. Identified species exceed the 2000 (Hugon P et al., 2015; Li J et al., 2014), coming together in an estimated number with up to $10^{14}$ cells, a tenfold of human cells in a body (Backhed et al., 2005; Gill et al., 2006). The commensal bacteria are distributed among nine phyla with Firmicutes, Bacteroidetes and Actinobacteria representing approximately 75% of the microbial diversity (Peris-Bondia et al., 2011). Certainly, the gut microbiota from one individual to another shows differences and is constantly in exchange to keep its balance known as microbiota homeostasis.

A prominent role of gut microbiota is the transformation and provision of dietary nutrients. Other roles include the interaction with endogenous systems to maintain the metabolic and immune homeostasis, the integrity of intestinal epithelia by provoking its renewal, a protection role against pathogens (barrier effect) and serving the digestive peristaltic as a stimulator (Thursby and Juge, 2017).

Due to changes of environment, dietary habits, extrinsic pathogens and related intrinsic diseases, generalized use of antibiotics, the commensal intestinal microbiota is facing a number of individual challenges during human lifespan (Rodriguez et al., 2015; Koenig et al., 2011). However, the flexibility of the gut
microbiota has its limits and several members of the bacterial community are known and suspected to become opportunistic pathogens or pathobionts. Pathobionts live in symbiosis with the host under regular circumstances but can potentially turn into pathogens, causing disease under specific conditions. These promoting conditions include immunosuppression of the host, increased bacterial virulence or local factors that enhance host-pathogen interactions. Obligate pathogens form a separate group which constitutes a constant challenge for the defense mechanisms of the host organism.

However, all intestinal pathogens can generate a broad range of gastrointestinal infections with many different clinical courses. Most of them are self-limiting, causing temporary gastroenteritis of different severity with symptoms such as nausea, vomiting, diarrhea and abdominal pain. Others such as *Helicobacter pylori* are known to colonize the stomach on a long term, which can provoke somatic mutations and epigenetic modifications resulting in gastric cancer. Beyond local pathogenic actions, some pathogens can translocate across the intestinal barrier and spread in the blood resulting in generalized infections.

### 1.2 Colorectal cancer and microbes

#### 1.2.1 Colorectal cancer

Colorectal cancer (CRC) is the third most common cancer accounting for 10% of cancer related mortality in western countries today (Sobhani *et al.*, 2013; Boyle and Leon, 2002). Prognoses estimate colorectal tumors to increase by 60% in developed countries by 2030 (Arnold *et al.*, 2017). The age-standardized incidence rate (ASRi) of colorectal cancer is higher in men (20.6 per 100,000 individuals) than in women (14.3 per 100,000). While a hereditary form is rarely seen in younger patients, the risk of sporadic occurrence for both sexes increases crucially with age > 50 years old (Kuipers *et al.*, 2016).

The current knowledge indicates that both genetic and environmental factors play an important role in the etiology of somatic and epigenetic changes that lead to colorectal carcinomas. The cellular changes from healthy to neoplastic epithelia underlie a strict order of genetical changes. These stepwise mutational
activation of protooncogenes and inactivation of tumor suppressor genes is known as the adenoma-carcinoma sequence. The benign hyperproliferation to small, large and dysplastic adenomatous neoplastic polyps is comparatively slow and can take up to twenty years silently without any perceivable symptoms. Then suddenly, the evolution of the tumor accelerates, becomes malignant, and then can scatter through metastasis. Under certain genetic conditions, this time frame can be much narrow (Jones et al., 2008). The general lifetime risk in western populations for developing colorectal cancer is 3-5%. It is constituted predominantly of sporadic cases with three quarters of patients showing a negative family history. However, the risk increases if close family members had been diagnosed earlier. Two major mechanisms are indicated for the sporadic carcinogenesis: 1) chromosomal instability (CIN) and 2) microsatellite instability (MSI) (Figure 1).

![CIN - Chromosomal Instability pathway](image)

**CIN - Chromosomal Instability pathway**

1. APC inactivation
2. KRAS mutation
3. 18q, SMAD2/4 loss
4. TP53 mutation and loss

Normal mucosa → Early adenoma → Intermediate adenoma → Late adenoma → Carcinoma

**MSI - Microsatellite Instability pathway**

MMR mutation; MLH1 methylation → BRAF mutation → TGFβR2, IGF2R, BAX mutation

**Figure 1**  **Conventional adenoma - carcinoma sequence.** The chromosomal instability (CIN) pathway starts with mutations in the tumor suppressor gene *APC* (adenomatous polyposis coli) located in the normal colonic mucosa. The tissue dysplasia progresses from adenoma to carcinoma while additional mutations in the proto-oncogene *KRAS* and regulator proteins *SMAD2/4* and *TP53* lead to procarcinogen dysregulations. The microsatellite instability pathway (MSI) begins with mutations in the DNA mismatch repair (MMR) gene and MLH1 methylation, followed by mutations and dysregulations of *BRAF, TGFβR2, IGF2R* and *BAX* genes while the genesis of the tumor progresses (De Palma et al., 2019).
CIN are detected in 85% of sporadic colorectal cancer and is the most common form of genomic instability. Acquired mutations in the adenomatous polyposis coli (APC) tumor suppressor gene and the protooncogene KRAS initiate other alterations in regulating proteins SMAD2/4 and TP53 while the cell dysplasia progresses from normal mucosa to carcinomas. MSI occurs in 15% and is dominated by mutations in the mismatch repair (MMR) gene and methylation mechanisms in the MLH1 gene. In the progress of the adenoma–carcinoma sequence, mutations and dysregulations in BRAF, TGFβR2, IGF2R and BAX lead to tumor growth (De Palma et al., 2019; Gupta et al., 2018; Pino et al., 2010; Geigl et al., 2008).

A smaller group of patients (5-10%) show hereditary colorectal cancer syndromes that tremendously accelerate carcinogenic pathways. Two dominant hereditary disorders have been reported. 1) Hereditary nonpolyposis colorectal cancer (HNPCC) best known as Lynch syndrome, which is due to massively increased mutations in DNA mismatch repair genes (e.g., MSH2, MLH1, MSH6, PMS2, and PMS1) that lead to the accumulation of microsatellites and subsequent carcinogenesis. 2) Familial adenomatous polyposis (FAP) which results from the complete inactivation of the APC tumor suppressor gene and the accelerated activation of the WNT pathway (Vasen et al., 2015; Pérez-Carbonell et al., 2012; Van Lier et al., 2012). FAP patients tend to develop a high number of colorectal carcinomas at young age (Kuipers et al., 2016).

In addition to these genetic factors, environmental and diet factors are well known to promote the development of colorectal cancer. The major risks are smoking, alcohol intake, consumption of red and processed meat, as well as increased body weight. Every unit ascent in the body mass index increases the risk of colorectal cancer by 2-3%. Furthermore, inflammatory and metabolic diseases such as inflammatory bowel diseases (IBD) and type 2 diabetes appear to have a prooncogenic impact as well (World Cancer Research Fund, Continuous Update Project Report, 2018).
1.2.2 Microbes and colorectal cancer

The significance of a few microbes as promotive agents in colorectal cancer have been reported in several studies. With $10^9$ to $10^{12}$ bacterial cells/ml in the large intestine versus $10^3$ to $10^8$ bacterial cells/ml in the small intestine, a correlation between the density of the microbiological flora and the nearly hundredfold incidence of colorectal cancer has been postulated (Jemal et al., 2009).

The microbiome is now considered a prime suspect for triggering the initiation and/or progression of colonic carcinogenesis. The development of new detection methods (16s rRNA gene sequencing, metagenomics, transcriptomics, proteomics, metabolomics) has led to an incredible improvement in defining the composition and function of the intestinal microbiome. However, the variation of the microbiome along the colonic axis and between intraluminal and mucosal microbe communities have made the listing of a complete species catalogue difficult (Lagier et al., 2012, Sobhani et al., 2011). Nevertheless, several strains such as reactive oxygen species producing Enterococcus faecalis, colibactin-producing Escherichia coli, enterotoxigenic Bacteroides fragilis (ETBF) and Streptococcus galloyticus subspecies galloyticus (SGG) are strongly associated with colorectal cancer (Pasquereau-Kotula et al., 2018). Furthermore, the enrichment of two oral bacteria, Fusobacterium nucleatum and Parvimonas micra is repeatedly reported in the intestinal metagenomics of CRC patients (Whitmore and Lamont, 2014). The molecular mechanisms underlying the promotion of colonic cell proliferation have not been completely elucidated today.

E. faecalis strains, for example differ in their ability to produce radical oxygen species. Those producing superoxide anions seem to induce DNA damage and genomic instability which might leads to the initiation of colorectal cancer. Experiments suggested that E. faecalis carcinogenesis can be mediated through triggered mucosal macrophages producing diffusible clastogens (chromosomal-breaking-factors), inducing DNA damage resulting in uncontrolled cell proliferation (Wang et al., 2008; 2012; Yang et al., 2013).

Most E. coli subtypes are commensal members of the intestinal microbiome. The phylogenetic group B2 E. coli however, is a CRC promoting suspect. A study
postulated a prevalence of 55% in CRC patients (Buc et al., 2013). B2 E. coli is able to produce colibactin, a genotoxic molecule through its polyketide synthase (PKS) island (Cuevas-Ramos et al., 2010; Nougayrède et al., 2006). Colibactin is able to induce double-strand DNA breaks, mutations and DNA rearrangements which then lead to a c-Myc mediated cell cycle arrest and cellular senescence. Excessed cellular senescence results in a strong upregulation of growth factors and consequent cell growth (Cougnoux et al., 2014).

Another suspect bacterium is B. fragilis producing BFT/fragilysin toxin. This toxin is a zinc-dependent metalloprotease which cleaves the adherence junction protein E-cadherin resulting in β-catenin release and thus activation of Wnt/β-catenin signaling pathway, leading to protooncogene c-Myc expression and dysregulated colonic cell formation (Soler et al., 1999; Wu et al., 2003).

A third suspect is Fusobacterium nucleatum a Gram-negative anaerobic bacterium which is found generally in the oral cavity. Several studies have now established the high abundance of this bacterium in CRC microbiota. F. nucleatum displays a unique surface adherence protein FadA is able to bind to the extracellular domain of an E-cadherin leading to enhanced β-catenin/Wnt signaling pathway, which in turn leads to increased cell proliferation and tumor growth (Rubinstein et al., 2013). Very interestingly, F. nucleatum is also able to suppress anti-tumoral immunity through Fap2/ TIGIT interaction (Gur et al., 2015).

Two types of mechanisms have emerged to explain how bacteria contribute to the development of human colorectal cancer. The first mechanism derived from E. faecalis and E. coli indicates carcinogenesis induction through DNA-damage and/or interference with DNA-repair mechanisms. The second mechanism derived from ETBF, F. nucleatum and SGG points towards Wnt/β-catenin signaling pathway which leads to cell hyperproliferation and thus neoplasia. Furthermore, it has been shown that close contact between bacteria and host tissues is necessary for induction of oncogenic modifications (Soler et al., 1999; Grivennikov et al., 2012).

On the contrary, there are life style habits exercising a protective effect against CRC development such as consumption of milk, whole grains, fresh fruits and vegetables, as well as intake of calcium, fiber, multivitamins and vitamin D,
physical activity and frequent intake of low-dose NSAIDs (COX-2 inhibitors) (Song et al., 2015; Dahm et al., 2010).

The onset of symptoms such as bloody stools, abdominal pain, appears quite late in CRC development compromising a favorable outcome. Therefore, it seems very important to find innovative tools and strategies for the early detection of CRC, for example through molecular detection of key bacterial species often associated with CRC development.

1.3 *Streptococcus bovis/Streptococcus equinus complex* (SBSEC)

1.3.1 Characteristics of SBSEC

The *Streptococcus bovis*/Streptococcus equinus complex (SBSEC) includes several species colonizing the animal and human gastrointestinal tracts. SBSEC belong to the group D streptococci according to Rebecca Lancefield’s classification (Lancefield, 1933). They consist of Gram-positive *cocc**i* growing in pairs or chains (*Figure 2*). They are catalase and oxidase negative, non-motile, non-sporulating, lactic acid producing cocci. It is a large group containing commensal bacteria used in fermentation of dairy products as well as opportunistic pathogens for animals and humans.

![Figure 2](image_url)
Over a decade ago, the taxonomy of the group was revised (Poyart et al., 2002; Schlegel et al., 2003). The former classification system distinguished between *Streptococcus bovis* biotype I able to use mannitol and *Streptococcus bovis* biotype II/1 and biotype II/2 unable to use this sugar as carbon source (Schlegel et al., 2003; Jans et al., 2015).

Schlegel et al. proposed to rename the various biotypes as indicated in Figure 3, *Streptococcus gallolyticus* subspecies *gallolyticus* (biotype I), *Streptococcus infantarius* subspecies *infantarius* (biotype II/1), *Streptococcus infantarius* subspecies *coli* (biotype II/1), and *Streptococcus gallolyticus* subspecies *pasteurianus* (biotype II/2), based on molecular characteristics (Schlegel et al., 2000, 2004; Beck et al., 2008).

The lack of a uniform nomenclature in the literature complicates the comparison between various studies (Poyart et al., 2002; Schlegel et al., 2003; Jans and Boleij, 2018).

SBSEC are detected as commensals in a various number of species colonizing the rumen, crop and cloaca of animals. It is found in ruminants and other herbivore mammals including livestock and companion animals and also in
INTRODUCTION

birds such as turkeys, pigeons and chickens (De Herdt et al., 1994 a,b; Schulz et al., 2015). In the food chain, SBSEC is present in various fermented dairy products (Jans et al., 2011).

Humans also carry a number of SBSEC members, known to colonize the gastrointestinal tract as commensals. The fecal carriage of SBSEC ranges between 5 - 60%. The prevalence in humans is still controversial probably mostly depending on the detection methods used (Klein et al., 1977; Noble, 1978; Potter et al., 1998; Abdulamir et al., 2010; Al-Jashamy et al., 2010; Chirouze et al., 2013; Lopes et al., 2014; Boltin et al., 2015; Dumke et al., 2017; Kaindi et al., 2017).

The route of transmission is presumed to be zoonotic meaning from animal to humans as an increased carriage prevalence has been demonstrated in rural living communities (Corredoira et al., 2015; Giannitsioti et al., 2007; Jans and Boleij, 2018).

1.3.2 The pathogenic potential of SBSEC

Most SBSEC strains are regarded as safe inhabitants of the digestive system in animals and humans but some of them are known to cause severe diseases such as ruminal acidosis in herbivores or infective endocarditis, meningitis and septicemia in the elderly and immunocompromised humans (Chow et al., 2011; Boleij and Tjalsma, 2013; Jans et al., 2015, 2016). Furthermore, studies have associated S. bovis biotype I frequently with colorectal cancer for more than sixty years, suggesting a promoting role in the disease. Indeed, McCoy and Mason were the first to report this association in 1951 (McCoy and Mason, 1951). Later on, Hoppes and Lerner confirmed this finding by showing that out of a group of 15 endocarditis patients infected with S. bovis, 9 (64%) had a concomitant colorectal neoplasia (Hoppes et al., 1974). Another study reported a 5-fold increased fecal carriage of S. bovis in CRC patients (Klein et al., 1977).

A literature survey, analyzing 52 case reports and 31 case series published in PubMed between 1970 – 2010 revealed a mean association of 65% between SGG infected patients and concomitant colorectal neoplasia (Boleij et al., 2011). SGG
could be found in the feces of about 75% of CRC patients, which is much higher than in healthy population (Klein et al., 1977; Abdulamir et al., 2010). The prevalence of \textit{SGG} under physiologic conditions in the human intestinal tract was considered low (2.5-15%) mostly based on culture techniques. A recent study reports a much higher carriage of 62.5% in the healthy German population based on a sensitive PCR technique (Dumke et al., 2017). Although the prevalence varies between studies, the link between \textit{SGG} and CRC is unambiguous. The high variability in the prevalence reported in the literature is probably a consequence of the different geographical locations where the samples were collected and the various methods (mostly culture based) used to identify \textit{SGG}.

1.3.3 Detection methods used to identify \textit{S. galloyticus subsp. galloyticus} (SGG)

The investigation of \textit{S. bovis} species in infective endocarditis and their first association with colorectal cancer was performed with classic enrichment cultures (McCoy and Mason, 1951). Patients suffering from infections with \textit{S. bovis} were often pretreated with antibiotics which effectively reduced the number of vital bacteria. As a consequence, the bacterial cultivation from a blood, tissue or fecal sample was difficult. The ability to identify bacteria based on lowest amounts of DNA helped molecular detection methods such as PCR analysis to outplay conventional microbiological techniques as detection methods (Holland et al. 2000, Dumke et al., 2017). The sequence analysis of 16S rRNA or the superoxide dismutase (\textit{sodA}) gene was often used for the identification by PCR (Sasaki et al., 2004). But while a molecular analysis in blood is easier, an analysis of feces is tremendously more challenging due to the incredible variety of the intestinal microbiota. To investigate different human reservoirs, SBSEC strains were as well included in analyses with MALDI-TOF-MS (matrix-assisted laser desorption / ionization time of flight mass spectrometry) (Hinse et al., 2011). Simplified, this method is based on the detection of peptides and the comparison of their masses with a data bank to receive a specific identification of the bacteria (Dieckmann et al., 2008; Nagy
et al., 2009; Seng et al., 2009). As well specific qualitative microarrays for *S. gallolyticus* subsp. *gallolyticus* detection have been designed (Hinse et al., 2011). Although MALDI-TOF-MS and microarrays allow the simultaneous detection of different bacteria of the intestinal microbiota, PCR techniques still set the diagnostic standard for high sensitivity and specificity (Ngassam Tchamba et al., 2019).

### 1.3.4 SGG versus colorectal cancer

The main pending question about *SGG* association with colorectal cancer is, whether *SGG* benefits from the alternate tumor environment to colonize the host colon at higher density or whether *SGG* can promote cell proliferation under certain conditions.

### 1.3.5 SGG as an opportunist passenger in colorectal cancer

The colorectal environment changes under cancerous conditions that can benefit to *SGG*. First, the increased glycolysis in tumor epithelia cells produces several metabolites, notably F6P, 3PG and alanine, which are efficiently metabolized by *SGG*. Processing these particular metabolites allows an increased *SGG* multiplication and gives a competitive advantage over other species (Boleij et al., 2012). Secondly, *SGG* is able to defeat close competitors, such as close related *Enterococci* commensals by producing a specific bacteriocin, named gallocin. Gallocin activity is strongly increased in the presence of secondary bile salts such as deoxycholic and lithocholic acid, which are found to be increased in CRC conditions (Aymeric et al., 2017). Thirdly, *SGG* possesses surface pili mediating interactions with host tissues. In particular, *Pil1* allows the binding to collagen (Danne et al., 2011) while *Pil3* mediates the binding to colonic mucins (Martins et al., 2015). It is tempting to speculate that cancerous mucins may represent better substrates for *SGG* adhesion and colonization, a hypothesis which is currently tested in the laboratory.

These three features of improved tumor-metabolite processing, bacteriocin-dependent competitor decimation and pilus mediated bacteria-tumor-tissue
adherence form a triangle that seems to promote SGG replication, particularly in tumorous conditions (Pasquereau-Kotula et al., 2018).

1.3.6 SGG as a promoting agent for colorectal cancer

In 2010, an important study showed that SGG accumulates in neoplastic tissue, displaying a much higher density on the tumor sites as compared to adjacent healthy tissues in the same patient. SGG colonization was correlated to a higher mRNA expression of genes encoding proinflammatory molecules such as IL-1, COX-2, IL-8. This led to the assumption that the inflammatory response could favor tumor progression (Abdulamir et al., 2010). More recently, another research group was able to demonstrate the long suspected pro-oncogenic potential of some SGG isolates (TX20005) using in vitro human cancer cell lines and murine models (Kumar et al., 2017).

This study reports that SGG TX20005 is able to increase cell proliferation in human colon cancer cell lines (HCT116, HT29, LoVo) but not in other colon cell lines (SW480, SW1116), nor in normal colonic cells (CCD841 CoN, FHC). Responsive cells infected with TX20005 displayed an increased expression of β-catenin, resulting in upregulation of its downstream targets (c-Myc/cyclin D) and proliferating cell nuclear antigen (PCNA). The proliferation effect correlates with adherence of SGG to the human colon cancer cells lines suggesting the necessity of a close bacteria-cell interactions. Azoxymethane (AOM) is a carcinogenic compound used in biological research to induce colon cancer (Sohn et al., 2001). In AOM · induced CRC murine models, chronic oral gavage for 12 weeks with SGG TX20005 induced a higher number of tumors as compared to mice infected with Lactococcus lactis (Kumar et al., 2017). It should be noted that the promoting effect of SGG is restricted to pre-transformed cells.

1.3.7 SGG plays a dual role in colorectal cancer

Taken together these data indicate that SGG benefits from the tumorous conditions but also it can also accelerate the development of colorectal cancer under certain conditions (Pasquereau-Kotula et al., 2018) (Figure 4).
Figure 4 Two working models explaining the dual association between *Streptococcus gallolyticus* subspecies *gallolyticus* (SGG) and colorectal cancer. 1) *SGG* as a passenger bacterium: in preneoplastic epithelium, the activation of the WNT signaling transduction pathway leads to the downregulation of the bile acids transporter Slc10A2 and a subsequent accumulation of bile acids. Accumulated bile acids seem to enable *SGG* to unleash a bacteriocin, killing related commensals (e.g. *Enterococci*). As a driver bacteria, high colonization of *SGG* individuals during premalignant lesion can cause an inflammatory response (IL-1, COX-2, IL-8) and cell proliferation mediated through upregulation of β-catenin and its oncogenic downstream targets (*C-Myc* and *cyclin D*) which might accelerates the adenoma → carcinoma sequence (Pasquereau-Kotula *et al.*, 2018)

1.4 Aim of the present study

SBSEC harbors three subspecies *S. gallolyticus* subspecies *gallolyticus* (*SGG*), *S. gallolyticus* subspecies *macedonicus* (*SGM*) and *S. gallolyticus* subspecies *pasteurianus* (*SGP*). Majorly *SGG* has been associated with colorectal cancer for a long period and was recently confirmed to have an accelerating effected on colorectal tumor development. To date, literature data about the prevalence of *SG* subspecies in healthy and diseased populations is strongly heterogenic due to the usage of different detection methods. A better knowledge about the distribution of *SG* subspecies in the population is required to gain better insights about its pathologic potential. Therefore, the aim of the present study was the development of a convenient and sufficient detection method for the
presence of the respective microbes in human feces and its application to investigate the feces of a clinical representative cohort. This experimental work investigated the following points:

- Setup of a molecular biological detection protocol for the sufficient identification of *SG* subspecies (*SGG, SGM, SGP*) in human feces.
- Investigation of *SG* subspecies (*SGG, SGM, SGP*) in the feces of a cohort consisting of CRC, colorectal adenoma (CRA) patients and healthy volunteers.
- Comparison of the results (data, method choice) with previous published studies.
2 MATERIAL AND METHODS

2.1 Material

2.1.1 Cohort and Patient material

The collection of fecal samples from colorectal healthy controls, colorectal adenoma and colorectal cancer patients was provided by professor Iradj Sobhani, director of the liver and gastro-intestinal department at the Henri Mondor hospital Creteil-APHP in Paris. The usage of samples for corresponding experiments has been approved by the French ethics committee. Respective statements are registered under EGY/FLR/AR103592 (08/07/2010), AOM09268/N108021 (15/11/2011) and AOM09268/CPP10-006 (31/10/2013).

The initial cohort included the stool samples from 93 volunteers, constituted of 62 negative (non-CRC) and 31 colorectal cancer positive (CRC) volunteers collected between 2011 - 2016. Patients with stool samples below the required weight of 150 mg for the DNA extraction were discarded. As well a few samples that failed initial RT-PCR qualitative and quantitative tests were excluded. The final results concern 71 participants in total with 48 colorectal cancer negatives and 23 colorectal cancer positive patients.

2.1.2 Oligonucleotides

All oligonucleotides (*Table 1, Table 2*) were purchased from Eurofins Genomics, les Ullis, France, diluted to 20 picomol/µl and stored at −20°C.
### Table 1  Oligonucleotides used for classic polymerase chain reaction in our study.

(SBSEC) *Streptococcus bovis/Streptococcus equinus* complex, (bp) base pair

<table>
<thead>
<tr>
<th>Target</th>
<th>Primer</th>
<th>Amplification length (bp)</th>
<th>Primer forward (5’-3’)</th>
<th>Primer reverse (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gramm positive bacteria</td>
<td>AllBacteria</td>
<td>146</td>
<td>CGGTGAATACGTTCCCCG</td>
<td>AAGTCGTAACAGGATACCCGTA</td>
</tr>
<tr>
<td>SBSEC</td>
<td>BIS Gallo 2020/2021</td>
<td>497</td>
<td>TGGGGAGGTATTTAATAATGATAAC</td>
<td>CTATAAATATCTGAAAATACCTGA</td>
</tr>
<tr>
<td>SBSEC</td>
<td>SodA 1/2</td>
<td>408</td>
<td>CAATGACAAATTCCCACATGAG</td>
<td>CACAGGAAAIGACCAAA</td>
</tr>
<tr>
<td>SBSEC</td>
<td>16S-SBSEC-FW</td>
<td>1119/20</td>
<td>GTAACGACATTTAACACCATGTTAG</td>
<td>16S-inf-rev</td>
</tr>
<tr>
<td>SBSEC</td>
<td>16S-SBSEC-3-FW</td>
<td>1119</td>
<td>CATACAGTGTITAAAACATGTTAG</td>
<td>16S-inf-rev</td>
</tr>
<tr>
<td>SBSEC</td>
<td>16S-SBSEC-4-FW</td>
<td>1119/20</td>
<td>GCATAATAGTTTITACATGTTAG</td>
<td>16S-inf-rev</td>
</tr>
<tr>
<td>SBSEC</td>
<td>16S-SBSEC-5-FW</td>
<td>1119/20</td>
<td>GTAACAGCTTTTTGACACATGTTAG</td>
<td>16S-inf-rev</td>
</tr>
<tr>
<td>SBSEC</td>
<td>16S-inf-rev</td>
<td></td>
<td></td>
<td>CCGCAGCAATCTCCACTAAAG</td>
</tr>
<tr>
<td>SBSEC</td>
<td>Gallo 2020/2021</td>
<td>1353</td>
<td>ATCTGTCAGTTGCAACGAC</td>
<td>GCCATAAATATGAATCTG</td>
</tr>
<tr>
<td>SBSEC</td>
<td>qPCR COMX</td>
<td>182</td>
<td>TGTTTCGTTCGTTCCTCCCT</td>
<td>TGACGTGCGGCAATTTG</td>
</tr>
<tr>
<td>SBSEC</td>
<td>Gallo2178</td>
<td>173</td>
<td>CACTATTGGACGTACCTGGTCGAT</td>
<td>CAATGAAAATGATCACGGGTG</td>
</tr>
</tbody>
</table>

### Table 2  Oligonucleotides used for real-time polymerase chain reaction in our study.

(SBSEC) *Streptococcus bovis/Streptococcus equinus* complex, (bp) base pair

<table>
<thead>
<tr>
<th>Target</th>
<th>Primer</th>
<th>Amplification length (bp)</th>
<th>Primer forward (5’-3’)</th>
<th>Primer reverse (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gramm positive bacteria</td>
<td>AllBacteria</td>
<td>146</td>
<td>CGGTGAATACGTTCCCCG</td>
<td>AAGTCGTAACAGGATACCCGTA</td>
</tr>
<tr>
<td>SBSEC</td>
<td>SodA 1/2</td>
<td>408</td>
<td>CAATGACAAATTCCCACATGAG</td>
<td>CACAGGAAAIGACCAAA</td>
</tr>
<tr>
<td>SBSEC</td>
<td>SodA 3/4</td>
<td>334</td>
<td>GCATACATGTGCAAAAT</td>
<td>AGACGACTACACTGCAATG</td>
</tr>
<tr>
<td>SBSEC</td>
<td>16S-SBSEC-3-FW</td>
<td>1119</td>
<td>CATACAGTGTITAAAACATGTTAG</td>
<td>16S-inf-rev</td>
</tr>
<tr>
<td>SBSEC</td>
<td>16S-inf-rev</td>
<td></td>
<td></td>
<td>CCGCAGCAATCTCCACTAAAG</td>
</tr>
<tr>
<td>SBSEC</td>
<td>Gallo2178</td>
<td>125</td>
<td>ACTGTGAGAGGAGGTGAGTTAAG</td>
<td>TGACATCAGATCGCAGGTCAAC</td>
</tr>
<tr>
<td>SBSEC</td>
<td>Gallo2179</td>
<td>173</td>
<td>CACTATTGGACGTACCTGGTCGAT</td>
<td>CAATGAAAATGATCACGGGTG</td>
</tr>
</tbody>
</table>
2.2 Methods

2.2.1 DNA isolation methods

2.2.1.1 DNA extraction from bacteria cultures

Genomic DNA from bacteria cultures was prepared using the *DNeasy® Blood & Tissue Kit* (50) according to the manufacturers protocol. DNAs were verified through agarose gel electrophoresis and measured with the *NanoDrop™ 2000c Spectrophotometer*.

2.2.1.2 DNA extraction from fecal samples

The *Promega Wizard® Genomic DNA Purification Kit* was used to extract the DNA content from fecal samples. The protocol was improved and adapted from Ahlroos and Tynkkynen (2009).

2.2.2 Amplification and analysis of DNA

2.2.2.1 Polymerase chain reaction methods

Classic and RT-PCR assays have been used with different settings in study. The basic set up are presented in the appendix.

2.2.2.2 Sequencing

Selected PCR products were controlled with agarose gel electrophoresis, purified with the *MP Biomedicals GENECLEAN® Turbo Kit* and finally send for nucleotide sequencing with the respective primers. The nucleotide sequences were analyzed and compared using the *ApE plasmid software* (2.0.53).
2.2.2.3 Restriction digest

The reaction mix contained the following components:

1. PCR-product 5 µl
2. FastDigest Green Buffer (10x) 1 x
3. Distilled DNase/RNase free Water 3.5 µl
4. Restriction enzyme 0.5 U

The sample was incubated for 30 min at 37°C. The whole sample was analyzed by agarose gel electrophoresis using a low voltage for a prolonged time (50V/45 min) to obtain an optimal resolution.

2.2.2.4 Agarose gel electrophoresis

Pure DNA, PCR and digested products were analyzed with agarose gel electrophoresis. DNA fragments were visualized with the Gel Doc XR+ System (Bio-Rad) and digital processed with the software application “Quantity One 4.6.2”.
3 RESULTS

3.1 Methodology

The aim of this study was to evaluate the prevalence of *S. gallolyticus* (*SGG*, *SGM*, *SGP*) in feces of a cohort consisting of colorectal cancer (CRC) patients, colorectal adenoma (CRA) patients and healthy volunteers (Périchon et al., 2022).

Fecal samples are quite difficult and heterogeneous specimens for DNA extraction and for DNA amplification. In particular, feces consistence depends on diet habits, metabolism and health status of the individual. The DNA content in feces shares eukaryotic- and prokaryotic origins as it derives from aliments, allogenic dead cells, and the individual human microbiome. Some complex plant polysaccharides can also interfere or inhibit conventional molecular techniques (Monteiro et al., 1997). Our protocol for the detection of *S. gallolyticus* subspecies in human feces was thus challenging and needed experimental optimization.

3.1.1 Selection of oligonucleotides to identify *S. gallolyticus*

A collection of oligonucleotide pairs (issued from previous publications or especially designed for this study) were tested (*Table 1* / *2*). Their specificity and sensitivity were determined by optimizing PCR conditions (*Table 3*). Based on this evaluation, the most satisfactory oligonucleotides were selected for the subsequent examination of the clinical fecal samples.

3.1.1.1 Determination of the oligonucleotide specificity

All oligonucleotides were submitted to same control tests. At first, the oligonucleotides specificity was virtually checked. The nucleotide sequences of
the PCR products were blasted on NCBI database to search for non-specific products. The virtual specificity was then controlled in wet laboratory. Oligonucleotides were first tested on genomic DNA from bacterial monocultures and mixed bacterial cultures and then on DNA isolated from feces.

Presented exemplarily is Gallo2179. The primer pair targets a SGG specific gene encoding the Pil1 pilus adhesin and was initially designed for RT-PCR experiments. The expected PCR product is 173 bp (Figure 5). A non-specific band of about 600 bp can be detected in S. gallolyticus subspecies macedonicus (well 7) and E. faecalis (well 11) as shown in Figure 5.

![Figure 5](image-url) Polymeric chain reaction analysis on genomic DNA from pure cultures using the oligonucleotides specific for Streptococcus gallolyticus subspecies gallolyticus (SGG) gene gallo 2179. A DNA fragment of 173 base pairs (bp) was expected for SGG samples marked in yellow. (M) 1 kilo base plus size marker, (1) negative control (H₂O), Genomic DNA extracted from (2) Bacteroides fragilis, (3) Parvimonas micra, (4) Fusobacterium nucleatum, (5) SGG, (6) SGG, (7) Streptococcus galloyticus subspecies macedonicus, (8) Streptococcus galloyticus subspecies pasteurianus, (9) Streptococcus infantarius subspecies infantarius, (10) Streptococcus infantarius subspecies coli (lutetiensis), (11) Enterococcus faecalis, (12) SGG strain UCN34 positive control.
After initial testing on pure genomic DNA, the oligonucleotides were tested on DNA extracted from feces. We started with feces of healthy volunteers that were tested negative for *SG*, that were deliberately contaminated with *SGG*. The best available method to confirm *SG* negativity at the time was *SodA1/2* PCR. Feces potentially contains many bacterial species as well as some factors that also could alter the efficiency of PCR reactions. Indeed, several primers gave multiple bands of unknown origin during the experiments with feces DNA. We showed that the three different pair of primers for 16S-SBSEC-3/-4/-5/16S-inf-rev recognized *SG* species, *SI* and *SL* (data not shown). As shown in Figure 6, the oligonucleotides 16S-SBSEC-3/-5/16S-inf-rev produced a non-specific band of about 1400 bp in *SG* free fecal DNA and a fainter band of 4 kbp with the *E. faecalis* sample. The correct and specific fragment of 1120 bp is detected in genomic *SGG* DNA as well as in *SGG* contaminated fecal DNA (DNA\textsuperscript{fecoal} + *SGG*).

---

**Figure 6** Oligonucleotides (16S-SBSEC-3/-5 / 16S-inf-rev) allow the amplification of the desired fragment from genomic DNA extracted from feces (*Streptococcus gallolyticus* strains) in feces extracted DNA. (M) 1 kilo base plus size marker, (E.F.) *Enterococcus faecalis* (negative), (SGG) *Streptococcus gallolyticus* subspecies *gallolyticus* (1 ng/µl), (DNA\textsuperscript{fecoal}) feces DNA, (DNA\textsuperscript{fecoal} + *SGG*) feces contaminated with *SGG*, (bp) base pair. Size of the expected fragment is indicated in red.
The primer 16S-SBSEC/16S-inf-rev was sensitive for SI and SL and hence did not match our targets (data not shown). Gallo2020/2021, BIS-Gallo2020/2021, Gallo2178, Gallo2179 were specific for SGG. SodA1/2, SodA3/4 and qPCR COMX were specific for SG. Generally, the results showed a loss of specificity during the test with fecal DNA. The oligonucleotides BIS-Gallo2020/2021 and SodA1/2 maintained their high target specificity and performed best on fecal DNA.

3.1.1.2 Determination of the oligonucleotide sensitivity

The primers BIS-Gallo2020/2021 and SodA1/2 qualified in the previous specificity tests and were chosen to determine their sensitivity. Less sensitive but beneficial due to their 16s rRNA target sequence, the oligonucleotides 16S-SBSEC-3/-5/16S-inf-rev were included as well.

The sensitivity was initially evaluated with decreasing concentrations of genomic DNA. A decreasing dilution series of genomic DNA concentrations (SGG) 1 ng/µl down to 0.5 pg/µl was set up and audited with the respective oligonucleotide pairs.

SodA1/2 was able to detect genomic DNA of SGG clearly with a concentration of 1pg/µl and even showed a faint band at 0.5pg/µl (Figure 7).
RESULTS

Figure 7 Detection limits using SodA1/2 on Streptococcus gallolyticus subspecies gallolyticus (SGG) genomic DNA samples. To determine sensitivity threshold of the oligonucleotide pair (expected size at ~408 base pair (bp)), classic PCR was performed on aliquots with decreasing concentrations (c) of genomic SGG DNA (M) 1 kilo base plus size marker, (1) SGG 0.0005 ng/µl, (2) SGG 0.001 ng/µl, (3) SGG 0.01 ng/µl, (4) SGG 0.05 ng/µl, (5) SGG 0.1 ng/µl, (6) SGG 0.25 ng/µl, (7) SGG 0.5 ng/µl, (8) SGG 1 ng/µl, (9) Lactobacillus subspecies (negative control). The vertical red arrow (c1) indicates the minimal significant detection limit in this experiment.

BIS-Gallo2020/2021 was able to detect genomic DNA of SGG with a concentration of 0.05 ng/µl and showed a faint band at 0.01 ng/µl (Figure 8).
RESULTS

Figure 8  Detection limits using BIS-Gallo2020/2021 on Streptococcus galloyticus subspecies galloyticus (SGG) genomic DNA samples. To determine sensitivity threshold of the oligonucleotide pair (expected size at ~497 base pair (bp)), classic polymerase chain reaction was performed on aliquots with decreasing concentrations (c) of genomic SGG DNA. (M) 1 kilo base plus size marker, (1) Lactobacillus subspecies (negative control), (2) SGG 1ng/µl, (3) SGG 0,25 ng/ µl, (4) SGG 0,1ng/ µl, (5) SGG 0,05 ng/ µl, (6) SGG 0,01 ng/ µl, (7) SGG 0,001 ng/ µl. The vertical red arrows (c1) indicate the minimal detection limits in this experiment.

The results in Figure 8 show that SBSEC-3/16S-inf-rev was able to detect genomic DNA of SGG with a concentration of 1pg/µl. In another assay it was able to detect 0.5pg/µl SGG. SBSEC-5/16S-inf-rev detected SGG with 10pg/µl and showed a faint band at 1pg/µl (Figure 9).
RESULTS

Figure 9 Detection limits using 16S-SBSEC-3/-5 + 16S-inf-rev on Streptococcus galloyticus subspecies galloyticus (SGG) genomic DNA samples. To determine sensitivity threshold of each oligonucleotide pairs (expected size at ~1119/20 base pair (bp)), classic PCR was performed on aliquots with decreasing concentrations (c) of genomic SGG DNA (M) 1 kilo base plus size marker, (1) *Lactobacillus* subspecies (negative control), (2) SGG 1ng/µl, (3) SGG 0.25 ng/µl, (4) SGG 0.1ng/µl, (5) SGG 0.05 ng/µl, (6) SGG 0.01µl, (7) SGG 0.001 ng/µl, (8) *Lactobacillus subspecies* (negative control) plus unfortunate application of an unknown size marker. The vertical red arrows (c1/c2) indicate the minimal detection limits in this experiment.

Subsequently, the detection limit in feces isolated DNA was determined. Feces samples from *SG* negative tested healthy volunteers were contaminated with specific amounts of *SGG* in decreasing quantities (1.25x10⁹ – 1.25x10² CFU/100mg feces). The DNA content was then extracted and the detection threshold for each oligonucleotide was set by classic PCR (*Figure 10, Table 3*). 16s rRNA is widely used as a target structure for the identification of bacteria (Keller *et al.*, 2010). Due to multiple repetitive operons in the genome, oligonucleotides with target sequences in the 16s rRNA are estimated to be highly sensitive. The operon is represented with 6 copies in the genome of *SG*.
Consequently, the amplification signal of 16s oligonucleotides are naturally enhanced by the factor 6. The high sensitivity was not confirmed for 16S-SBSEC-5/16-S-inf-rev but for 16S-SBSEC-3/16-S-inf-rev by our experiments. The oligonucleotides 16S-SBSEC-3/16-S-inf-rev presented itself as the most sensitive, allowing the detection of SGG in feces to concentrations of $1.25 \times 10^2$ CFU / 100 mg feces (Figure 10). As previously observed, the two primers 16S-SBSEC-3/-5/16-S-inf-rev reveal their deficient specificity by showing multiple aspecific bands beside the expected.

**Figure 10** Detection limits using 16S-SBSEC-3/-5 + 16S-inf-rev on fecal DNA samples contaminated with known amounts (CFU – colony forming unit) of *Streptococcus gallolyticus* subspecies *gallolyticus* (SGG) strain UCN34. Classic polymerase chain reaction on fecal DNA contaminated with SGG in decreasing concentrations was performed to determine the sensitivity threshold of the respective oligonucleotides. Universal 16s rRNA oligonucleotides (*All Bacteria*) able to amplify any bacterial DNA was used as control for DNA extraction (expected size at 146 base pair (bp)). (M) 1 kilo base plus size marker (1) SGG free fecal DNA (negative control), (2) genomic SGG 1ng/µl (positive control), (3) SGG $1.25 \times 10^8$ CFU/100mg feces, (4) SGG $1.25 \times 10^9$ CFU/100mg feces, (5) SGG $1.25 \times 10^7$ CFU/100mg feces, (6) SGG $1.25 \times 10^6$ CFU/100mg feces, (7) SGG $1.25 \times 10^5$ CFU/100mg feces, (8) SGG $1.25 \times 10^4$ CFU/100mg feces, (9) SGG $1.25 \times 10^3$ CFU/100mg feces, (10) SGG $1.25 \times 10^2$ CFU/100mg feces; The red arrows (c1/c2) indicated the respective detection limits.
Intriguingly, the oligonucleotide pair *SodA1/2* reached the same sensitivity for *S. gallolyticus* as *16S-SBSEC-3/16S-inf-rev*, despite it targets the single represented superoxide dismutase gene in *S. gallolyticus* (Figure 11).

**Figure 11** Detection limits of *SodA1/2* on fecal DNA samples contaminated with known amounts (CFU – colony forming unit) of *Streptococcus gallolyticus* subspecies *gallolyticus* (SGG) strain UCN34. Classic polymerase chain reaction on fecal DNA contaminated with SGG in decreasing concentrations was performed to determine the sensitivity threshold of the respective oligonucleotides. (bp) base pair, (M) 1 kilo base plus size marker, (1) genomic SGG 1ng/µl (positive control), (2/3) SGG 1.25x10^2 CFU/100mg feces, (4/5) SGG 1.25x10^3 CFU/100mg feces, (6/7) SGG 1.25x10^4 CFU/100mg feces, (8/9) SGG 1.25x10^5 CFU/100mg feces, (10/11) SGG 1.25 x 10^6 CFU/100mg feces, (12/13) SGG 1.25x10^6 CFU/100mg feces, (14/15) SGG 1.25 x 10^6 CFU/100mg feces, (16) SGG free fecal DNA (negative control). The red arrows in sample 4/5 (c1/c2) indicate the respective detection limits. Sample (3) shows a faint and sample (2) a dislocated band which is not valued SGG positive.

*BIS-Gallo2020/2021* detect a minimal concentration of 1.25x10^7 CFU/100mg feces.

Based on these evaluating experiments a ranking considering sensitivity and specificity was generated (*Table 3*). The establishment of highly sensitive and specific oligonucleotides was a crucial step during the preparations for this
RESULTS

study. The oligonucleotide pairs 16S-SBSEC-3-FW/16S-inf-rev and SodA1/2 appeared to be most significant with a high sensitivity (< 1,25x10^2 CFU/100mg) but major differences in their specificity. 16S-SBSEC-3/16S-inf-rev recognizes the SBSEC members SG and SI, while SodA1/2 only recognize SG strains. Eventually, we used SodA1/2 in our workflow with RT-PCR due to its favorable short length of 408 bp and its better performance in feces extracted DNA. To compensate the lack of information about the subspecies, continuative molecular methods (RFLP assays, PCR product sequence analyzing) were tested to provide the missing data.

Table 3  Oligonucleotide properties. The table shows various oligonucleotides that were tested. The primers selected for the clinical investigation are marked in yellow. The threshold is specified for genomic DNA (gDNA) and for Streptococcus gallolyticus (SG) subspecies gallolyticus (SGG) colony forming units (CFU) dissolved in feces. No threshold is indicated if the primer was discarded due to its specificity and the sensitivity was not tested. (SBSEC) Streptococcus bovis/ Streptococcus equinus complex, (SGM) SG subspecies macedonicus, (SGP) SG subspecies pasteurianus, (SI) Streptococcus infantarius, (SL) Streptococcus lutetiensis

<table>
<thead>
<tr>
<th>Target</th>
<th>Primer</th>
<th>Specificity</th>
<th>Length (bp)</th>
<th>Threshold gDNA (ng/µl)</th>
<th>Threshold SGG in feces (CFU/100mg)</th>
<th>Annealing Temperature (°C)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>General bacterial DNA</td>
<td>AllBacteria</td>
<td>all</td>
<td>146</td>
<td>-</td>
<td>-</td>
<td>50</td>
<td>Furet et al. 2009</td>
</tr>
<tr>
<td>SBSEC specific</td>
<td>16S-SBSEC-3/16S-inf-rev</td>
<td>SGG, SGM, SGP, SI, SL</td>
<td>1119</td>
<td>&lt; 0,0005</td>
<td>&lt; 1,25x10^2</td>
<td>55</td>
<td>Jans et al., 2012</td>
</tr>
<tr>
<td></td>
<td>SodA1/2</td>
<td>SGG, SGM, SGP</td>
<td>408</td>
<td>&lt; 0,001</td>
<td>&lt; 1,25x10^2</td>
<td>50</td>
<td>Sasaki et al., 2004</td>
</tr>
<tr>
<td></td>
<td>16S-SBSEC-5/16S-inf-rev</td>
<td>SGG, SGM, SGP, SI, SL</td>
<td>1119/20</td>
<td>&lt; 0,001</td>
<td>&lt; 1,25x10^7</td>
<td>55</td>
<td>Jans et al., 2012</td>
</tr>
<tr>
<td></td>
<td>BIS-Gallo 2020/2021</td>
<td>SGG</td>
<td>497</td>
<td>&lt; 0,005</td>
<td>&lt; 1,25x10^7</td>
<td>50</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>16S-SBSEC/16S-inf-rev</td>
<td>SI, SL</td>
<td>1119/20</td>
<td>-</td>
<td>-</td>
<td>50</td>
<td>Jans et al., 2012</td>
</tr>
<tr>
<td></td>
<td>16S-SBSEC-4/16S-inf-rev</td>
<td>SGG, SGM, SGP, SI, SL</td>
<td>1119/20</td>
<td>-</td>
<td>-</td>
<td>50</td>
<td>Jans et al., 2012</td>
</tr>
<tr>
<td></td>
<td>Gallo 2020/2021</td>
<td>SGG</td>
<td>1353</td>
<td>-</td>
<td>-</td>
<td>50</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>Gallo2178</td>
<td>SGG</td>
<td>125</td>
<td>-</td>
<td>-</td>
<td>52.5</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>Gallo2179</td>
<td>SGG</td>
<td>173</td>
<td>-</td>
<td>-</td>
<td>52.5</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>qPCR COMX</td>
<td>SGG, SGM, SGP</td>
<td>182</td>
<td>-</td>
<td>-</td>
<td>50</td>
<td>Designed for this study</td>
</tr>
<tr>
<td></td>
<td>SodA 3/4</td>
<td>SGG, SGM, SGP</td>
<td>334</td>
<td>-</td>
<td>-</td>
<td>50</td>
<td>Designed for this study</td>
</tr>
</tbody>
</table>
3.1.2 DNA isolation of clinical fecal samples

The protocol that was used for the DNA isolation from feces favored the isolation of gram-positive bacteria (Ahlroos and Tynkkynen, 2009).

A challenge was to measure and equalize the bacterial DNA concentration in the samples. The extracted DNA compound of a specific volume of stool can be very different due to its high variability. Feces contains predominantly eukaryotic host DNA of dead cells and genetic material from food components. In contrast, the bacterial DNA share is comparatively low. Therefore, nonspecific DNA monitoring methods such as electrophoretic fragmentation of pure DNA or direct measurements with UV/VIS spectrophotometer (NanoDrop™ 2000c Spectrophotometer) resulted in a false detection of bacterial and non-bacterial DNA. Even indirect concentration measurements of specific bacterial DNA-binding fluorophores (Qubit 4 Flurometer™) failed to correspond with the content of bacterial DNA. The wrong measurement happened most likely due to alternated binding of the fluorophores in the milieu of stool-isolated DNA. RT-PCR performed with a universal bacterial 16s rRNA oligonucleotide pair (AllBacteria) is specific for all bacterial DNA (Furet et al., 2009). This method was used for relative quantification of bacterial DNA in the samples on which bases the concentrations of bacterial DNA fractions were equalized in each sample.

A major falsifying risk factor for every step from the sample recovery to the sample examination was the contamination with genetical traces of *S. gallolyticus*. The contamination could happen via direct contact with vectors or indirectly via aerosols containing genomic material from *S. gallolyticus*. General precautions according to the WHO guideline “Dos and Don’ts for molecular testing” encompassed a low working temperature, a minimum of mechanical manipulation, and minimal open exposure time to avoid contamination. To avoid contamination, the samples were immediately separated in multiple aliquots to prevent an excessive use of single tubes. Used DNA aliquots were discarded after single use and consumable supplies were exchanged frequently.
Consequently, the risk of false detection due to contamination was reduced significantly (WHO, 2018).

3.1.3 PCR in combination with a RFLP assay

We tested a PCR / RFLP assay for the identification of SG subspecies that was previously developed (Jans et al., 2012). This approach consisted of a specific amplification of 16s RNA genes using 16S-SBSEC-3-FW/16S-inf-rev primers and a following characterization of subspecies by enzymatic digestion (Msel/Xbal). Working with 16S-SBSEC-3/16S-inf-rev products based on genomic DNA from broth cultures, we indeed obtained similar results as Jans et al., 2012. Using the Msel/Xbal RFLP assay, we were able to identify SG subspecies in 16S-SBSEC-3/16S-inf-rev PCR products. During the analysis of feces extracted DNA, 16S-SBSEC-3/16S-inf-rev frequently displayed aspecific or weak bands which then disturbed RFLP assays. Since the results with the 16s rRNA oligonucleotides were not convincing, a PCR analysis with a second pair of highly specific primers SodA1/2 (Sasaki et al., 2004) was used. SodA1/2 performed more specifically than 16S-SBSEC-3-FW/16S-inf-rev and replaced the 16s rRNA primers (Figure 12).
RESULTS

Figure 12  Comparison of the polymerase chain reaction results obtained on feces samples either with 16S-SBSEC-3/16S-inf-rev (1119/20 base pair (bp)) or SodA1/2 (408bp). (M) 1 kilo base plus size marker, (1) - (5) clinical fecal samples from randomized cohort members (6) H2O (negative control), (7) Streptococcus gallolyticus subspecies gallolyticus (SGG) strain UCN34 genomic DNA 1 ng/µl (positive control). Samples (1), (2) and (6) were negative in both experiments. Samples (3), (5) and (7) were positive in both experiments. Sample (4) gave a divergent result according to the primers used: weakly positive with 16S-SBSEC-3/16S-inf-rev and negative with SodA1/2 primers.

We indeed obtained discrepant results between 16S-SBSEC-3/16S-inf-rev and SodA1/2 (Figure 13). Thus, we developed a new approach based on a RFLP assay with the second, more specific primers SodA1/2. The DNA sequences of SGG, SGP, SGM amplified by SodA1/2 were compared and analysis revealed that a PCR mapping using HindIII and SfaNI restriction enzymes could be used to discriminate between the 3 subspecies (Figure 13). Restriction maps revealed that SGG and SGP PCR products exhibit two recognition sites for HindIII and only one for SGM. To distinguish between SGG and SGP, SfaNI restriction enzyme could be used.
RESULTS

Figure 13  Restriction site analysis of the 408 base pair (bp) SodA1/2 DNA fragment of *Streptococcus gallolyticus* subspecies *gallolyticus* (SGG), *Streptococcus gallolyticus* subspecies *pasteurianus* (SGP) and *Streptococcus gallolyticus* subspecies *macedonicus* (SGM). Nucleotide sequence of sodA of SGG and SGP show two HindIII restriction sites and only one for SGM. Consequently, three fragments (102, 152, and 154 bp) will be obtained with SGG and SGP polymerase chain reaction products and two fragments (154 and 254 bp) with SGM. SodA1/2 SGG and SGM sequences do not harbor SfaNI restriction site, whereas one site is present in SGP SodA sequence. Therefore, HindIII is able to discriminate SGM out of the *Streptococcus gallolyticus* (SG) family and SfaNI distinguishes SGP out of the group. Combining the two restriction fragment length polymorphism assays, it is possible to differentiate the SG SodA1/2 PCR products into its subspecies SGG, SGP and SGM.

Applied to genomic DNA of single subspecies the verification was accurate for the three subspecies SGG, SGM and SGP (Figure 14). Verification on DNA from fecal samples contaminated with a single SG subspecies were quite promising as well.
Electrophoretic separation of DNA fragments after *Streptococcus gallolyticus* (SG) specific polymerase chain reaction amplification with SodA1/2 and restriction fragment length polymorphisms assay. Amplified DNA fragments were digested using HindIII (A) and SfaNI (B) restriction enzymes in two separate reactions. (bp) base pair, (M) 1 kilo base plus size marker. (SGG) *Streptococcus gallolyticus* subspecies *gallolyticus*, (SGP) *Streptococcus gallolyticus* subspecies *pasteurianus*, (SGM) *Streptococcus gallolyticus* subspecies *macedonicus*.

Theoretically the RFLP assay should have been accurate enough to discriminate if multiple subspecies were present in individual samples. Unfortunately, this could not be confirmed in tests with genomic DNA and feces extracted DNA. Since the possibility of multiple SG subspecies per sample was a conceivable scenario in clinical samples, the disability of the HindIII/SfaNI RFLP assay to work in such conditions would have been a clear limitation to our study. This weakness was eliminated by introducing a new protocol.
3.1.4 Real-time PCR / PCR product nucleotide sequence assay

3.1.4.1 Investigation of *S. gallolyticus* by RT-PCR

Based on the previous experiences, the final assay featured a RT-PCR based identification of *SG* with the established *SodA1/2* oligonucleotides combined with a nucleotide sequence analysis of the PCR products. Despite the high performance of *SodA1/2* with classic PCR during our tests we observed problems with *SG* unclear samples during the investigation of our clinical samples, the extended reviewing possibilities of RT-PCR helped to identify *SG* positive and negative samples better. Besides the quantification of DNA, a key tool was the analysis of the melt curve and the determination of the amplicon-specific melt temperature. The melt temperature is an excellent tool to review the DNA amplification of the right templates and to recognize wrongly replicated sequences. The temperature-dependent dissociation of double stranded DNA (dsDNA) during DNA denaturation can be visualized with the melt curve. DNA-intercalating fluorophores reach their peak fluorescence during the maximum presence of dsDNA. When the temperature is raised, dsDNA of a certain length dissociates around a characteristic temperature. As the quantity of single stranded DNA (ssDNA) increases, dsDNA is reduced equally to the emitted fluorescence. The melt curve measures the decreasing fluorescence, equivalent to the dissociating DNA double strands, against temperature. The melt temperature defines the point where 50% of dsDNA is denatured during heating. The expected melt temperature for an amplicon derives from its base pair length and base composition. The amplicon of the oligonucleotides *sodA1/2* with 408 bp melts at 75.9 °C. The results could then easily be monitored for the appropriate melt temperature peaks. All different signals were accordingly discriminated.

The sample results were then controlled by repeating the experiment with another aliquot set, reagents and equipment. The PCR products of random samples were spot checked by electrophoretic fragmentation to control the digital results analogously. As shown in *Figure 15*, positive and negative results
were well-defined and fully matched expected results from RT-PCR. Further, the PCR did not produce unwanted byproducts.

Figure 15    Electrophoretic migration of DNA fragments after real-time polymerase chain reaction (RT-PCR) amplification with SodA1/2 (expected size 408 base pair (bp)). RT-PCR was performed on the clinical fecal samples. Random test of RT-PCR by agarose gel electrophoresis to control the previously gained digital results. (M) 1 kilo base plus size marker, (1)-(14) RT-PCR products (SodA1/2) of randomly chosen clinical feces samples. Samples marked with a green arrow were positive. The results obtained from the electrophoretic analysis matched those obtained during RT-PCR.

3.1.4.2 Identification of the *S. gallolyticus* subspecies by sequence analyzing

We used a subsequent sequence analysis to search for specific single nucleotide polymorphisms (SNP) in the amplified *soda* gene in order to distinguish between *S. gallolyticus* subspecies. The sodium oxide dismutase gene is a ubiquitous represented gene in all *Streptococci*. *SodA1/2* primers are able to specifically amplify a part of the *soda* gene of *SG* species. In order to obtain the complete subspecies identification, the RT-PCR product of *SG* positive samples was purified, and its nucleotide sequence determined. The respective subspecies present nucleotide differences in the *soda* encoding region. Twelve and 23
nucleotide differentiate *SGG* and *SGM*, and *SGG* and *SGP* sequences, respectively (Figure 16). Even in concomitant presence of multiple strains it was possible to determine each subspecies affiliation. Oligonucleotides *SodA3* and *SodA4* were used for nucleotide sequencing.

In summary, different PCR techniques in combination with restriction fragment length polymorphism (RFLP) analysis and nucleotide sequence analyzing were used to identify *SG* subspecies based on the *sodA* gene in human feces (Figure 17). The initial experiment concept featured the examination of the clinical fecal samples by classic PCR and RFLP analysis. Several technical obstacles led to the usage of RT-PCR in combination with DNA fragment sequence analysis.
Figure 17  Flowchart of the protocols used for detection of *Streptococcus gallolyticus* (SG) in human fecal samples. Details are discussed in the respective chapters. (A.) This schema indicates sample preparation and examination. This method combines classic polymerase chain reaction (c-PCR) and subsequent RFLP assay. Detection process started by classic PCR with the...
RESULTS

1st oligonucleotide pair (onp) 16S-SBSEC-3/16S-inf/reverse (Jans et al., 2011) and controlled with the 2nd SodA onp (repeated twice). The PCR products were investigated with restriction fragment length polymorphisms (RFLP) assays to discriminate each subspecies. Due to inconsistent results and not clear results this protocol was dropped, and a second protocol shown in (B) was adapted for real-time (RT)-PCR and showed consistent and reproducible results. *(SGG)* Streptococcus galloyticus subspecies galloyticus, *(SGP)* Streptococcus galloyticus subspecies pasteurianus, *(SGM)* Streptococcus galloyticus subspecies macedonicus

3.2 The prevalence of *S. galloyticus* among clinical samples

3.2.1 General characteristics of our cohorts

The clinical data for each patient group is summarized in Table 4. The final results included 71 participants (♂43, ♀28) in total with 48 colorectal cancer negative (♂27, ♀21, sex ratio 1.3:1.0) and 23 colorectal cancer positive patients (♂16, ♀7). The selected sex ratio in the CRC patient group (♂2.3: ♀1.0) reflected approximately the higher CRC incidence among men in western Europe of ♂2.8 : ♀1.0 (Rawla et al., 2019). The 48 colorectal cancer negative participants are subdivided into 2 groups consisting of 25 healthy controls (CN; ♂13, ♀12; sex ratio 1.0: 1.0) without any colorectal lesions and 23 colorectal adenoma patients (CRA; ♂14, ♀9; sex ratio 1.6: 1.0). The mean age of this cohort is 62.9 ± 2 years. The mean body mass index (BMI) is 26.0 ± 0.9.

An adenoma is a benign neoplasia from epithelial origin that potentially can progress into a malignant carcinoma. This sequence is described as the adenoma-carcinoma-sequence. Our CRA patients harbor small hyperproliferations, small and intermediate adenomas, but no severe dysplasia (pre-cancerous polyps). At the time of the sample isolation, the CRC patients displayed various stages of the primary tumor, but no signs of lymph node invasion or metastasis. According to the Union of International Cancer Control (UICC) classification the CRC patients correspond to stages I and II *(Table 5)* (O’Sullivan et al., 2017).
Table 4  General characteristics of the patient groups. (CRC) colorectal carcinoma bearing patients, (CRA) colorectal adenoma patients, (CN) healthy control volunteers, (non-CRC) all healthy control volunteers (CN) and non-malignant colorectal adenoma harboring patients (CRA) combined. “Mean age” refers to the age of the person when the fecal sample was isolated. BMI, body mass index.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total (n)</th>
<th>CRC</th>
<th>CRA</th>
<th>CN</th>
<th>non-CRC</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (n)</td>
<td>71</td>
<td>23</td>
<td>23</td>
<td>25</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women (n)</td>
<td>28</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men (n)</td>
<td>43</td>
<td>16</td>
<td>14</td>
<td>13</td>
<td>27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total</th>
<th>CRC</th>
<th>CRA</th>
<th>CN</th>
<th>non-CRC</th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age (years)</td>
<td>62.9</td>
<td>64.7</td>
<td>60.8</td>
<td>63.9</td>
<td>62.1</td>
<td>64.8</td>
<td>61.7</td>
</tr>
<tr>
<td>Mean BMI</td>
<td>26.0</td>
<td>25.9</td>
<td>26.1</td>
<td>26.0</td>
<td>26.0</td>
<td>25.4</td>
<td>26.9</td>
</tr>
</tbody>
</table>

Table 5  UICC stages are used for prognosis and subsequent treatment of a tumor. The stages are determined through the TNM (Tumor, Regional Lymph Nodes, Distant Metastasis) class of a tumor. TNM is a globally accepted standard to classify the anatomical extend and spread of cancer. This table shows a simplified presentation. Tis (Carcinoma in situ): intraepithelial or invasion of lamina propria, T1: tumor invaded submucosa, T2: tumor invaded muscularis propria, T3: tumor invaded peri-colorectal tissues through the muscularis propria, T4a: tumor penetrates to the surface of the visceral peritoneum, T4b: tumor directly invades or is adherent to other organs or structures. N0: No regional lymph node metastasis, N1: Metastasis in 1-3 regional lymph nodes, N2: Metastasis in 4 or more regional lymph nodes. M0: No distant metastasis, M1: Distant metastasis (Brierley et al., 2018). Untagged fields are relevant for this study.

<table>
<thead>
<tr>
<th>UICC Stage</th>
<th>Tumor (T)</th>
<th>Lymph Nodes (N)</th>
<th>Distant Metastasis (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Tis</td>
<td>N0</td>
<td>M0</td>
</tr>
<tr>
<td>I</td>
<td>T1, T2</td>
<td>N0</td>
<td>M0</td>
</tr>
<tr>
<td>IIa</td>
<td>T3</td>
<td>N0</td>
<td>M0</td>
</tr>
<tr>
<td>IIb</td>
<td>T4a</td>
<td>N0</td>
<td>M0</td>
</tr>
<tr>
<td>IIc</td>
<td>T4b</td>
<td>N0</td>
<td>M0</td>
</tr>
<tr>
<td>III</td>
<td>Any T</td>
<td>N1/N2</td>
<td>M0</td>
</tr>
<tr>
<td>IV</td>
<td>Any T</td>
<td>Any N</td>
<td>M1</td>
</tr>
</tbody>
</table>
3.2.2 Detection of *S. gallolyticus* member bacteria in fecal DNA

The prevalence of each *S. gallolyticus* member bacteria (SGMB) was determined in the fecal DNA samples of our control, adenoma and colorectal cancer groups.

### 3.2.2.1 Overview of *S. gallolyticus* prevalence within the final cohort

In a first step we gain an overview of all samples that contained any *S. gallolyticus* member bacteria in the various groups (*Table 6, Figure 18*). This was accomplished by the detection of the *sodA* gene using RT-PCR. *SodA* is a ubiquitous gene encoding the superoxide dismutase frequently used for bacterial identification by PCR techniques (Jans *et al.*, 2012).

**Table 6** Detection of *Streptococcus gallolyticus* member bacteria (SGMB) in feces of healthy control (CN), adenoma (CRA) and colorectal cancer bearing groups (CRC). SGMB includes the subspecies *Streptococcus gallolyticus* subspecies *gallolyticus*, *Streptococcus gallolyticus* subspecies *pasteurianus* and *Streptococcus gallolyticus* subspecies *macedonicus*.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of persons</th>
<th>Positive isolation of fecal SGMB - n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>25</td>
<td>15 (60.0)</td>
</tr>
<tr>
<td>CRA</td>
<td>23</td>
<td>15 (65.2)</td>
</tr>
<tr>
<td>CRC</td>
<td>23</td>
<td>12 (52.2)</td>
</tr>
</tbody>
</table>

Among healthy controls, we find an overall prevalence of 60.0 % SGMB. The frequency indicated in colorectal adenoma bearing patients increases slightly up to 65.2 %. In colorectal cancer patients, the prevalence of SGMB was lowest, 52.2%.
3.2.2.2 Detection of *S. gallolyticus* at subspecies level

After SGMB positive samples were identified in step one, the second phase should give an answer to which exact *S. gallolyticus* subspecies we were detecting in the sample. Therefore, the nucleotide sequence of the PCR product was determined. Through this analysis, we identified the respective subspecies in the sample (*Table 7; Figure 19*). We were also able to detect multiple SGMB subspecies within one sample.
Table 7  Detection and subspecies classification of *Streptococcus gallolyticus* member bacteria (SGMB) in feces of healthy control (CN), adenoma (CRA) and colorectal cancer groups (CRC). Number defines the cohort size.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number</th>
<th>Positive isolation of fecal SGMB - n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><em>Streptococcus gallolyticus</em> subspecies <em>gallolyticus</em></td>
</tr>
<tr>
<td>CN</td>
<td>25</td>
<td>11 (44.0)</td>
</tr>
<tr>
<td>CRA</td>
<td>23</td>
<td>10 (43.5)</td>
</tr>
<tr>
<td>CRC</td>
<td>23</td>
<td>9 (39.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Streptococcus gallolyticus</em> subspecies <em>pasteurianus</em></td>
</tr>
<tr>
<td>CN</td>
<td>25</td>
<td>6 (24.0)</td>
</tr>
<tr>
<td>CRA</td>
<td>23</td>
<td>7 (30.4)</td>
</tr>
<tr>
<td>CRC</td>
<td>23</td>
<td>3 (13.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Streptococcus gallolyticus</em> subspecies <em>macedonicus</em></td>
</tr>
<tr>
<td>CN</td>
<td>25</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>CRA</td>
<td>23</td>
<td>2 (8.7)</td>
</tr>
<tr>
<td>CRC</td>
<td>23</td>
<td>3 (13.0)</td>
</tr>
</tbody>
</table>

Among the control cohort of healthy volunteers, we detect *S. gallolyticus subsp. gallolyticus* (*SGG*) in 44 % of the samples. Same prevalence was found in colorectal adenoma patients. In the feces of CRC patients, the prevalence decreases to 39.1 %.

*S. gallolyticus subsp. pasteurianus* (*SGP*) was found in the healthy control group with a prevalence of 24.0 %. The *SGP* colonization in adenoma bearing patients is more frequent (30.4 %). The lowest *SGP* carriage is in the colorectal cancer group, 13.0 %.

The detection of *S. gallolyticus subsp. macedonicus* (*SGM*), the third subspecies considered as non-pathogenic was quite rare in the cohort that we have analyzed. The highest prevalence of *SGM* was found in the CRC cohort, 13% versus 8.7% in CRA and 0% in CN.
3.2.2.3 Detection of multiple \textit{S. gallolyticus} subspecies within individual samples

The analysis of the PCR products identified DNA of one or multiple SGMB within particular fecal samples. We also have analyzed the simultaneous colonization by several \textit{S. gallolyticus} subspecies of individuals and possible associated characteristics.

Our analysis shows, that only a small part of the identified \textit{S. gallolyticus} subspecies mutually exists in the samples (\textit{Figure 20}). Due to the limited size of our cohorts, the identification of multiple strains within samples could result in low case numbers. The low numbers struggle to be representative as \textit{Table 8} shows.
Table 8  
*Streptococcus galloyticus* member bacteria (SGMB) in the feces of control (CN), adenoma (CRA) and colorectal cancer bearing groups (CRC). The classification shows if the strain was found alone in a sample or in cohabitation with another *Streptococcus galloyticus* subspecies. Number defines the cohort size, total counts the overall incidences of a subspecies, alone specifies samples with only one subspecies, + “subspecies” identifies cases with combinations of subspecies. *Streptococcus galloyticus* subspecies *galloyticus* (SGG), *Streptococcus galloyticus* subspecies *pasteurianus* (SGP), *Streptococcus galloyticus* subspecies *galloyticus macedonicus* (SGM).

<table>
<thead>
<tr>
<th>Group</th>
<th>Number</th>
<th>SGG</th>
<th>SGP</th>
<th>SGM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>total</td>
<td>alone</td>
<td>+ SGP</td>
</tr>
<tr>
<td>CN</td>
<td>25</td>
<td>11</td>
<td>9 (82)</td>
<td>2 (18)</td>
</tr>
<tr>
<td>CRA</td>
<td>23</td>
<td>10</td>
<td>6 (60)</td>
<td>3 (30)</td>
</tr>
<tr>
<td>CRC</td>
<td>23</td>
<td>9</td>
<td>6 (67)</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>

Due to their low significance, combinations of subspecies that were only sighted once are not considered in percentage in the following text.

Within the healthy control group, we find *SGG* in 82 % of its cases as the only *S. galloyticus* subspecies. *SGP* is detected in 67 % of its manifestations solitarily. A fraction of samples is colonized by both strains. The relative percentages of *SGG* plus *SGP* sightings are calculated for each subspecies. Thus, 18 % of *SGG* subspecies coexist with *SGP* and, vice versa, 33% of *SGP* subspecies are found together with *SGG*. As earlier presented, *SGM* is not detected among healthy participants.

60 % of *SGG* cases are not accompanied by other *S. galloyticus* subspecies in adenoma bearing patients. The same applies to 57 % of *SGP*. Our analysis shows that 30 % of *SGG* and 43 % of *SGP* cohabitate with each other. The combination *SGG* + *SGM* is only detected in one case.

Only a low number of *SGM* and *SGP* was detected in the group of CRC patients. Although crossovers of *S. galloyticus* subspecies have been observed, the low case numbers must be emphasized. *SGG* as well as *SGP* are both detected in 67 % of their cases as the only colonizing *S. galloyticus* subspecies. We find that
22% of *SGG* cohabitate with *SGM*. Vice versa, 67% of *SGM* is found in combination with *SGG*. *SGM* solitarily as well as the combination of *SGG* and *SGP* are observed in only one case.

**Figure 20** Frequency of solitary growth versus co-colonization of different *Streptococcus galolyticus* subspecies (SG). *Streptococcus galolyticus* subspecies *galolyticus* (*SGG*), *Streptococcus galolyticus* subspecies *pasteurianus* (*SGP*), *Streptococcus galolyticus* subspecies *macedonicus* (*SGM*). Their prevalence in the healthy control group in grey (CN), colorectal adenoma bearing group in orange (CRA) and group of colorectal cancer in red (CRC) is compared.
4 DISCUSSION

The increased risk for colorectal cancer is associated with demographic, behavioral and environmental factors. To date, important associated factors are first-degree relatives suffering from CRC, genetic predispositions, the prevalence of IBDs, increased BMI, red meat intake, cigarette smoking, low physical activity and low fiber consumption (Jones et al., 2008). Besides these factors, the individual risk to develop CRC is driven by multiple other effects of unknown relevance. Among those, a hazardous category are specific members of the intestinal microbiome of which some are confirmed to promote and accelerate the colorectal tumor progression (Pasquereau-Kotula et al., 2018).

*S. gallolyticus* subspecies *gallolyticus* causes various infective diseases in humans and is associated with the growth of colorectal cancer. People diagnosed with infective endocarditis caused by *SGG* are advised to undergo a prophylactic colonoscopy to identify potential neoplasia. Studies outline the role of *SGG* as promoter and profiteer of colorectal cancer (Boleij et al., 2011; Kumar et al., 2017; Pasquereau-Kotula et al., 2018). Despite the progressing knowledge about the interaction between *SGG* and the human body, information about its epidemiology remains scarce. To date, literature provides heterogeneous data about the prevalence of *SGG* and its phylogenetic relatives (*Streptococcus gallolyticus* member bacteria; SGMB) in the healthy population and such with a history of colorectal neoplasia. Different detection methods among studies, the inconsistent use of nomenclatures for *S. gallolyticus* subspecies, and unfavorable selection of cohorts in many studies are significant limitations for a comparison or conclusion about the prevalence of *SGG*. Dependent on the study, an *SGG* carriage in the healthy population ranges between 2.5 – 62.5 % whereas CRC bearing patients seem to show a colonization between 47 – 85 %, depending on the detection method used (Klein et al., 1977; Waisberg et al., 2002; Kok et al., 2007; Gupta et al., 2010; Corredoira et al., 2015; Chand et al., 2016; Dumke et al., 2017). A meta review of 52 case reports published between
1951 and 2010 attempted to summarize earlier studies (Boleij et al., 2011). In this review 20 studies that differentiate colorectal neoplasia in early adenomas and late carcinomas were separately reviewed and compared. The results clearly show an increased *S. bovis* carriage in early stages (median of 43 %) towards late tumor stages (median of 18 %). Though, only one study among the 20 included an age-matched control population. The final results were only able to conclude that *S. bovis* infected patients show increased rates of adenomas and carcinomas. A definite limitation of these studies is clearly the frequent absence of differentiation into *S. bovis* subspecies. When a subclassification was performed, a clear peak in infections with *S. bovis* biotype I (*SGG*) is observed in CRC patients while *S. bovis* biotype II strains (*SGP, SII, SIL*) do not seem to exceed the general asymptomatic population (Boleij et al., 2011).

### 4.1 The prevalence of *S. gallolyticus* infections

Our study complements the existing data on the prevalence of *S. gallolyticus* subspecies in the feces of healthy people and such with different conditions of colorectal neoplasia. Previous limitations are minimized by using a strictly age and weight matched cohort consisting of 25 healthy, 23 benign colorectal adenoma and 23 malign carcinoma bearing individuals. The sex ratio reflects higher CRC incidence among men in western Europe of 2.8 ♂ : 1.0 ♀ (Rawla et al., 2019). The colorectal health of every participant is clinically confirmed by colonoscopy. With a pointer towards previous studies, the detection method is crucial to obtain clear results. Consequently, a new protocol using different molecular assays was established to gain and process DNA from the fecal samples.

#### 4.1.1 *S. gallolyticus* subsp. *gallolyticus*

Our results reveal a high *SGG* incidence in the feces of our healthy cohort. Almost half healthy individual is colonized with *SGG*, as we find the subspecies in 44% of our control group. The prevalence of *SGG* in the colorectal adenoma bearing cohort reaches almost the same frequency (43%). Our data emphasizes
that *SGG* is found regularly in the alimentary system of the healthy population as well as such with benign colorectal hyperplasia. Our data correlates with a late German publication (Dumke *et al.*, 2017). Using highly sensitive PCR assays on feces as well, the study postulated an even higher *SGG* carriage of 62.5% in the healthy population.

Due to the link between the subspecies and colorectal neoplasia we hypothesized elevated numbers in the colonization among this cohort. Differently than expected, the prevalence of *SGG* in CRC patients is slightly lower than the equal incidence among the cohorts. The number of surveys that do not demonstrate a link between colorectal cancer and the colonization of the gut with *SGG* are few. Only two studies using a culture-based and RT-PCR based approach suggest no conspicuous association between *S. gallolyticus* and tumorous tissue (Boltin *et al.*, 2015; Viljoen *et al.*, 2015).

In the follow-up, the cohorts were reexamined by a different researcher. Using the described experimental set-up on the same set of samples, the results show some intriguing differences to our first analysis. Possible explanations and considerations for the discrepancy are discussed in *Paragraph 4.2*. The previous / new experiment series are labeled as Series A / B.

As observed in Series A, Series B finds *SGG* in healthy and adenoma bearing patients at similar frequencies. In contrast to Series A the prevalence of *SGG* detected by Series B in healthy (28.0%) and adenoma bearing volunteers (30.4%) is lower by an average of 14.5% points in both groups. The most striking contrast to Series A is seen in colorectal cancer patients. Series B detects *SGG* in 43.5% of the cases. As opposed to series A, the high case number in Series B emphasizes a favored colonization of *SGG* in the gastrointestinal system of colorectal tumor patients. This is in agreement with other surveys that associated *SGG* with colorectal tumors. An earlier study examining colorectal tissue with PCR techniques found SGMB in 39.5% of the colorectal tissue of tumor bearing patients and in 28.5% of the healthy controls (Abdulamir *et al.*, 2010). A more recent publication compared the prevalence of *SGG* in the colorectal tissue of CRC patients by using *SGG*-specific RT-PCR. The authors found the bacteria in 74% of colorectal tumor cells and 47% of adjacent physiologic tissue (Kumar *et al.*, 2017). Both studies demonstrate a dense
colonization of *SGG* in association with colorectal neoplasia affected cells. By detecting an increased load of *SGG* in the feces of CRC patients, we complement the records about local tissues with luminal data. Unfortunately, the study of Dumke et al. from 2017 does not provide any data about the feces of tumorous patients for further comparison.

### 4.1.2 *S. gallolyticus* subsp. *macedonicus* and subsp. *pasteurianus*

Thanks to the application of a sophisticated detection method, our results cover individual data for every SGMB subspecies. Whereas data on *SGG* confirms its hypothesized link with colorectal tumor growth, *SGM* and *SGP* show no elevated activity during the disease.

*SGM* is rarely spotted among all cohorts in both Series A and B. The differences between the cohorts are extremely low, ranging from 0 to 3 positively tested participants per group. These results confirm that *SGM* has a low pathogenic potential (Papadimitriou et al., 2014) and is not involved in CRA nor CRC.

*SGP* is identified more frequently than *SGM*. Among healthy and adenoma bearing volunteers the prevalence ranges from 24.0 to 30.4% in Series A and 16.0 to 17.4% in Series B. Both experiment series find *SGP* with a lower frequency in CRC patients. In Series A incidences are cut in half (13.0%) while they quarter in Series B (4.3%). The association of *S. bovis* type II with colonic cancer has been already described and we do not value the lower prevalence of *SGP* in the group of tumor patients as a pathologic abnormality (Ruoff et al., 1989; Takamura et al., 2014). Yet, the importance of *SGP* in colorectal cancer is not established.

The results outline the importance of detection methods that allow to discriminate *S. galloyticus* biotypes. Our results strongly question studies that use imprecise detection methods or an unclear taxonomy of *S. galloyticus* subspecies. In agreement with our opinion, a meta-analysis including 6 studies that discriminated *S. bovis* biotype I and II revealed a significant correlation between the presence of *S. bovis* biotype I and the risk of CRC in comparison with *S. bovis* type II infection (Boleij et al., 2011). We emphasize that a majority
of the inhomogeneous data about the colonization of SGG and CRC exist due to a lack of detection methods that allow to distinguish subspecies. Further, detailed information about the tumor stage is needed for a clear classification.

4.1.3 Simultaneous presence of S. gallolyticus subspecies

Further, we analyzed if S. gallolyticus subspecies occur as single- or co-colonizers with other SGMB strains in individual samples. Concomitant colonization of individuals with two subspecies exists (Dumke et al., 2017; Lopes et al., 2014). Although we find overlapping subspecies in most of the cohorts, the numbers in the groups are low and statically not significant, no suspicious correlation is registered. Consequentially, we do not assume any mutual interaction or influence between SGG, SGM, and SGP.

4.2 Review about Series A and Series B

As described above, the evaluation of the different cohorts preceded a long test period of countless experiments to set up and adjust the final protocol for this work. Preparations and processing of DNA extractions, innumerable PCRs and a diversity of other molecular assays were conducted. In a large part the process took place in the same laboratories. Albeit meticulous precautions against contamination were taken, genetic traces of S. gallolyticus seemed to accumulate over time. Series A was carried out in direct chronology to the preparation phase and most likely suffered from its influence. Despite the effort, it seems like contaminations falsified the results of a few samples from Series A. Series B was conducted with a time lap in a different laboratory using new materials only. Hence, we believe that Series B correctly reflects the distribution of S. gallolyticus among our groups. Generally, to minimize such risks different rooms should be used for:

a. master mix preparation,

b. nucleic acid extraction and DNA template addition,

c. amplification and handling of amplified product, and

d. product analysis, e.g. gel electrophoresis.
A setting with four rooms is not always available. A less preferable option is a laminar flow cabinet for DNA processing equipped with UV light as it is used for clinical diagnosis (clean room) (WHO, 2018).

4.3 Outlook

The genesis of CRC is a multifactorial process. In the everyday clinical practice, it is often impossible to define the individual causes that induced CRC in a patient. Though, the knowledge about the interaction of risk factors with the human body helps in the prevention, screening and treatment of CRC. The cellular stress of various internal and external factors accumulates over a lifetime and can activate the malignant transformation of colorectal cells (Jones et al., 2008). The investigated risk factor in this study, a colorectal colonization with SGG, could be associated with patients suffering from CRC. However, SGG was found in healthy volunteers as well as CRC could be present in SGG negative patients. The results highlight that the colonization of the big bowl by SGG could be one among many risk factors that increase the probability to develop CRC.

SGG’s carcinogenic potential in the colorectal tumor progression depends on many specific factors. Characteristics such as the density of SGG colonies, their individual virulence, and existing premalignant lesions in the big bowl need be taken in account. A recent model suggests that SGG acts as a non-severe colonizer of the big bowl in healthy humans and might turn into a cell proliferating agent in consequence to premalignant cell changes (Kumar et al., 2017). Following this idea, the quantification of SGG in healthy and sick people should be analyzed to investigate its hypothesized double role. Applied to the protocol of this study, new SodA oligonucleotides with a shorter product could be more precise for a quantification of SGG with RT-PCR. The sodA gene excels in the detection of Streptococci with molecular methods and its different sequences allow a perfect identification of subspecies. Such investigations could include other bacteria linked with CRC as well, such as Parvimonas micra, Bacteroides fragilis, Fusobacterium nucleatum and genotoxic strains of Escherichia coli. Side symptoms in colorectal cancer encompass alternating bowel
movement, pending between obstruction and diarrhea. Accordingly, the excretion of bacteria can oscillate between defecations. Fecal samples of one individuum, isolated at different points in time should determine an average fecal *SGG* carriage and prevent non-representative snapshots. As described by previous studies, *SGG* favors the proximity to neoplastic tissue, is able to adhere to it and to profit in several ways from the tumor underlying environmental changes. The proximity to neoplastic colorectal tissue seems to initiate a concatenation of growth beneficial circumstances for *SGG*. The favored adhesion of *SGG* to neoplastic tissue might lead to a dense colonization in the area and displacement of other microbes. *SGG* is conversely enabled to influence the neoplastic tissue in a cell proliferating manner. It has been observed that *SGG* can be observed densely attached to neoplastic colon tissue while this does not apply for adjacent tissue samples (Abdulamir *et al.*, 2010, Boleij *et al.*, 2012, Kumar *et al.*, 2017). However, an elaborate colonoscopy is the only access to the intestinal mucosa and diagnostic biopsy is only arguable with reasonable suspicion. As a result, *SGG* in vivo studies of colorectal tissues are complicated to realize in healthy humans due to a lack of biopsies. Yet, clinical evidence of *SGG* in biopsies of colorectal neoplasia could help to define its prognosis. Our method is a highly sensitive and specific detection method for *S. gallolyticus* subspecies in feces. Though, our study does not allow any statement whether the luminal prevalence of *SGG* appropriately reflects the tissue associated colonization. Yet, if *SGG* colonization of the bowl will become more relevant as a risk assessment parameter for colorectal cancer in the future, practicable screening methods such as stool investigations will be crucial.
5 SUMMARY

*Streptococcus gallolyticus* subspecies *gallolyticus* (*SGG*) is an opportunistic pathogen in humans that causes various infective diseases and is strongly associated with colorectal tumor growth. Other *Streptococcus gallolyticus* subspecies (*SG*), *Streptococcus gallolyticus* subspecies *macedonicus* (*SGM*) and subspecies *pasteurianus* (*SGP*), are not linked with colorectal cancer (CRC). The significance to distinguish between the subspecies has not always been clear which might has led to ambiguous data about the prevalence of *SG* subspecies in the population. We aimed to establish a reliable feces-based detection technique to investigate *SG* strains on a subspecies level which should help to increase our knowledge about their prevalence among healthy persons and such with colorectal conditions. In our study we have investigated the luminal *SGG* carriage using the stools of 71 French volunteers, coholed in a healthy, colorectal adenoma and CRC bearing group. A real-time PCR based assay was used to detect *SG* subspecies in feces extracted DNA. Various oligonucleotides were tested for their specificity and sensitivity which established *SodA1/2* as the best performing pair. *SodA1/2* targets the *sodA* gene which encodes the superoxide dismutase enzyme in all *Streptococci* subspecies. Due to the ubiquitous presence of *sodA*, its informative value is limited to the identification of *SG* at species level. To distinguish *SG* at subspecies level, the nucleotide sequence of the *sodA* PCR product was determined to define the subspecies in the stool samples. Our investigation revealed an elevated colonization of *SGG* in the gastrointestinal tract of patients with a history of CRC (43.5%). *SGG* was less frequent in healthy (28.0%) and colorectal adenoma bearing volunteers (30.4%), nevertheless indicated a significant carriage among these groups as well. The prevalence of *SGM* and *SGP* is not associated with colorectal neoplasia which outlines the importance of an *SGG* specific identification. Our study is limited to bacteria detection in stool, corresponding to the luminal *SG* carriage in the gut of a small French population. An additional examination of tissue
samples that reflect adherent mucosal bacteria as well as a larger population of broader geographical background is needed to gain a comprehensive view about the prevalence of *S. gordonii* and to establish *S. gordonii* as a recognized risk factor for CRC.
6 REFERENCES


REFERENCES


36. Jans C. (2012) Biodiversity of lactic acid bacteria in raw camel milk products of East Africa including genomic and functional characterization of


APPENDIX

1. Additional information about Materials

1.1 Bacteria

1.1.1 Monocultures

*Bacteroides fragilis*  
CIP 77.16T

*Enterococcus faecalis*  
murine isolate

*Enterococcus faecium*  
Com12

*Enterococcus hirae*

*Fusobacterium nucleatum*  
*subspecies animalis*

*Parvimonas micra*  
BRFIo.2017

*Streptococcus gallolyticus*  
*subspecies gallolyticus*

*Streptococcus gallolyticus*  
*subspecies macedonicus*

*Streptococcus gallolyticus*  
*subspecies pasteurianus*

*Streptococcus infantarius*  
*subspecies coli (S. lutetiensis)*

*Streptococcus infantarius*  
*subspecies infantarius*
1.1.2 Mix cultures

*Streptococcus gallolyticus subspecies gallolyticus + E. faecalis + E. faecium + E. hirae*

*Streptococcus gallolyticus subspecies gallolyticus + Lactobacilli subspecies*

*Streptococcus gallolyticus subspecies gallolyticus + complete bacteria collection*

1.2 Proteins

1.2.1 Enzymes

Lysozyme, from chicken egg white

Sigma-Aldrich, St. Louis, Missouri, USA

Mutanolysin from *Streptomyces globisporus* ATCC 21553

Sigma-Aldrich, St. Louis, Missouri, USA

1.2.2 Restriction enzymes

*Bsm* I

Fermentas, Thermo Fisher Scientific, Marietta, OH, USA

*Hind*III

Thermo Science, Thermo Fisher Scientific, Marietta, OH, USA

*Mse* I

Fermentas, Thermo Fisher Scientific, Marietta, OH, USA

*SfaN* I

Thermo Science, Thermo Fisher Scientific, Marietta, OH, USA

*Xba* I

BioLabs, Ipswich, MA, USA
### 1.3 Chemicals reagents

<table>
<thead>
<tr>
<th>Item</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distilled DNase/RNase free Water</td>
<td>Invitrogen, Thermo Fisher Scientific, Marietta, OH, USA</td>
</tr>
<tr>
<td>Molecular DNA marker</td>
<td>Thermo Fisher Scientific, Marietta, OH, USA</td>
</tr>
<tr>
<td>GeneRuler 1kb</td>
<td></td>
</tr>
<tr>
<td>DNA Loading Dye 6x</td>
<td>Thermo Fisher Scientific, Marietta, OH, USA</td>
</tr>
<tr>
<td>DNA polymerase BIOTAQ</td>
<td>Bioline, London, UK</td>
</tr>
<tr>
<td>dNTP Mix, 1ml, 25 mM each</td>
<td>Thermo Fisher Scientific, Marietta, OH, USA</td>
</tr>
<tr>
<td>EDTA</td>
<td>Sigma-Aldrich, St. Louis, Missouri, USA</td>
</tr>
<tr>
<td>Ethanol, absolute</td>
<td>Fisher Scientific UK, Loughborough, UK</td>
</tr>
<tr>
<td>Ethidium bromide 0.5mg/ml solution</td>
<td>Euromedex, Straßburg, France</td>
</tr>
<tr>
<td>FastDigest Green Buffer 10x</td>
<td>Thermo Fisher Scientific, Marietta, OH, USA</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>Sigma-Aldrich, St. Louis, Missouri, USA</td>
</tr>
<tr>
<td>MgCl₂ Solution 50mM</td>
<td>Bioline, London, UK</td>
</tr>
<tr>
<td>NH₄ (10x) Buffer</td>
<td>Bioline, London, UK</td>
</tr>
<tr>
<td>SsoFast™ EvaGreen® Supermix</td>
<td>Bio-Rad Laboratories GmbH, München, Germany</td>
</tr>
<tr>
<td>SYBR™ Gold Nucleic Acid Gel Stain</td>
<td>Invitrogen, Thermo Fisher Scientific, Marietta, OH, USA</td>
</tr>
<tr>
<td>TAE buffer (50x)</td>
<td>Invitrogen, Thermo Fisher Scientific, Marietta, OH, USA</td>
</tr>
</tbody>
</table>
APPENDIX

Taq DNA Polymerase, recombinant Invitrogen, Thermo Fisher Scientific, Marietta, OH, USA
Tris, Trizma Base Sigma-Aldrich, St. Louis, Missouri, USA
UltraPure Agarose Invitrogen, Thermo Fisher Scientific, Marietta, OH, USA

1.4 Consumables

CapitolBrand® Genomics SP-0027 Capitol Scientific, Inc, Adhesive PCR plates film seals Austin, TX, USA

Cell culture flask 50ml, 100ml, 250ml Fisher Scientific GmbH, Schwerte, Germany
Falcon conical tubes 25ml, 50ml Corning Science, Reynosa, Mexico
Filter tips, Neptune Bt 1250.N, 200, 20, 10XL Dutscher, Brumath, France
FrameStrip® 8 Strip PCR tubes 4titude® Limited, Wotton, UK
FrameStar® 96-well PCR plate 4titude® Limited, Wotton, UK
Incubation loops 10µl Sarstedt, Nürnbrecht, Germany
Microtubes, 1,5 and 2,0 ml, PP with attached PP cap Sarstedt, Nürnbrecht, Germany
Micro tube 2ml, PP, PP screw-cap Sarstedt, Nürnbrecht, Germany
Non-filter tips, SureOne 20µl, 200µl, 1000µl Thermo Fisher Scientific, Marietta, OH, USA
1.5 Kits

BIOTAQ DNA Polymerase Kit  
Bioline, London, UK

DNeasy® Blood & Tissue Kit (50)  
Qiagen GmbH, Hilden, Germany

GENECLEAN® Turbo Kit  
MP Biomedicals, Santa Ana, California, USA

UltraClean® GelSpin® DNA Extraction Kit  
Qiagen GmbH, Hilden, Germany

Wizard® Genomic DNA Purification Kit  
Promega Corporation Madison, WI, USA

1.6 Technical equipment

Applied Biosystems 2720 Thermal Cycler  
Thermo Fisher Scientific, Marietta, OH, USA

BioSpectrometer® basic  
Eppendorf AG, Hamburg, Germany

Centrifuge 5424 R  
Eppendorf AG, Hamburg, Germany  
(Rotor: FA-45-24-11)

CFX96 Real-Time System  
Bio-Rad Laboratories GmbH, München, Germany

Eppendorf Research® plus pipettes  
Eppendorf AG, Hamburg, Germany

EP-2117: RunOne™ Electrophoresis Unit  
Embi Tec, San Diego, CA, USA

Fuego SCS pro  
WLD-TEC GmbH, Göttingen, Germany

Gel Doc XR+ System  
Bio-Rad Laboratories GmbH, München, Germany

Mettler PE 160 Scale  
Mettler-Toledo, Columbus, OH, USA
MiniSpin® plus centrifuge  Eppendorf AG, Hamburg, Germany
NanoDrop™ 2000c Spectrophotometers  Thermo Fisher Scientific, Marietta, OH, USA
Pipetboy acu pipetting system  Integra Biosciences, St Ouen l'Aumône, France
Qubit 4 Fluorometer  Thermo Fisher Scientific, Marietta, OH, USA
Top-Mix 94323 vortex mixer  Heidolph, Schwabach, Germany
Thermomixer® comfort  Eppendorf AG, Hamburg
Thermomixer® compact  Eppendorf AG, Hamburg

1.7 Bacteria culture media

BHI (Brain Heart Infusion broth)  Sigma-Aldrich, St. Louis, Missouri, USA

CHB (Columbia 5% Horse Blood agar)  Bio-Rad Laboratories GmbH, München, Germany

DMEM (Dulbecco's Modified Eagle Medium) High glucose agar + 10% Gibco Bovine Serum (heat inactivated, New Zealand Origin)  Thermo Fisher Scientific, Marietta, OH, USA

Difco™ Lactobacilli MRS (De Man, Rogosa and Sharpe) Agar, BD  Fisher Scientific, Hampton, NH, USA

TGY (Tryptone Glucose Yeast Extract)  Fisher Scientific, Hampton, NH, USA
Broth, m - DIFCO® Dehydrated
Culture Media and Ingredients

THY (Todd Hewitt Yeast Broth)  
– DIFCO® Dehydrated  
Fisher Scientific, Hampton, NH, USA
Culture Media and Ingredients

1.8 Software

ApE Plasmid Editor for MacOSX: 2.0.53  
M. Wayne Davis

CFX Manager 3.0 Software  
Bio-Rad Laboratories GmbH, München, Germany

DNA Strider for MacOSX: 1.59.2  
Wordpress.org

Quantity One 4.6.2.  
Analysis software  
Bio-Rad Laboratories GmbH, München, Germany

2. Additional information about methods

2.1 Bacteria cultivation, conditions and media

All bacteria stocks are regularly stored at -80°C.

a. *Bacteroides fragilis* (CIP 77.16T) has been cultivated for 48h at 37°C under anaerobic conditions on a Columbia 5% horse blood agar.

b. *Streptococcus gallolyticus subspecies gallolyticus + E. faecalis + E. faecium + E. hirae*

*Streptococcus gallolyticus subspecies gallolyticus + Group B Streptococci*
Bacteria were cultivated for 18 h at 37°C static under microaerophilic conditions in a final volume of 10 mL of BHI in a 15 mL Falcon tube

c. *Fusobacterium nucleatum* (CIP 104879) has been cultivated for 24 h at 37°C under anaerobic conditions on a Columbia 5% horse blood agar.

d. *Streptococcus gallolyticus subspecies gallolyticus* + *Lactobacilli subspecies* were cultivated for 18 h at 37°C static under microaerophilic conditions in 10 mL of MRS medium in a 15 mL Falcon tube.

e. *Lactobacilli subsp.* was cultivated for 18 h at 37°C static under anaerobic conditions (anaerobe jar) in MRS medium.

f. *Parvimonas micra* (BRFIo2017) has been cultivated for 48 h at 37°C under anaerobic conditions on a Columbia 5% horse blood agar.

g. *Enterococcus faecalis*
   *Enterococcus faecium*
   *Enterococcus hirae*
   *Streptococcus gallolyticus subspecies gallolyticus*
   *Streptococcus gallolyticus subspecies macedonicus*
   *Streptococcus gallolyticus subspecies pasteurianus*
   *Streptococcus infantarius subspecies coli* (*S. lutetiensis*)
   *Streptococcus infantarius subspecies infantarius*

The bacteria were cultivated for 18 h at 37°C static under microaerophilic conditions in a medium volume of 10 mL THY in a 15 mL Falcon tube.
h. *Streptococcus gallolyticus subspecies gallolyticus*

*Streptococcus gallolyticus subspecies gallolyticus + All other named bacteria*

Bacteria were cultivated for 18 h at 37°C static under microaerophilic conditions in a medium volume of 10mL DMEM HG + 10% SVF in a 15 mL Falcon tube

### 2.2 DNA extraction from fecal samples

The modified protocol from Ahlroos and Tynkkynen (2009) for the Promega Wizard® Genomic DNA Purification Kit

#### I. Chemicals

EDTA 50mM, Lysozyme 50mg/ml (-20°C), Mutanolysin 5U/µl (-20°C), Nuclei Lysis Solution (kit), Rnase Solution (kit), Protein Precipitation Solution (kit), Isopropanol, Ethanol 70%, DNA Rehydration Solution (kit)

#### II. Material

Mini-spatulas, Micropipettes with filter tips, Microcentrifuge, Microtubes Safelock with 1.5 – 2mL, Vortex, Thermomixer at 37°C, Water bath at 65°C and 80°C, ice.

#### III. Principle of the method

After homogenizing the samples, the cells were lysed, bacterial RNA was eliminated, and proteins precipitated. The process was completed by precipitating, washing and rehydrating the DNA.

#### IV. Procedure

The stool samples (150 mg) were slowly defrosted on ice, transferred in 2 mL tubes and stored on ice until further processing. 750 µl of EDTA 50 mM were added to the samples and vortexed until partial homogenization. After adding another 750µlEDTA 50mM the homogenization was continued until completion.
The samples were centrifuged at 14,000 rpm for 2 min and the supernatant removed. The bacterial pellet was resuspended in the lysis buffer (480 µl EDTA 50mM, 100 µl lysozyme 50 mg/ml and 20 µl mutanolysin 5U/µl) and incubated at 37°C/800 rpm/1 h.

After another centrifugation at 14,000 rpm for 2 min 600 µl of Nuclei Lysis Solution it was incubated at 80°C for 5 min. 3 µl of RNAse Solution was added the samples were then incubated 1 hour at 37°C/800 rpm.

200 µl of Protein Precipitation Solution was added, the sample were vortexed and then centrifuged at 16,000 rpm for 3 min. The supernatant was carefully transferred in 1.5 mL Eppendorf tubes containing 600 µl of isopropanol at room temperature.

The samples were then centrifuged at 16,000 rpm for 2 min and the precipitated DNA was washed with ethanol 70%.

A last centrifugation step at 16,000 rpm for 2 min was required before removing the ethanol. The tubes were left open at 37°C until complete ethanol evaporation. DNA was rehydrated in provided DNA Rehydration Solution at 4°C overnight. After verification of the quality and quantity measurement, samples were aliquoted.

V. **Quality and quantity review of fecal extracted DNA**

The DNA quality was measured through the ratio of absorbance at 260-280 nm. The DNA quantity was defined by RT-PCR performed with a 16s rRNA oligonucleotide pair (AllBacteria) specific for all gram-positive bacterial DNA. The share of utilizable gram-positive bacterial DNA was measured in each sample and eventually the concentrations were equalized using an array of well-defined samples.
2.3 Polymerase chain reaction methods

2.3.1 Classic PCR

1. Protocol

The reaction mix contained the following components:

1. PCR buffer (10x) 1x
2. MgCl₂ (10x) 1.5 mM
3. Primer 1 20 pmoles
4. Primer 2 20 pmoles
5. dNTP 10 mM
6. DNA polymerase 0.5 U
7. Template-DNA 1 µg

The standard PCR program is listed below:

1. Initial denaturation 94°C 3:00 min
2. Denaturation 94°C 0:20 min
3. Annealing x 0:30 min (depends of primers used)
4. Elongation 72°C 0:30 min / 1kb
5. Terminal Elongation 72°C 5:00 min
6. storage 4°C -

At the end of the program, 5 µl of each PCR product was analyzed by electrophoresis on agarose gel.
2.3.2 Real-time PCR

I. Protocol

The reaction mix contained the following components:

1. SsoFast™ EvaGreen® Supermix 1 x
2. Primer 1 10 pmoles
3. Primer 2 10 pmoles
4. Template-DNA 50 ng

Each experiment was set up on a 96-well PCR plate containing a positive control (genomic DNA SGG UCN 34, 10 ng /µl), a negative control (DNase/RNase free water) and the sample mixes. Each assay was performed in triplicates. To minimize the contamination risk, the positive control was always set at last.

The standard real-time PCR program is listed below:

1. Initial denaturation 95°C 3:00 min
2. Denaturation 95°C 0:10 min
3. Annealing 55°C 0:15 min 40 cycles
4. Elongation 72°C 0:25 min + plate read
5. Terminal elongation 72°C 3:00 min
6. Melt Curve 55°C to 95°C with a temperature increment of 0.5°C/0:05 min + plate read
I would like to express my deep gratitude to everyone who has directly or indirectly supported me throughout this project.

First of all, I cannot express enough my thankfulness to Prof. Dr. med. Barbara Spellerberg and Shaynoor Dramsi PhD who made the cooperation between the Ulm University and the Institut Pasteur Paris possible. This international research cooperation was the foundation for the success of my work and all the stunning experiences that have grown of it.

Furthermore, special thanks to the whole unit of Gram-positive pathogens led by Prof. Patrick Trieu-Cuot at the Institut Pasteur Paris for the wonderful and productive time. I am especially grateful to Shaynoor Dramsi PhD who has provided me with invaluable guidance throughout the research. I also sincerely thank Bruno Perichon for his relentless patience and effort.

At Ulm University I would like to thank my first and second advisor Prof. Dr. med. Barbara Spellerberg and Prof. Dr. rer. med. Alexander Kleger for their consultancy.

The Semmelweis University Budapest, especially the department of microbiology, deserves my deepest gratitude. My Alma Mater has supported and encouraged me to take my education beyond boarders and to never stop exploring.

Last but not least, I could not have finalized this project without the help of my loved ones who have stood by me unconditionally over great distances during my years of studying in Budapest and Paris. Thank you for always having my back.
CURRICULUM VITAE

Name: Julian Lichtl-Häfele
Year and place of birth: 1993, Kirchheim unter Teck, Baden-Württemberg
Nationality: German

Education and Career

2012: Abitur at Schlossgymnasium Kirchheim unter Teck
2012-2013: Apprenticeship as emergency medical technician at Medakademie Stuttgart and Daschow
2013-2020: Medical studies at Semmelweis University, German Faculty of Medicine, Budapest, Hungary
2019-2020: Thesis to obtain the degree of Doctor Medicinae (dr. med.) at Semmelweis Universität, Institute for Transplantation Medicine and Surgery led by Dr. med habil. Eva Toronyi, Budapest, Hungary
2017-2021: Dissertation to obtain the Doctoral Degree of Medicine (Dr. med.) at Ulm University, Faculty of Medical Microbiology and Hygiene led by Prof. Dr. med. Steffen Stenger