Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Julia Butt, M.Sc.
born in Cuxhaven, Germany
Oral-examination: 26.04.2017

Seroepidemiology of
Streptococcus gallolyticus subspecies *gallolyticus*
and *Fusobacterium nucleatum*
with colorectal cancer

Referees: Prof. Dr. Martin Müller
Prof. Dr. Ralf Bartenschlager

Acknowledgement

This thesis was performed under the supervision of Dr. Michael Pawlita from October 2013 to January 2017 in the German Cancer Research Center (DKFZ, Heidelberg), research program Infection, Inflammation and Cancer in the division of Molecular Diagnostics of Oncogenic Infections.

I would like to thank Prof. Dr. Martin Müller and Prof. Dr. Ralf Bartenschlager for reviewing this thesis and for their support as members of my thesis advisory committee.

Special thanks go to Dr. Michael Pawlita, for initializing the project and providing me the opportunity to work in his lab. I am thankful for the excellent supervision, our weekly discussions, your support and guidance. Moreover, I am grateful for the opportunity to work abroad to gain inside in research and culture at other places outside the DKFZ.

I would like to thank Dr. Tim Waterboer for his help and support in the project not only as part of the thesis advisory committee but also whenever else I had questions.

I would like to thank Meira Epplein, PhD, not only for providing study samples and data for the project but also for being a great mentor in epidemiology and beyond. It was a great pleasure to do research at Vanderbilt University under her supervision.

I further would like to thank collaboration partners Dr. Mazda Jenab and Dr. David Hughes as well as Prof. Dr. Hermann Brenner and Simone Werner who gave me the opportunity to analyze their study samples and data as well as supported me with their expertise and knowledge in epidemiological analyses. Thanks also go to Dr. Annemarie Boleij and Dr. Harold Tjalsma, who not only shared specimens but also their extensive knowledge on bacterial associations with colorectal cancer with me. I also would like to thank the colleagues from the Microbiological Diagnostics laboratory at the DKFZ who never hesitated to help.

I would like to thank all colleagues that make our lab a great place to work. I would like to specifically thank Ute Koch, Monika Oppenländer, Claudia Brandel and all the "Azubis" and students, especially Indra-Jasmin Gierse, for their support and help in experimental issues. I would also like to thank Dr. Angelika Michel, Dr. Martina Willhauck-Fleckenstein and Ute Koch for answering questions without losing patience and for the great (and sometimes challenging) times during conferences and in India. And I would like to thank Dr. Daniela Höfler for her company and support, not only in the lab, but also with her friendship.

I am grateful for my friends in- and outside Heidelberg that accompanied me throughout the PhD time and beyond. I also thank the Cassidy's for giving me a lovely home in Nashville. And I thank Jan-Eric Meissner, who gave constant support especially in the last tedious times of the thesis. Last but not least, möchte ich meiner Familie danken, meinen Eltern Walter und Gisela Butt und meinen Geschwistern Sabine Lenz und Eike Butt für ihr Vertrauen, ihre Unterstützung, Rückzugsorte und die Ablenkung. Ein besonderer Dank geht ebenso an die lieben Kleinen, die einfach immer für gute Laune sorgen.

Summary

Colorectal cancer (CRC) is among the most frequently diagnosed cancers worldwide. Recent research focused on the association of CRC with an altered microbiome. More specifically, two bacteria, *Fusobacterium nucleatum* (*F. nucleatum*) and *Streptococcus gallolyticus* subspecies *gallolyticus* (*S. gallolyticus*) were individually brought in context with CRC. *F. nucleatum* is predominantly present in oral plaques and was found to be abundant in stool and tumor tissue of CRC patients. *S. gallolyticus* is a rare commensal in the human intestine and inducer of infective endocarditis that is associated with presence of CRC.

The aim of this thesis was to explore potential serological associations of *F. nucleatum* and *S. gallolyticus* with CRC using multiplex serology, a high-throughput technology that allows the analysis of large seroepidemiological studies. Multiplex serology was to be developed for *F. nucleatum* and *S. gallolyticus* and applied in a retrospective case-control study to analyze potential serological associations with CRC. Prospective studies were to be analyzed to give information on temporality of the association: if serological associations are present prior to diagnosis, these antibodies might serve as early marker for risk of developing CRC.

Eleven proteins for each, *F. nucleatum* and *S. gallolyticus*, were selected, recombinantly expressed and applied in multiplex serology. Serological validation of the assays was possible only to a limited extent due to a lack of a gold standard assay for comparison. Cut-offs for antibody-positivity to the individual proteins were arbitrarily defined to allow for 10% of controls as positive. Antibody responses to *F. nucleatum* and *S. gallolyticus* were analyzed in a retrospective case-control study conducted in Germany and two independent case-control studies nested within multi-center prospective cohorts from Europe and southern United States. Positivity to any of the *F. nucleatum* proteins was not associated with CRC, neither retro- nor prospectively. In contrast, odds for prevalent and incident CRC in the German case-control study as well as the European prospective study were significantly 2-fold increased with positivity to two or more proteins of a *S. gallolyticus* 6-marker panel. However, this association was not found in the southern United States study.

In conclusion, antibody responses to *S. gallolyticus*, but not *F. nucleatum*, were significantly associated with CRC prior to diagnosis and might serve as marker for CRC development. A causal relationship of *S. gallolyticus* with CRC cannot be inferred from the generated data, however, results of this thesis might stimulate research on the involvement of *S. gallolyticus* in CRC development as well as risk factors leading to *S. gallolyticus* colonization.

Zusammenfassung

Dickdarmkrebs ist eine der am häufigsten diagnostizierten Krebsarten weltweit. Aktuelle Studien untersuchen die Assoziation von Dickdarmkrebs mit einer Veränderung des bakteriellen Mikrobioms. Zwei bakterielle Spezies, *Fusobakterium nucleatum* (*F. nucleatum*) und *Streptokokkus gallolyticus* subspezies *gallolyticus* (*S. gallolyticus*), sind dabei besonders im Zusammenhang mit Dickdarmkrebs aufgefallen. *F. nucleatum* wird im Menschen hauptsächlich in Zahnbelägen gefunden, ist aber auch in Stuhl- und Tumorgewebe-Proben von Dickdarmkrebspatienten identifiziert worden. *S. gallolyticus* ist ein seltener Kommensal im menschlichen Gastrointestinaltrakt, aber auch ein Auslöser der infektiösen Endokarditis, welche wiederum assoziiert mit Dickdarmkrebs auftritt.

Das Ziel der vorliegenden Arbeit war die Untersuchung der serologischen Assoziation von *F. nucleatum* und *S. gallolyticus* mit Dickdarmkrebs. Dafür wurden pro Bakterium jeweils elf Antigene entwickelt und mittels Multiplex-Serologie zur Antikörper-Analyse in einer deutschen retrospektiven Fall-Kontroll-Studie eingesetzt. Des Weiteren wurden zwei unabhängige prospektive Studien, eine multinationale europäische und eine US-amerikanische Kohorte, auf Antikörper gegen *F. nucleatum* und *S. gallolyticus* untersucht, um den zeitlichen Zusammenhang der Assoziation zu analysieren. Wenn serologische Assoziationen vor der Dickdarmkrebs-Diagnose messbar sind, wären die Antikörper potentielle Marker für ein erhöhtes Dickdarmkrebs-Risiko.

Da keine serologischen Goldstandard-Testverfahren für *F. nucleatum* und *S. gallolyticus* verfügbar sind, wurde der Cut-Off für Antikörper-Positivität arbiträr festgelegt und definiert 10% der Kontrollen als Antikörper-positiv. Antikörper-Positivität für *F. nucleatum* war nicht assoziiert mit Dickdarmkrebs, weder retro- noch prospektiv. Im Gegensatz dazu waren die Antikörper-Antworten gegen *S. gallolyticus* in der deutschen Fall-Kontroll-Studie und der prospektiven europäischen, aber nicht der US-amerikanischen, Kohorte mit einem 2-fach erhöhten Risiko für Dickdarmkrebs assoziiert.

Zusammenfassend waren Antikörper-Antworten gegen *S. gallolyticus,* aber nicht *F. nucleatum,* assoziiert mit einem erhöhten Dickdarmkrebs-Risiko und sind potentielle Marker für die Erkrankung. Ein kausaler Zusammenhang einer *S. gallolyticus* Infektion mit Dickdarmkrebs kann aus den präsentierten Analysen nicht hergeleitet werden, jedoch sind die Ergebnisse der Arbeit ein möglicher Anknüpfungspunkt für zukünftige Forschung zum Einfluss einer *S. gallolyticus* Infektion auf die Tumor-Entwicklung im Dickdarm.

Contents

ACKNOWLEDGEMENT **I**

SUMMARY **III**

ZUSAMMENFASSUNG **IV**

CONTENTS **V**

1. INTRODUCTION **1**

1.1. Colorectal cancer (CRC) **1**

1.1.1. Epidemiology and risk factors 1

1.1.2. CRC development 2

1.1.3. CRC screening 3

1.2. Gastrointestinal Bacteria in Health and Disease **4**

1.2.1. *Fusobacterium nucleatum* and CRC 6

1.2.2. *Streptococcus gallolyticus* subsp. *gallolyticus* and CRC 9

1.3. Seroepidemiology of Infection-Associated Cancers **12**

1.3.1. Epidemiological study designs and measures of association 12

1.3.2. Infection-associated cancers and serology 15

1.4. Multiplex Serology **16**

1.5. Objectives **17**

2. MATERIALS AND METHODS **18**

2.1. Materials **18**

2.1.1. Chemicals 18

2.1.2. Buffers 19

2.1.3. Consumables 20

2.1.4. Laboratory equipment 21
2.1.5. Antibodies 21
2.1.6. Bacterial strains 22
2.1.7. Enzymes and reaction buffers 22
2.1.8. Antigens 22
2.1.9. Sera and respective study data 22
2.1.10. Software and websites 23

2.2. Methods 24

2.2.1. Selection and cloning of proteins for *F. nucleatum* and *S. gallolyticus* multiplex serology 24
2.2.2. Recombinant expression of selected *F. nucleatum* and *S. gallolyticus* proteins in *E. coli* BL21 29
2.2.3. Bacterial lysis 30
2.2.4. Quantification of lysate protein concentration by Bradford assay 30
2.2.5. Analytical DNA digestion 30
2.2.6. Agarose gel electrophoresis 31
2.2.7. Polymerase chain reaction with subsequent sequencing 32
2.2.8. Western blot 33
2.2.9. Anti-tag ELISA 34
2.2.10. Multiplex serology 35
2.2.11. Study designs 36
2.2.12. Statistical analyses 47

3. RESULTS 49

3.1. Expression and quality control of *F. nucleatum* and *S. gallolyticus* proteins 49

3.1.1. Cloning and expression 49
3.1.2. Verification of the expression constructs by analytical digestion and PCR with subsequent sequencing 50
3.1.3. Verification of full-length expression by Western blot and anti-tag ELISA 54
3.1.4. Loading of GST-X-tag fusion proteins onto beads 58

3.2. Validation of *F. nucleatum* and *S. gallolyticus* multiplex serology 60

3.2.1. Antibody responses to *F. nucleatum* in comparison to *F. nucleatum* DNA in colon tissue 60

3.2.2. *S. gallolyticus* multiplex serology in sera from experimentally infected mice 64

3.2.3. *S. gallolyticus* multiplex serology in bovine sera 65

3.2.4. *S. gallolyticus* multiplex serology in humans with *S. gallolyticus* bacteremia 68

3.3. Search for serological associations of *F. nucleatum* with CRC by multiplex serology in a retrospective case-control study and prospective nested case- control studies 72

3.3.1. Cut-off definition 72

3.3.2. Risk factors for antibody-positivity to any *F. nucleatum* protein among controls 76

3.3.3. Antibody responses to *F. nucleatum* and prevalent CRC in the BliTz-DACHSplus study 78

3.3.4. Antibody responses to *F. nucleatum* and incident CRC in the SCCS study 82

3.3.5. Antibody responses to *F. nucleatum* and incident CRC in the EPIC study 85

3.4. Search for serological associations of *S. gallolyticus* with CRC by multiplex serology in a retrospective case-control study and prospective nested case-control studies 88

3.4.1. Cut-off definition 88

3.4.2. Risk factors for antibody-positivity to any *S. gallolyticus* protein among controls 92

3.4.3. Antibody responses to *S. gallolyticus* and prevalent CRC in the BliTz-DACHSplus study 94

3.4.4. Antibody responses to *S. gallolyticus* and incident CRC in SCCS 101

3.4.5. Antibody responses to *S. gallolyticus* and incident CRC in the EPIC study 104

3.4.6. Exploration of *S. gallolyticus* multiplex serology in precancerous colorectal lesions 107

4. DISCUSSION 111

4.1. *F. nucleatum* multiplex serology and CRC 112

4.1.1. Development of *F. nucleatum* multiplex serology 112

4.1.2. Antibody responses to *F. nucleatum* and CRC in a retrospective case-control study and two independent case-control studies nested within prospective cohorts 115

4.2. *S. gallolyticus* multiplex serology and CRC 118

4.2.1. Development of *S. gallolyticus* multiplex serology 118

4.2.2. Antibody responses to *S. gallolyticus* and CRC in a German case-control study and an independent case-control study nested within a European prospective cohort 120

4.2.3. Antibody responses to *S. gallolyticus* and CRC in a case-control study nested within a southern US prospective cohort 125

4.3. Natural history of *F. nucleatum* and *S. gallolyticus* infection **127**

4.4. Conclusions and Outlook **128**

5. REFERENCES **132**

APPENDIX **146**

I. Vector maps **146**

II. Alignments of sequenced inserts to reference insert sequences **158**

III. Alignment of *S. gallolyticus* proteins Gallo0577, Gallo1570 and Gallo2179 **190**

IV. Alignment of *F. nucleatum* proteins Fn0387, Fn1449 and Fn1893 **191**

V. Different models for estimating the association of antibody responses to *F. nucleatum* with CRC in BliTz/DACHSplus, SCCS and EPIC **192**

VI. Different models for estimating the association of antibody responses to *S. gallolyticus* with CRC in BliTz/DACHSplus, SCCS and EPIC **197**

VII. Abbreviations **202**

VIII. Publications **205**

1. Introduction

1.1. Colorectal cancer (CRC)

1.1.1. Epidemiology and risk factors

Colorectal cancer (CRC) is among the most frequently diagnosed cancers. According to the World Health Organization (WHO) it accounted for the third most common cancer worldwide with 1.36 million newly diagnosed cases in 2012. Mortality was lower with 693,933 deaths worldwide in 2012. The incidence was higher in men (746,000 new cases in 2012) than in women (614,000 new cases in 2012) (Figure 1) [1] and CRC risk increases with older age [2]. CRC incidence varied strongly between developed (737,000 new cases in 2012) and developing regions (624,000 new cases in 2012) in the world. Thereby, age-standardized incidence rates (ASR) were highest in Australia and New Zealand (44.8 and 32.3 per 100,000 in men and women) and lowest in Western Africa (4.5 and 3.8 per 100,000) (Figure 1) [1]. The higher incidence in more developed regions has been attributed to risk factors associated with a more „Western“ lifestyle, including changes in diet and other lifestyle factors. Indeed, smoking [3], excessive alcohol consumption [4], high consumption of processed red meat [5], lowered physical activity [6] and associated morbidities like obesity and diabetes [7, 8] were identified as risk factors for CRC. Interestingly, also inflammatory diseases of the bowel were found to increase CRC risk [9], which is emphasized by the finding that long-term use of non-steroidal anti-inflammatory drugs (NSAID) like Aspirin has protective effects [10]. Apart from these environmental risk factors, family history of CRC was identified as another strong risk factor indicating a hereditary component of the disease [11]. Approximately 5% of all CRC cases have been even attributed to hereditary syndromes. The most frequent is the so-called Lynch-syndrome or hereditary non-polyposis CRC [12], followed by familial adenomatous polyposis coli [13].

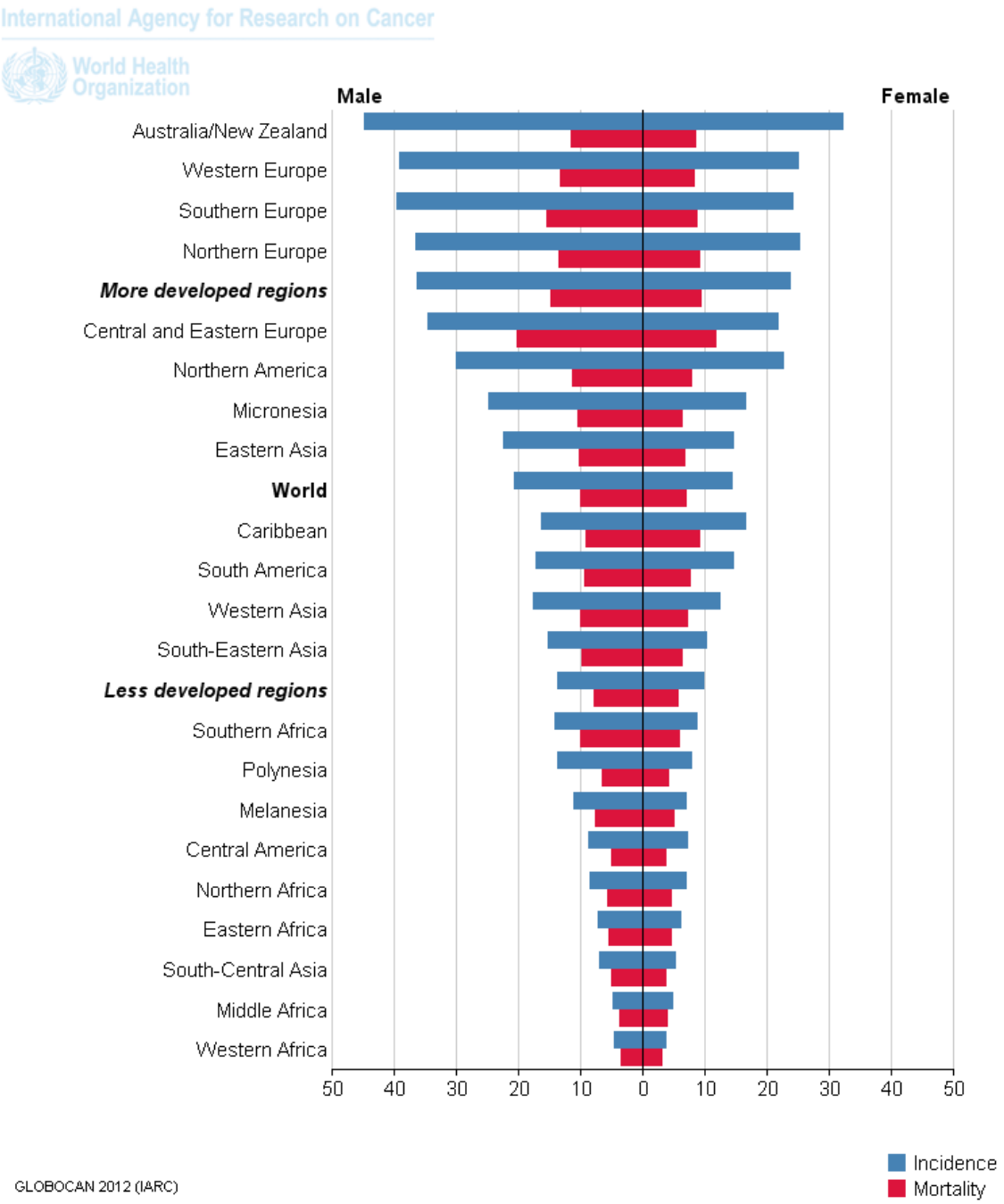


Figure 1: Estimated age-standardized incidence and mortality rate (new cases per 100,000 and year) for CRC, by sex and region [1].

1.1.2. CRC development

The majority of CRC cases (~70%), however, develop sporadically with a multistep process of genetic and morphological changes over a long period of time. Two pathways were described: the conventional adenomatous and the serrated pathway [14, 15]. The majority of

sporadic CRC (85-90%) follow the conventional adenomatous pathway [14]. Chromosomal instability leads to mutations, classically in the *APC* (adenomatous polyposis coli) gene, which results in hyperproliferation of the tissue and development of polyps [16]. Acquisition of additional mutations, mostly in the *KRAS* (Kirsten rat sarcoma) and tumor suppressor *p53* genes, then further promotes progression from adenoma to CRC [16]. The histology and morphology are thereby indicators for the risk of CRC development out of adenomatous polyps: Adenomatous polyps larger than 1 cm, those with high-grade dysplasia as well as those consistent of mainly villous, in contrast to tubular architecture, are often termed advanced adenoma and considered to have a faster progression [14].

Approximately 10-15% of sporadic CRC cases are attributed to the serrated pathway [14]. This pathway is characterized by initial activating mutations in the *BRAF* oncogene but also by the CpG island methylator phenotype (CIMP) resulting in hypermethylation and thus deactivation of relevant promotors [17]. A frequent epiphenomenon in the serrated pathway is a high degree of microsatellite instability (MSI-H) resulting from an inactivation of mismatch repair genes. MSI-H adenomas have an increased susceptibility to the acquisition of additional cancer-relevant mutations [17]. Based on their morphology serrated adenomas can be further subdivided into three different types: most serrated adenomas (70-95%) are considered hyperplastic polyps, which are small (< 5 mm) and most frequently occur in the rectosigmoid colon; traditional serrated adenomas are rare (< 1%), often pedunculated and most frequently occur in the distal colon or rectum; sessile serrated adenomas, which account for 5-25% of serrated adenomas, are flat and mostly found in the proximal colon [17]. Whether adenomas of the serrated pathway have a similar risk of progression to CRC as the conventional adenomatous pathway is unclear [17].

1.1.3. CRC screening

Early identification of colorectal neoplasm has led to a decrease in incidence but also mortality of CRC in the recent years [18]. Several countries introduced screening programs based on mainly two diagnostic methodologies: highly sensitive but invasive colonoscopy/sigmoidoscopy and less invasive but also less sensitive blood in stool tests [15]. In Germany, for example, blood in stool test by guaiac-based fecal occult blood test (gFOBT) is covered by health insurances for all individuals above 50 years in 2 years intervals [19]. Colonoscopy is recommended every 10 years for individuals between age 55 and 74. Other European countries like Denmark or Netherlands have not introduced a CRC screening

program by 2008, others like France based their screening programs only on gFOBT [19]. Colonoscopy offers a very sensitive method with the major advantage that identified precancerous lesions can be removed directly during the procedure [15]. However, it is also more unpleasant for the patient and more cost- and time-intensive for the health system than testing for blood in stool. Blood in stool tests, especially fecal immunochemical test (FIT), provide a good sensitivity in detecting CRC [15]. A German study compared gFOBT and FIT and found sensitivities of 33.3% and 73.3% in detecting CRC at a specificity of 95%. Sensitivity for advanced adenomas, however, was low with 8.6% and 23.4% [20]. A meta-analysis by Niedermaier et al. showed the same with sensitivities for detecting CRC by FIT ranging from 48% to 95% at specificities above 90% and lowered sensitivities for detecting advanced adenomas ranging from 21% to 63% [21]. A more detailed analysis by type of adenoma revealed that the likelihood of detecting small adenomas (< 1 cm) was two-fold lower compared to adenomas larger than 1 cm. In addition, pedunculated adenomas were more likely to be detected than sessile adenomas. Apart from the morphology, also the localization influenced the diagnostic potential with proximal adenomas being less likely detected than distal adenomas [22]. Different approaches were addressed trying to identify a screening test that is effective, inexpensive and succeeds in a high compliance in patients. They are mostly based on the identification of biomarkers in either blood or fecal samples and try to measure aberrant genetics, epigenetics, protein expression, or bacterial composition [21, 23, 24]. None of the approaches have been shown to outcompete blood in stool tests in test-performance but also cost-effectiveness so far. However, it could be attempted to combine different blood- or fecal-based tests to increase sensitivity in detection of colorectal neoplasm, especially at the early stage.

1.2. Gastrointestinal Bacteria in Health and Disease

The gut microbiome has raised increasing interest in the recent years with regard to hosts health and disease. The human colon harbors an enormous number of microorganisms including up to 10^{12} bacterial cells per 1 g of feces. The most predominant phyla in the colon are Firmicutes and Bacteroidetes. Differences in the ratio between both phyla but also in abundance of more specific bacterial species have been related to disease. Obesity-related diseases, inflammatory bowel diseases (IBD) and CRC are thereby of special interest [25].

The gut is colonized with bacteria shortly after birth and the resulting microbiome is in an important symbiosis with the host [26]. The microbiome contributes to the host by digestion

of food and consequently liberation of essential nutrients such as the degradation of complex polysaccharides into short chain fatty acids [27]. It also protects the host from pathogenic bacteria by outcompeting them. The intestinal barrier, given by a mucous layer, a tight epithelium and a sophisticated immune system, retains the microbiome in the intestinal lumen, however, also enables tolerance of the commensal bacteria by the host immune system [28, 29]. Disturbances in this equilibrium may correlate with disease, e.g. IBD, obesity-related disease or CRC [25]. These disturbances might result from host factors, including genetics or hormonal status, as well as from the environment, including diet, drug use (antibiotics, NSAIDs) or lifestyle [25].

Several metagenomic attempts have been made to identify microbial compositions that are associated with CRC [29]. For example, a study by Zeller et al. found especially high abundance of Fusobacterial species associated with CRC [30]. Another study by Flemer et al. tried to generate CRC-associated microbiota profiles [31]. These metagenomic studies are overall very comprehensive since they regard the complete microbiome, however, consequently they are also highly complex and cost-intensive. This thesis focuses on two specific bacterial species that have been brought into context with CRC by different means that will be described below. So far it is unclear if and how bacteria might influence tumorigenesis. The alpha-bug theory hypothesizes that certain pro-oncogenic bacteria, e.g. *Bacteroides fragilis*, are able to induce changes in the microbiome, in the immune response of the host and finally in the colonic epithelium resulting in CRC [32]. A second theory, the so-called bacterial driver-passenger model (Figure 2), involves bacterial drivers that initiate tumor formation by DNA damage followed by passenger bacteria that benefit from the changes in the epithelium and are enabled to invade the tissue and potentially promote tumorigenesis [33]. However, so far it is unclear whether bacteria can be causally related to CRC or whether presence of an altered composition of the microbiome simply is a consequence of changes in the epithelial tissue.

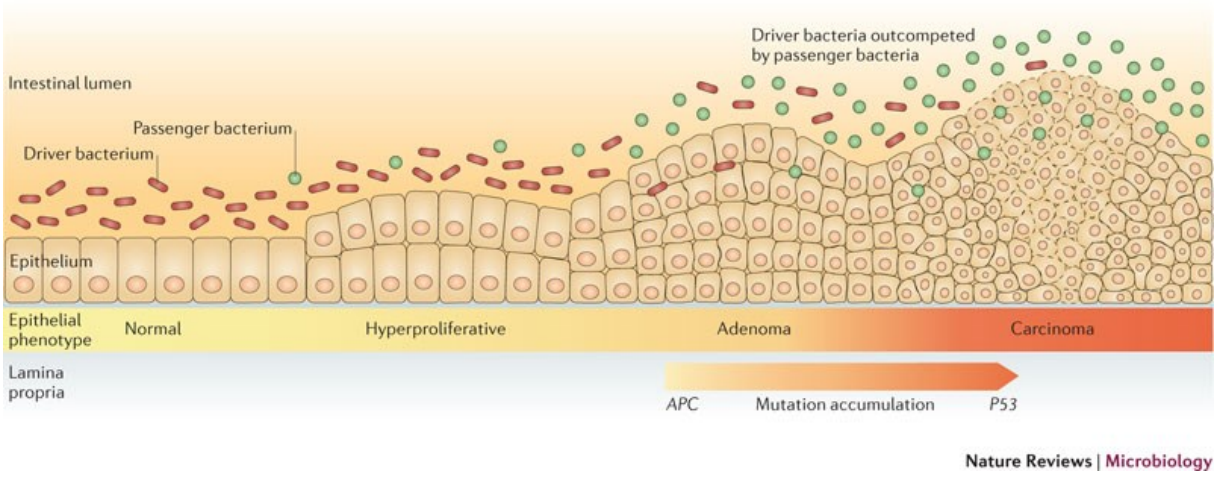


Figure 2: Bacterial driver-passenger model [33]. Driver bacteria initiate tumor formation and passenger bacteria later colonize the tumor tissue.

1.2.1. *Fusobacterium nucleatum* and CRC

Fusobacterium nucleatum (*F. nucleatum*) is an anaerobic gram-negative bacterium that belongs to the phylum Fusobacteria and the family of Fusobacteriaceae. *F. nucleatum* can be further subdivided into four subspecies (subsp.): *F. nucleatum* subsp. *nucleatum*, subsp. *polymorphum*, subsp. *vincentii* and subsp. *animalis* [34]. For simplicity I will restrict the designation to *F. nucleatum*. It is a predominant species in oral plaques and was found in about 80% of periodontal samples in a study by Moore et al., and supposed to be involved in the initiation of periodontitis [35]. It thereby not only acts pro-inflammatory with the ability to invade host cells but also functions as a bridging bacterium binding other potentially pathogenic bacterium to the oral cells [36]. Besides its presence in the oral cavity *F. nucleatum* was identified in several other entities including lung and urinary tract infections, sinusitis and abscesses in the brain, liver and skin but also in context with CRC [36]. Zeller et al., for example, found *F. nucleatum* among those bacteria with higher abundance in fecal samples of CRC patients compared to that of healthy controls [30]. Other studies focused more specifically on the detection of *F. nucleatum* instead of the composition of the whole microbiome (Table 1). Application of mostly 16S rDNA or *nusG* based PCR in fecal or rectal mucosal samples identified a higher abundance of *F. nucleatum* DNA in adenoma and CRC cases compared to controls [37-42].

Table 1: Case-control studies on the association of *F. nucleatum* with CRC

First Author	Year	Country	Sample size	Method	Result
McCoy [39]	2013	USA	Controls[1] (n=67) Adenoma (n=48)	16S rDNA qPCR rectal mucosal samples	*F. nucleatum* DNA more abundant in rectal mucosal samples of adenoma cases compared to controls
Mira-Pascual [40]	2015	Spain	Controls[1] (n=10) Adenoma (n=11) CRC (n=7)	16S rDNA qPCR Fecal samples	*F. nucleatum* DNA more abundant in fecal samples of CRC cases compared to controls
Fukugaiti [37]	2015	Brasil	Controls[1] (n=10) CRC (n=7)	16S rDNA qPCR Fecal samples	*F. nucleatum* DNA more abundant in fecal samples of CRC cases compared to controls
Yu [42]	2015	China	Controls[2] (n=109) CRC (n=47)	Butyryl-CoA dehydrogenase qPCR Fecal samples	*F. nucleatum* DNA more abundant in fecal samples of CRC cases compared to controls
Wong [41]	2016	China	Controls[1] (n=102) Adenoma (n=103) CRC (n=104)	nusG qPCR Fecal samples	*F. nucleatum* DNA more abundant in fecal samples of adenoma and CRC cases compared to controls
Liang [38]	2016	China	Controls[2] (n=236) CRC (n=203)	nusG qPCR Fecal samples	*F. nucleatum* DNA more abundant in fecal samples of CRC cases compared to controls

[1]colonoscopy-negative; [2]Healthy controls

Distinct studies compared the abundance of *F. nucleatum* DNA directly in tumor tissue and adjacent normal tissue of the same patient (Table 2) [43-46]. They all found significantly higher levels of *F. nucleatum* DNA specifically in the tumor tissue. These findings led to further studies assessing a potential causal role of *F. nucleatum* in CRC development. Studies by Flanagan et al., Mima et al. and Wei et al. found that a higher amount of *F. nucleatum* DNA in tumor tissue was associated with worse clinical outcome and larger tumor size [44, 47, 48]. Moreover they found that higher amount of *F. nucleatum* DNA associates with MSI-H tumors. These findings were confirmed by Nosho et al. and Ito et al. in independent studies and extended by the finding that high *F. nucleatum* DNA levels are significantly less frequent in premalignant lesions compared to CRC and more frequent in tumors expressing the *BRAF* mutant and those of the serrated pathway [49, 50].

Table 2: Studies comparing *F. nucleatum* DNA abundance in CRC tumor and normal adjacent tissue.

First Author	Year	Country	Sample size	Method	Result
Kostic [45]	2012	Spain	CRC (n=95)	16S rDNA qPCR	*F. nucleatum* DNA more abundant in tumor than adjacent normal tissue
Castellarin [43]	2013	USA	CRC (n=99)	nusG qPCR	*F. nucleatum* DNA more abundant in tumor than adjacent normal tissue
Flanagan [44]	2014	Czech Republic Germany Ireland	CRC (n=49) CRC (n=45) CRC (n=28) Adenoma (n=52)	nusG qPCR	*F. nucleatum* DNA more abundant in tumor than adjacent normal tissue in CRC cases
Li [46]	2016	China	CRC (n=101)	16S rDNA qPCR	*F. nucleatum* DNA more abundant in tumor than adjacent normal tissue

More mechanistic studies addressed a potential mode of action of the bacterium in tumor progression and found that *F. nucleatum* might interact with and inhibit the adaptive immune system in the tumor. High abundance of *F. nucleatum* DNA was found to be inversely associated with presence of CD3+ T-cells [51]. In vitro studies showed that outer membrane proteins Fap2 (systematic name: Fn1449) and RadD (Fn1526) induced cell death in human lymphocyte cell lines [52] and that Fap2 binds to inhibitory receptors of Natural killer cells [53]. In the contrary, a study in mice showed that tumors confronted with *F. nucleatum* infection exhibited a pro-inflammatory microenvironment [54]. Wei et al. found an association of high abundance of *F. nucleatum* with NFκB (nuclear factor κB) expression [48]. In addition it was found that infection of CRC cell lines with *F. nucleatum* enhanced their proliferation by the induction of NFκB [55]. These findings indicate an inhibitory effect on the adaptive immune system and a pro-inflammatory and pro-proliferative effect on the tissue. However, these studies should be further confirmed to gain a comprehensive evidence for an immunity modulating influence of *F. nucleatum*. Apart from a potential influence on the immune system an adhesion protein FadA (Fn0264) was identified [56] to be important in invasion of the bacterium but also promotion of tumorigenesis by binding to E-cadherin and activation of β-catenin signaling [57]. RadD and Fap2 belong to a type V secretion system with potential virulence properties [58]: Fap2 was shown to adhere to D-galactose-β(1-3)-N-

acetyl-D-galactosamine, which is highly expressed by tumor cells [59] and RadD was shown to be important for biofilm formation [60]. Altogether recent research found several indications for a potential causal involvement of *F. nucleatum* in CRC development and progression, however, clear evidence is not given so far.

1.2.2. *Streptococcus gallolyticus* subsp. *gallolyticus* and CRC

Streptococcus gallolyticus subsp. *gallolyticus* (*S. gallolyticus*) is a gram-positive bacterium that belongs to the *S. bovis*/*S. equinus* complex and the family of Streptococcaceae in the phylum Firmicutes. The S. *bovis/S. equinus* complex is a diverse group of human and animal commensals but also pathogens and comprises besides *S. gallolyticus* also *S. gallolyticus* subsp. *pasteurianus* (*S. pasteurianus), S. gallolyticus* subsp. *macedonicus* (*S. macedonicus*), *S. infantarius* subsp. *infantarius* (*S. infantarius*), *S. lutetiensis, S. alactolyticus* and strains originally and not further subdifferentiated as *S. equinus* and *S. bovis*. As the name of the complex indicates, these bacteria, also *S. gallolyticus*, are frequently identified from sources of animal origin, including ruminants, especially cattle, chickens, pigeons and pigs indicating a potential for zoonotic transmission [61-65]. The taxonomy and nomenclature shown above was proposed by Schlegel et al. in 2003 [66]. Prior to that the classification was based on the ability of the bacteria to ferment mannitol and subdivided the *S. bovis* species into biotype I (now *S. gallolyticus*), biotype II/1 (now *S. infantarius* and *S. lutetiensis*), biotype II/2 (now *S. pasteurianus*) as well as *S. macedonicus* and *S. equinus* [67]. The nomenclature is still under debate but for consistency I will use the nomenclature proposed by Schlegel et al. [66].

Early before species of the complex were further subclassified, studies linked an infective endocarditis inducing bacterium named *S. bovis* to CRC [68-71]. Ruoff et al. in 1989 were the first to describe that bacteremia induced specifically by *S. gallolyticus* was highly correlated with colorectal neoplasm [72]. Table 3 summarizes studies that compare the presence of adenoma and/or CRC in patients with bacteremia induced by *S. gallolyticus* compared to those induced by other bacteria.

Table 3: Cohorts of bacteremia patients and association with CRC.

First Author	Year	Country	Bacteremia induced by	Method	Result
Corredoira [73]	2005	Spain	*S. gallolyticus* (n=42) *S. pasteurianus/ S. infantarius* (n=20) *S. salivarius* (n=17)	Microbial typing	Presence of adenoma/CRC significantly more frequent in patients *with S. gallolyticus* bacteremia (57% vs. 15% and 0%, respectively)
Corredoira [74]	2012	Spain	*S. gallolyticus* (n=98) No bacteremia (n=196)	Microbial typing	Presence of advanced adenoma/CRC significantly more frequent in patients with *S. gallolyticus* bacteremia (70% vs. 32%, respectively)
Corredoira [75]	2014	Spain	*S. gallolyticus* (n=99) *S. pasteurianus/ S. infantarius* (n=36)	Microbial typing	Presence of adenoma/CRC significantly more frequent in patients with *S. gallolyticus* bacteremia (70% vs 22%, respectively)
Sheng [76]	2014	Taiwan	*S. gallolyticus* (n=31) *S. pasteurianus* (n=126) *S. infantarius* (n=15)	Microbial typing	Presence of adenoma/CRC was not significantly more frequent in patients with *S. gallolyticus* bacteremia (16% vs 15% and 0%, respectively)
Corredoira [77]	2015	Spain	*S. gallolyticus* (n=224) *S. pasteurianus/ S. infantarius* (n=270)	Microbial typing	Presence of adenoma/CRC significantly more frequent in patients with *S. gallolyticus* bacteremia (51% vs 16%, respectively)

[1] colonsocopy-negative

Four of the five studies were conducted by Corredoira et al. in Spain and showed a significantly higher fraction of adenoma/CRC cases among *S. gallolyticus* bacteremia patients (51-70%) than among patients with bacteremia induced by other bacteria (0-32%) [73-75, 77]. The fifth study by Sheng et al. from Taiwan, however, did not see this difference in adenoma/CRC presence between patients with bacteremia induced by *S. gallolyticus*, *S. pasteurianus* and *S. infantarius* [76].

The association of *S. gallolyticus* presence in fecal or tumor tissue with CRC was only rarely described in the current literature, as compared to the *F. nucleatum* studies. Fecal carriage rates were assessed by PCR or bacterial isolation of *S. gallolyticus* DNA and overall fecal carriage rates in healthy individuals were low (1-11%) [78, 79]. Chirouze et al. could not

identify a significant difference in the fecal carriage rate between controls and CRC cases [78]. Abdulamir et al. isolated species of the so-called *S. gallolyticus* member bacteria (SGMB), not differentiating between the different subspecies *gallolyticus*, *pasteurianus* and *macedonicus*, from fecal, mucosal and tissue samples of CRC patients and controls [80]. There was no difference between CRC cases and controls in the number of isolates from fecal and mucosal samples but in the number of isolates from tissue samples. Other studies addressed the association of *S. gallolyticus* with CRC and premalignant lesions with serological methods (Table 4). Independent of the detection method they all found significantly more *S. gallolyticus* positive CRC cases than controls. A study performed in our laboratory applied multiplex serology (1.4) using four *S. gallolyticus* pilus proteins [81], two of them, Gallo2178 and Gallo2179, being further elucidated in this thesis. All of these studies were in a retrospective case-control design and although Abdulamir et al. [82] and Garza-Gonzalez et al. [83] regarded also polyps in addition to CRC cases there was no prospective study conducted so far addressing whether *S. gallolyticus* infection is present in the same individual prior to CRC diagnosis. This could be on the one hand of diagnostic value and on the other hand indicative for a causal relationship.

Table 4: Case-control studies on the association of *S. gallolyticus* with CRC

First Author	Year	Country	Sample size	Method	Result
Abdulamir [82]	2009	Malaysia	Controls[1,2] (n=60) CRC (n=50) Adenoma (n=14)	*S. gallyticus* cell wall protein ELISA	Significantly more *S. gallolyticus* positive adenoma/CRC cases
Boleij [84]	2010	Netherlands USA	Controls[1] (n=127) Early stage CRC (n=48) Advanced CRC (n=34) Controls[1] (n=48) Early stage CRC (n=35) Advanced CRC (n=22)	*S. gallolyticus* ribosomal protein RpL7/L12 ELISA	Higher antibody titers among early stage CRC cases compared to controls
Garza-Gonzalez [83]	2012	USA	Controls[2] (n=54) Polyps (n=133)	*S. gallolyticus* whole cell protein Western blot	Significantly more *S. gallolyticus* positive polyp cases
Butt [81]	2016	Spain	Controls[1] (n=576) CRC (n=576)	Multiplex serology with *S. gallolyticus* pilus proteins	Significantly more *S. gallolyticus* positive CRC cases

[1]Healthy controls; [2]colonoscopy-negative

The bacterial driver-passenger model developed by Tjalsma et al. [33] (1.2) describes *S. gallolyticus* as a potential passenger bacterium: a commensal that invades tissue and blood stream to finally cause a bacteremia or endocarditis specifically through the presence of CRC or a precursor. Several studies found that *S. gallolyticus* is able to bind to collagen I rich surfaces [85, 86], a collagen type that is expressed in heart valves as well as CRC tissue [87, 88] indicating a niche for *S. gallolyticus* adherence to CRC tissue in contrast to normal tissue. Pilus-structures, appendages of gram-positive bacteria, especially those composed of proteins Gallo2178 and Gallo2179, are assumed to be mediators of this adherence [89, 90]. Furthermore *S. gallolyticus* was shown to adhere to collagen type IV, fibrinogen and mucus [85, 86, 91-93]. Boleij et al. even showed that *S. gallolyticus* has a growth advantage in spent medium of malignant colonocytes [94]. Additionally, cytokine expression profiles in human CRC tissue positive for *S. gallolyticus* DNA indicate a pro-inflammatory and thus pro-carcinogenic potential [80, 82]. However, further studies are needed to confirm a role of *S. gallolyticus* in CRC carcinogenesis going beyond the possibility of a simple "passenger" in this process.

1.3. Seroepidemiology of Infection-Associated Cancers

1.3.1. Epidemiological study designs and measures of association

Epidemiology is "the study of the occurrence and distribution of health-related events, states, and processes in specified populations, including the study of determinants influencing such processes, and the application of this study to the control of health problems." [95]. With respect to this thesis this implies the study of an exposure to bacterial infections as potential determinant for the event/outcome CRC. The application of different epidemiological study designs thereby allows assessing whether exposures are related to outcomes. A relatively convenient study design in terms of cost- and time-effectiveness is given by retrospective case-control studies: individuals are chosen based on the outcome (and cases and controls are often matched by important confounders such as age and sex), and then the prevalence of the exposure is assessed retrospectively (Figure 3A). Consequently, the odds for the outcome at presence of exposure can be compared to the odds for the outcome in the absence of the exposure resulting in an odds ratio (OR) as a measure of association for the relationship between outcome and exposure (Figure 3A) [96].

Prospective cohorts, in contrast, provide incidences (number of newly identified cases per person-time at risk) for the outcome. In this study design the presence of the exposure in a

cohort of individuals is determined at baseline and then the individuals are followed-up over time for the occurrence of the outcome (Figure 3B). Incidences for the outcome among those individuals with the exposure at baseline are compared to those without the exposure to estimate the relative risk (RR) for the outcome in the exposed in relation to the unexposed (Figure 3B) [97].

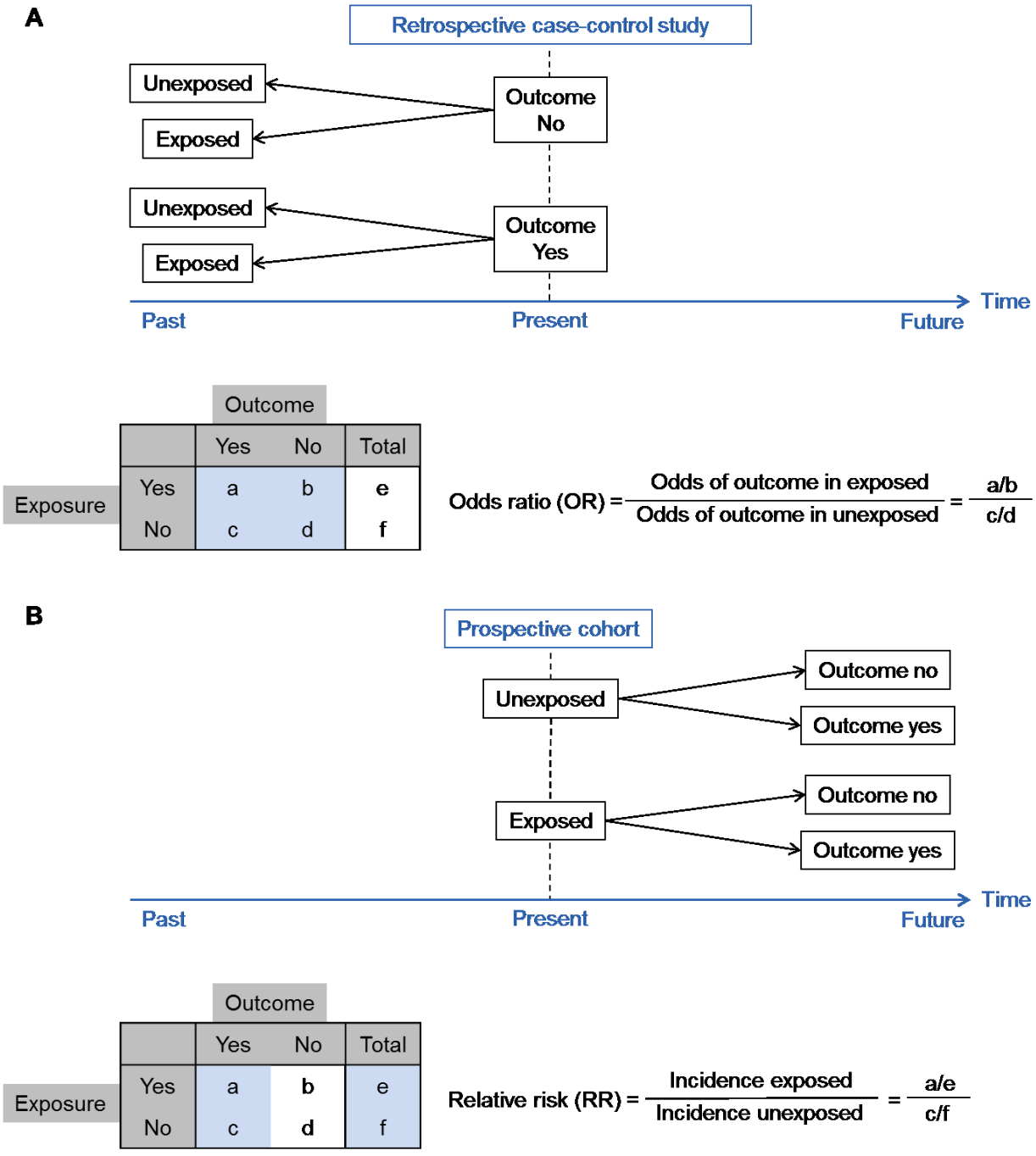


Figure 3: Principle and measures of association of different epidemiological study designs. A) retrospective case-control studies, B) prospective cohorts.

Prospective cohorts imply a temporality of the association since exposure happens prior to outcome detection. Temporality is one out of the nine Bradford Hill criteria and a first indicator for causation [98]. In contrast, retrospective case-control studies can only show an association between exposure and outcome. Prospective studies, however, are also cost- and time-intensive especially for rare outcomes and exposures, since study groups have to be large enough to guarantee statistical power, analyses of resulting large numbers of biological specimens are costly, and often the follow-up time lasts several years for the outcome to occur. An often applied design is therefore a case-control study nested within a prospective cohort. Frequency of exposure is assessed in either all or a selection of individuals with incident outcome under comparison to a respective group of controls. This subset analyses offers the advantage of a prospective setting, and consequently temporality of the association as well as data analyses in a minimal number of samples. However, also here the strength of the association is estimated as OR and not as relative risk [99].

In both study designs, retrospective case-control and case-control studies nested within a prospective cohort, the strength of an association between categorical exposure and outcome is given by the OR. In a simple univariate approach the OR can be calculated from a 2 by 2 table as depicted in Figure 3B. However, in the majority of cases exposure and outcome do not depend exclusively on each other. Other factors, so-called confounders, might influence the relation and have to be considered when estimating the strength of an association. Logistic regression models enable such multivariate modelling of the association [100].

The significance of the estimated OR is assessed by the 95% confidence interval (95% CI). This interval gives the upper and lower limits among which the true parameter (OR) will range in an unlimited repetition of the analysis at least as frequent as given by the confidence level (95%) [101].

As mentioned above other factors, confounders, might influence the association between exposure and outcome and it should be adjusted for confounding variables to minimize bias. Confounders are associated with both, the outcome and exposure of interest, and cannot be an intermediary step in the causal pathway [102]. Bias through confounding can be minimized already during study design: the selection of controls can be matched to cases by important and potentially confounding variables [103]. In contrast, if there is indication for effect modification, i.e. a biological assumption why one group of individuals should differ to another in the strength and/or direction of the association, the association should be estimated

stratified by the two groups of individuals [104]. In case of CRC potential associations are often assessed separately for males and females since males are more likely to develop CRC than females and may also differ in their association with risk factors for the outcome [105].

1.3.2. Infection-associated cancers and serology

Valid ascertainment of the exposure, in this case a bacterial infection, is crucial for assessing a potential association with the outcome. Acute bacterial infections can be diagnosed directly by bacterial culture, microscopic analysis or based on detection of species-specific DNA by PCR or sequencing [106]. However, these methodologies, especially the gold standard bacterial culture, are time-consuming and costly and do not provide the possibility to measure also past infections. A different, easy-to-apply method is serology, the measurement of antibodies against antigenic structures. The detected antibody responses serve as biomarkers for the indirect detection of acute and past infections [106]. Seroepidemiological studies provide the possibility to identify associations of this biomarker with the outcome. As described above serum collections of retrospective case-control studies are analyzed to assess associations, however, cannot be used to analyze causal inference. Serum collections of case-control studies nested within prospective cohorts, in contrast, provide the possibility to assess the temporality of an association. This knowledge may be the basis for further molecular analyses of a potential causal relationship of the infection with the outcome. Independent of a potential causality the identified antibody markers might be applicable in early diagnostics of the outcome.

Coherently, serological associations of infectious agents with cancer have been frequently identified and appeared in different ways:

i) Infection was more frequent in tumor patients leading to a higher seroprevalence, i.e. Hepatitis C infection and liver cancer [107]

ii) The infectious load was higher in tumor patients leading to higher antibody titers and perhaps also higher seroprevalence, i.e. *Helicobacter pylori* (*H. pylori*) infection and gastric cancer [108] or Epstein-Barr-virus infection and nasopharyngeal cancer [109]

iii) Specific antigens were mainly expressed in tumor development or in the established tumor and antibodies to them might serve as tumor markers, i.e. antibodies to human papillomavirus 16 E6 and E7 proteins in cervical [110] and oropharyngeal cancer [111], Merkel cell polyomavirus T-antigen in Merkel cell

carcinoma [112] or to *H. pylori* Cytotoxin-associated antigen A (CagA) and Vacuolating toxin A (VacA) [113].

1.4. Multiplex Serology

Conventional serology often uses the enzyme-linked immunosorbent assay (ELISA) or Western blot to detect antibody responses to antigens of infectious agents. Whole cell lysates, or purified endogenously or recombinantly expressed antigens are immobilized on microtiter plates or blotted on a membrane, respectively. These methods are limited to one antigen or one antigen pool analyzed per well/lane and analysis of large seroepidemiological studies is time- and material-consuming. A technique called multiplex serology, developed in our laboratory by Waterboer et al. [114] allows the analysis of approximately up to 2000 sera per day for up to 100 antigens. Polystyrene beads filled with two fluorescent dyes in various ratios provide an array of 100 different bead sets, each with its internal specific color. Mixing of the differently loaded bead sets results in a suspension array. The technique developed by Waterboer et al. [114] uses recombinantly expressed proteins (X) as antigens flanked by an N-terminal Glutathione-S-transferase (GST) and a C-terminal peptide including the seven C-terminal amino acids of SV40 large T antigen (tag). Glutathione-casein is crosslinked to the beads and binds to GST, which allows for in-situ affinity purification of the GST-X-tag fusion proteins. Antibodies in serum samples binding to the antigens can be detected by a biotin-labeled secondary antibody and streptavidin-R-phycoerythrin (strep-PE) as fluorescent reporter dye (Figure 4). A Luminex xMAP device, which is comparable to a flow cytometer, has two lasers: the red laser excites the internal fluorescence of the bead set and consequently identifies the loaded antigen and the green laser excites the reporter fluorescence (strep-PE), which is then quantified. The antibody reactivity is given as median fluorescent intensity (MFI) -value of at least 100 beads per set. Altogether, multiplex serology allows for quantification of antibody responses in large sets of sera against several pathogens simultaneously.

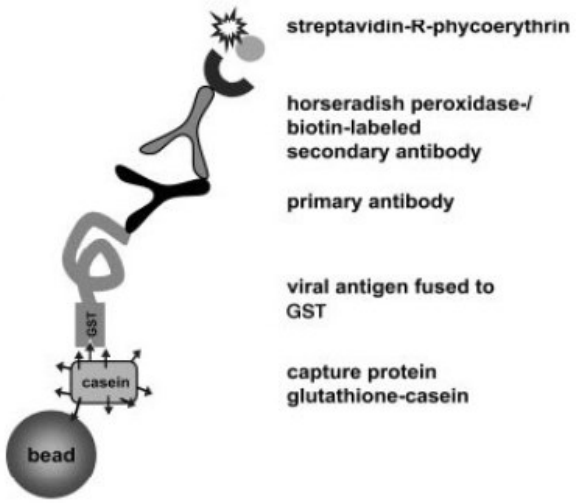


Figure 4: Principle of multiplex serology. (modified from [114])

1.5. Objectives

The aim of this thesis was to elucidate the potential serological association of *F. nucleatum* and *S. gallolyticus* with CRC. More specifically it was aimed at first developing a fluorescent bead-based multiplex serology assay for the detection of antibody responses against *F. nucleatum* and *S. gallolyticus* proteins. Second, using this assay it was aimed to assess whether published associations of *F. nucleatum* and S. *gallolyticus* with prevalent CRC can be reproduced on a serological basis in a retrospective CRC case-control study and whether specific bacterial antibody marker for CRC can be identified. At last, it was to be addressed whether antibody responses to the bacteria can be detected specifically prior to CRC diagnosis in two independent case-control studies nested within prospective cohorts. The prospective analyses thereby might give further insight into the temporality of the association and whether specific detection of such antibody markers prior to diagnosis might be of diagnostic potential in the identification of individuals at increased risk of (pre-) cancerous colorectal lesions.

2. Materials and Methods

2.1. Materials

2.1.1. Chemicals

Acrylamide/Bisacrylamide-solutions	Carl Roth (Karlsruhe)
Agarose	Sigma-Aldrich (Steinheim)
Ammoniumperoxodisulfate (APS)	Carl Roth (Karlsruhe)
Ampicillin	Roche (Mannheim)
β-Mercaptoethanol	Merck (Darmstadt)
Bacto agar	DIFCO Becton Dickinson (Sparks, MD, USA)
Bacto tryptone	DIFCO Becton Dickinson (Sparks, MD, USA)
Bacto yeast extract	GIBCO, Invitrogen (Karlsruhe)
Bradford reagent (Roti-Quant)	Carl Roth (Karlsruhe)
Bromphenol blue	Merck (Darmstadt)
Casein	Sigma-Aldrich (Steinheim)
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich (Schnelldorf)
DNA marker (Smart ladder)	Eurogentec (Seraing, Belgium)
1,4-Dithiothreitol (DTT)	Carl Roth (Karlsruhe)
Ethanol	Riedel-de Häen (Seelze)
Ethylendiamintetraacetat (EDTA)	GIBCO, Invitrogen (Karlsruhe)
Glutathione	Sigma-Aldrich (Taufkirchen)
Glycerol (100%)	Carl Roth (Karlsruhe)
Glycine	Gerbu (Gaiberg)
H_2O, DNase/RNase-free	GIBCO, Invitrogen (Karlsruhe)
Hydrochloric acid (HCl)	Riedel-de-Häen (Seelze)
Isopropanol	J.T. Baker (Deventer, Niederlande)
Isopropyl-β-D-thiogalactosid (IPTG)	Carl Roth (Karlsruhe)
Methanol	DKFZ (Heidelberg)
Milk powder	Carl Roth (Karlsruhe)
N,N,N',N'-Tetramethylethylendiamin (TEMED)	Merck (Darmstadt)
peqGreen	peqLab, VWR (Erlangen)
Polyvinylalcohol (PVA)	Sigma-Aldrich (Steinheim)
Polyvinylpyrrolidon (PVP)	Sigma-Aldrich (Steinheim)

Potassiumdihydrogenphosphate (KH_2PO_4)	Merck (Darmstadt)
Prestained protein ladder (Broad range)	Biolabs (Munich)
Protease Inhibitor Complete (1 tablet/ml)	Roche (Mannheim)
Smart ladder	Eurogentec (Cologne)
Sodium-acetate (NaAc)	Thomas Chemikalien (Heidelberg)
Sodium-azide (NaN_3)	Merck (Darmstadt)
Sodium-carbonate (Na_2CO_3)	Carl Roth (Karlsruhe)
Sodium-chloride (NaCl)	Sigma-Aldrich (Steinheim)
Sodium-dodecyl-sulfate (SDS)	Gerbu (Gaiberg)
di-Sodiumhydrogenphosphate (Na_2HPO_4)	Merck (Darmstadt)
Streptavidin-R-Phycoerythrin (strep-PE)	Moss Inc. (Pasadena, Maryland, USA)
Sucrose	Merck (Darmstadt)
Sulfuric acid (H_2SO_4), 95-97 %	AppliChem (Darmstadt)
Superchemiblock (CBS-K)	Chemicon (Temecula, CA, USA)
Tetramethylbencidine (TMB)	Sigma-Aldrich (Taufkirchen)
Tris(hydroxymethyl)-aminoethan (Tris)	Sigma-Aldrich (Steinheim)
Tween®-20	Gerbu (Gaiberg)
xMAP^{TM} Sheath fluid	Luminex Corp. (Austin, Tx, USA)

2.1.2. Buffers

Agarose gel electrophoresis buffer, 50x	2 M Tris, pH 7.8, 0.25 M NaAc water free, 0.05 M EDTA
DNA sample buffer, 6x	0.25% (w/v) Bromphenol blue, 40% (w/v) sucrose in H_2O
ELISA blocking buffer	0.2% (w/v) casein in PBS-T
ELISA coating buffer	2 ng/µl glutathione-casein in 50 mM carbonate buffer (50 mM Na_2CO_3, 50 mM $NaHCO_3$ 1:4, pH 9.6)
ELISA stop solution	1 M H_2SO_4
ELISA substrate solution	100 µg/ml TMB in 100 mM NaAc, pH 6.0, 0.015% H_2O_2
EMBL transfer buffer	48 mM Tris, 39 mM glycine, 0.0345% SDS, 20% methanol
LB medium	1% (w/v)) Bacto tryptone, 0.5% (w/v) Bacto yeast extract, 1% (w/v) NaCl; pH 7.5
LB_{amp} medium	LB medium, 10 µl/ml ampicillin

LB_{amp} agar	LB_{amp} medium, 1.5% (w/v) Bacto agar
Phosphate-buffered saline (PBS), 10x	124 mM NaCl, 22 mM Na_2HPO_4, 10 mM KH_2PO_4, pH 7.4
PBS-T	0.05% (v/v) Tween 20 in 1x PBS
SDS sample buffer, 4x	160 mM Tris, pH 6.8; 10% (v/v) glycerol; 2% (w/v) SDS; 5% (v/v) β-mercaptoethanol; 0.25% (w/v) bromphenol blue
SDS-PAGE running buffer, 10x	250 mM Tris, 14.4% (w/v) Glycine, 1% (w/v) SDS
Multiplex serology blocking buffer	1 mg/ml Casein in PBS
Serum pre-incubation buffer	2 mg/ml GST-tag lysate, 0.5% (w/v) polyvinyl alcohol, 0.8% (w/v) Polyvinyl pyrrolidone, 2.5% (v/v) CBS-K super chemiblock in multiplex serology blocking buffer
Storage buffer	0.05% (w/v) NaN_3 in multiplex serology blocking buffer
Western blot blocking buffer	10% milk in PBS-T

2.1.3. Consumables

Beside the general consumables, i.e. tips, gloves, tubes, the following specific consumables were applied:

96-well microtiter plates	Nunc (Wiesbaden)
96-well polystyrene flat-bottom plates	Greiner bio-one (Frickenhausen)
96-well filter plates	Millipore (Bredford, MA, USA)
ECL™ Western blotting Detection reagents	GE Healthcare (Freiburg)
Electroporation cuvettes	10 mm, Invitrogen (Karlsruhe)
Nitrocellulose membrane PROTRAN	Schleicher & Schuell (Dassel)
QiaPrep Spin MidiPrep kit	Qiagen (Hilden)
QiaPrep Spin MiniPrep kit	Qiagen (Hilden)
Qiagen Multiplex PCR kit	Qiagen (Hilden)
QiaQuick PCR purification kit	Qiagen (Hilden)
SeroMAP™ Microspheres (Fluorescent polysterene beads)	Luminex Corp. (Austin, TX, USA)
Whatman 3MM paper	Schleicher & Schuell (Dassel)
X-ray films X-Omat™ Blue XB-1	Kodak (Rochester, NY, USA)

2.1.4. Laboratory equipment

Beside the general laboratory equipment, i.e. shaker, vortex, heater and water baths, the following specific tools were used:

Agarose gel electrophorese system	Renner GmbH (Darmstadt)
Benchtop centrifuge (5415D)	Eppendorf (Hamburg)
Centrifuge RC-5C, Sorvall	Thermo Scientific (DuPont, DE, USA)
Developer Curix 60	Agfa (Cologne)
Gel Doc EZ Imager	BioRad (Munich)
Gene pulser	BioRad (Munich)
Luminex 100 analyzer	Luminex Corp. (Austin, TX, USA)
Luminex 200 analyzer	Luminex Corp. (Austin, TX, USA)
Luminex SD sheath fluid delivery system	Luminex Corp. (Austin, TX, USA)
Luminex XYP plate handler	Luminex Corp. (Austin, TX, USA)
Mini Trans-Blot Electrophoretic Transfer Cell	BioRad (Munich)
Multiskan PLUS MKII	Titertek (Pforzheim)
NanoDrop ND-1000	Thermo Scientific (DuPont, DE, USA)
PCR cycler (Eppendorf Mastercycler)	Eppendorf (Hamburg)
Power Pac 300	BioRad (Munich)
Pressure homogenizer EmulsiFlex-C5	Avestin (Mannheim)
Pulse controller	BioRad (Munich)
Rotor ThermoScientific F12-6x500 LEX	Thermo Scientific (DuPont, DE, USA)
Rotor Sorvall SA-600	Thermo Scientific (DuPont, DE, USA)
SDS-PAGE electrophoresis chamber (Mini-PROTEAN II)	BioRad (Munich)
Ultrasonic bath	Bandelin Sonorex (Berlin)
Underbench centrifuge (Varifuge RF)	Heraeus (Hanau)
Vacuum manifold	Millipore (Bredford, MA, USA)
Vacuum pump (Millivac ®)	Millipore (Bredford, MA, USA)

2.1.5. Antibodies

Biotinylated goat anti-human IgA, IgM, IgG	Dianova (Hamburg)
Biotinylated goat anti-mouse IgG	Dianova (Hamburg)
Biotinylated goat anti-bovine IgA, IgM, IgG	Dianova (Hamburg)

Biotinylated mouse anti-tag from KT3 hybridoma cell supernatant	[115], purified and biotinylated by Tim Waterboer
Goat anti-mouse IgG horseradish peroxidase (HRP) conjugate	Dianova (Hamburg)
Goat anti-mouse IgG (HRP) conjugate	Dianova (Hamburg)
Mouse anti-tag from KT3 hybridoma cell supernatant	[115], purified by Tim Waterboer
Rabbit anti-GST	Sigma-Aldrich

2.1.6. Bacterial strains

Escherichia coli (*E. coli*) BL21 wildtype bacteria were purchased from GE Healthcare (Freiburg).

2.1.7. Enzymes and reaction buffers

All restriction enzymes and respective reaction buffers were purchased from NEB (Frankfurt).

2.1.8. Antigens

The bacterial lysates containing recombinantly expressed GST-tag (without insert) and Polyomavirus JC VP1 as GST-X-tag fusion protein were prepared in the laboratory of Michael Pawlita (DKFZ, Heidelberg) by Ute Koch.

2.1.9. Sera and respective study data

S. gallolyticus reference sera from mice were kindly provided by Annemarie Boleij (Radboud University Medical Centre, Nijmegen, Netherlands). Mice had been inoculated with $4.5*10^6$ *S. gallolyticus* UCN34 cells (positive control) or PBS (negative control). Serum samples had been taken weekly for up to 7 weeks. Data on colonization status of the mice with *S. gallolyticus* were also kindly provided by Annemarie Boleij. *S. gallolyticus* DNA status in colon tissue and fecal samples was kindly provided by Indra-Jasmin Gierse (DKFZ, Heidelberg).

S. gallolyticus reference sera from cattle and respective fecal samples were collected by Indra-Jasmin Gierse in 10 farms located in Eastern Germany. *S. gallolyticus* DNA status in fecal samples was kindly provided by Indra-Jasmin Gierse.

S. gallolyticus reference sera from three individuals with diagnosed *S. gallolyticus* bacteremia and three healthy controls were kindly provided by Harold Tjalsma (Radboud Medical University Center Nijmegen, Netherlands).

Serum samples and study data of the BliTz and DACHSplus study were kindly provided by Hermann Brenner (DKFZ, Heidelberg).

Serum samples and study data of the SCCS were kindly provided by Meira Epplein (Vanderbilt University, Nashville, TN, USA).

Serum samples and study data of the EPIC study were kindly provided by Mazda Jenab (International Agency for Research on Cancer (IARC), Lyon, France).

Serum samples and study data of the Irish CRC case-control study were kindly provided by David Hughes (Royal College of Surgeons in Ireland (RCSI), Dublin, Ireland). For 52 participants colon tissue samples were analyzed for the presence of fusobacterial DNA by qPCR. Data were kindly provided by David Hughes as reference for *F. nucleatum* multiplex serology results.

2.1.10. Software and websites

BlastP	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Clustal Omega	https://www.ebi.ac.uk/Tools/msa/clustalo/
GraphPad Prism 6	GraphPad Software (La Jolla, USA)
Luminex 100 IS 2.2 SP1 Software	Luminex Corp. (Austin, TX, USA)
Microsoft Windows 7	Microsoft Corp. (Unterschleißheim)
Microsoft Office 2010	Microsoft Corp. (Unterschleißheim)
PubMed	https://www.ncbi.nlm.nih.gov/pubmed
Reverse Complement	http://www.bioinformatics.org/sms/rev_comp.html
SAS 9.4	SAS Institute Inc. (Cary, NC, USA)
SignalP 4.1	http://www.cbs.dtu.dk/services/SignalP/
TMpred	http://www.ch.embnet.org/software/TMPRED_form.html

2.2. Methods

2.2.1. Selection and cloning of proteins for *F. nucleatum* and *S. gallolyticus* multiplex serology

Extensive literature search ("PubMed") identified only few known immunogenic proteins for *S. gallolyticus* (pilus proteins Gallo1569, Gallo2178, Gallo2179 and Gallo2039) [116] and *F. nucleatum* (adhesin Fn0264 (FadA) [117] and porin Fn1859 (FomA) [118]). Both *F. nucleatum* proteins were included in the antigen selection. Based on the findings in Butt et al. [81] only Gallo2178 and Gallo2179 were included for *S. gallolyticus*.

Sequence data was available for the genomes of different strains of both, *S. gallolyticus* and *F. nucleatum*. Genomes of *S. gallolyticus* strain UCN34 [119] and *F. nucleatum* strain ATCC25586 [120] served as reference to select additional potential antigens for each bacterium (Table 5). Predictions of protein function and localization (cell wall/outer membrane or secretion) thereby served as major criterion for protein selection [119-122]. Based on these predictions I chose 18 additional proteins resulting in a total number of eleven proteins per bacterium (Table 5). Ten out of eleven *S. gallolyticus* proteins were predicted to be located at the cell wall and Gallo0933 as a putative enzyme degrading tannins was predicted to be secreted. Gallo0272, Gallo0577, Gallo1570, Gallo2178 and Gallo2179 were predicted to be involved in adhesion of the bacterium. Especially Gallo2178 and Gallo2179 were well described as parts of pilus structures that enable *S. gallolyticus* to adhere to collagen rich surfaces [89]. Gallo0112, Gallo0748 and Gallo2018 were predicted to have enzymatic function. The function of Gallo1675 was unknown. Proteins selected for *F. nucleatum* were mainly localized at the outer membrane. Two fragments of a putative hemolysin were predicted to be secreted (Fn1817_1 and Fn1817_2). Others were predicted to be important for adhesion (Fn0264), to be outer membrane proteins (Fn0253, Fn1859) or act as secretion system in the outer membrane (FN0131, Fn0387, Fn1426, Fn1449, Fn1526 and Fn1893).

Amino acid sequences for predicted signal peptides ("SignalP 4.1") or predicted transmembrane domains ("TMpred") were excluded from the final sequence to facilitate expression. If the full-length proteins were bigger than 100 kD they were either split up (e.g. Gallo0112) or only domains with a predicted function were selected (e.g. autotransporter domains in Fn0387, Fn1526, Fn0387 and Fn1893).

Table 5: Selected antigens and their characteristics for *F. nucleatum* and *S. gallolyticus* multiplex serology

Strain	Antigen	Putative function[1]	Predicted localization[1]	Protein accession no.[2]	Selected region (AA)	Protein size incl. GST-tag (kD)[4]
Fusobacterium nucleatum subsp. *nucleatum* (ATCC25586; Accession no.: NC_003454)	Fn0131	Hemolysin activator	OM	NP_603038	17-566	86
	Fn0253	Outer membrane protein A	OM	NP_603160	37-132	37
	Fn0264[3]	Adhesin (FadA)	OM	NP_603171	19-129	38
	Fn0387	Outer membrane protein, Type Va secretion system, autotransporter domain	OM/EC	NP_603291	1442-1714	56
	Fn1426	Outer membrane protein, Type Va secretion system, serine peptidase domain	OM/EC	NP_604320	25-374	65
	Fn1449	Outer membrane protein, Type Va secretion system, autotransporter domain	OM/EC	NP_604343	2884-3155	56
	Fn1526	Outer membrane protein, Type Va secretion system, autotransporter domain	OM/EC	NP_602353	1857-2135	57
	Fn1817_1	Hemolysin, filamentous haemagglutinin repeat	EC	NP_602617	205-276	34
	Fn1817_2	Hemolysin, filamentous haemagglutinin repeat	EC	NP_602617	839-909	34
	Fn1859[3]	Major outer membrane protein (FomA)	OM	NP_602659	21-368	64
	Fn1893	Outer membrane protein, Type Va secretion system, autotransporter domain	OM/EC	NP_602692	1079-1351	56
Streptococcus gallolyticus subsp. *gallolyticus* (UCN34; Accession no.: NC_013798)	Gallo0112A	Fructan hydrolase N-Terminus	CW	WP_012961337	44-816	111
	Gallo0112B	Fructan hydrolase C-Terminus	CW	WP_012961337	784-1275	80
	Gallo0272	Glucan binding protein C domain	CW	WP_012961389	500-997	81
	Gallo0577	CnaB domain	CW	WP_012961602	27-715	102
	Gallo0748	Cell-envelope proteinase A	CW	WP_012961731	36-800	110
	Gallo0933	Tannase	EC	WP_012961863	21-596	90
	Gallo1570	*Pil2* pilus subunit	CW	WP_012962246	24-605	90
	Gallo1675	Unknown function	CW	WP_012962333	40-724	102
	Gallo2018	Involved in bacteriocin synthesis	CW	WP_009855005	27-311	58
	Gallo2178[3]	*Pil1* pilus subunit (major pilin)	CW	WP_009855153	26-448	73
	Gallo2179[3]	*Pil1* pilus subunit (collagen-binding domain)	CW	WP_009855154	35-628	92

[1][52, 56, 58, 60, 119-123]; [2]NCBI Reference Sequence; [3]previously shown to be immunogenic [116-118]; [4]Molecular weight GST-tag 27.4 kD

Amino acid sequences of the selected antigens were analyzed for homologues in other bacterial species with "BLASTp" to assess specificity. The reference strain for *F. nucleatum* belonged to the *F. nucleatum* subsp. *nucleatum*. Proteins of other *F. nucleatum* subspecies were closely related to the selected proteins (> 74% identity) impeding a subspecies specific detection of antibody responses (Table 6). In 2015, after protein selection had been performed, a new fusobacterial species had been identified, *F. hwasookii* [124]. The selected proteins shared high sequence identity to proteins of this bacterium (up to 97%) and a serological distinction between *F. hwasookii* and *F. nucleatum* could not be guaranteed based on the antigen selection. However, the impact of *F. hwasookii* in human infection is unknown so far. *F. periodonticum* is the next bacterial relative and except for Fn0264 the selected proteins shared a sequence identity of less than 80% with proteins of this bacterium. Other, none fusobacterial species did not exceed an identity of 56%.

Several *S. gallolyticus* proteins shared high sequence identity with proteins of *S. equinus* (up to 97%) (Table 7). However, this bacterium is a frequent isolate of horses and only rarely identified in humans [66]. Other closely related species of *S. gallolyticus* were *S. pasteurianus* and *S. infantarius*. S. *infantarius* has a homologous protein to Gallo0748 (67% sequence identity). A protein of *S. pasteurianus* identified recently [125] shares a high sequence identity with Gallo1675 (99%). The sequence identity was below 75% for all other selected *S. gallolyticus* proteins to proteins of species of the *S. equinus*/*S. bovis* complex, the family of Streptococcacae or none streptococcal species.

Table 6: Amino acid sequence homologies of selected *F. nucleatum* subsp. *nucleatum* antigens to proteins of other bacteria

Antigen	Amino acid sequence homology (% query; % identity) to closest other *F. nucleatum* subspecies	Fusobacteriaceae	bacterial species
Fn0131	*F. nucleatum* subsp. *vincentii* (99%; 94%)	*F. periodonticum* (99%; 73%); *F. hwasookii*[1] (99%; 92%)	*Klebsiella michiganensis* (93%; 28%)
Fn0253	*F. nucleatum* subsp. *vincentii* (100%; 99%)	*F. periodonticum* (100%; 79%); *F. hwasookii*[1] (100%; 97%)	*Bordetella ansorpii* (98%; 55%)
Fn0264	*F. nucleatum* subsp. *polymorphum* (100%; 98%)	*F. periodonticum* (100%; 96%); *F. hwasookii*[1] (100%; 97%)	-
Fn0387	*F. nucleatum* subsp. *fusiforme* (100%; 96%)	*F. periodonticum* (100%; 73%); *F. hwasookii*[1] (100%; 72%)	*Campylobacter ureolyticus* (98%, 33%)
Fn1426	*F. nucleatum* subsp. *vincentii* (100%; 95%)	*F. necrophorum* (99%; 44%); *F. hwasookii*[1] (100%; 47%)	*Haemophilus ducreyi* (97%; 28%)
Fn1449	*F. nucleatum* subsp. *polymorphum* (100%; 93%)	*F. periodonticum* (100%; 85%); *F. hwasookii*[1] (100%; 89%)	*Campylobacter ureolyticus* (88%, 33%)
Fn1526	*F. nucleatum* subsp. *vincentii* (100%; 93%)	*F. russii* (100%; 63%)	*Campylobacter hominis* (89%, 31%)
Fn1817_1	*F. nucleatum* subsp. *animalis* (100%; 74%)	*F. periodonticum* (100%; 68%)	*Proteus mirabilis* (70%; 56%)
Fn1817_2	*F. nucleatum* subsp. *animalis* (100%; 99%)	*F. necrophorum* (100%; 75%)	*Bordetella bronchiseptica* (97%; 39%)
Fn1859	*F. nucleatum* subsp. *vincentii* (100%; 87%)	*F. periodonticum* (100%; 70%); *F. hwasookii*[1] (100%; 74%)	*Bordetella trematum* (92%; 26%)
Fn1893	*F. nucleatum* subsp. *fusiforme* (100%; 99%)	*F. periodonticum* (100%; 75%)	*Campylobacter ureolyticus* (98%, 33%)

[1]identified 2015 [124]; % query: percentage of query coverage that overlaps the subject sequence; % identity: percentage identity between the query and subject sequences over the coverage area

Table 7: Amino acid sequence homologies of selected *S. gallolyticus* antigens to proteins of other bacteria

Antigen	Amino acid sequence homology (% query; % identity) to closest other species of the		
	S. equinus/bovis complex	Streptococcaceae	bacterial species
Gallo0112A	*S. equinus*[1] (100%; 90%)	*S. uberis* (97%; 75%)	*Atopobium parvulum* (94%; 60%)
Gallo0112B	*S. equinus*[1] (100%; 91%)	*S. uberis* (96%; 71%)	*Lactobacillus equi* (96%; 50%)
Gallo0272	-	*S. suis* (73%; 38%)	*Atopobium parvulum* (74%; 33%)
Gallo0577	-	*S. parasanguinis* (95%; 42%)	*Lactobacillus fermentum* (97%; 56%)
Gallo0748	*S. equinus*[1] (100%; 86%) *S. infantarius* (100%; 67%)	*S. suis* (95%; 61%)	*Enterococcus cecorum* (97%; 56%)
Gallo0933	*S. equinus*[1] (100%; 97%)	*S. oralis* (99%; 69%)	*Butyrivibrio* sp. NC2007 (99%; 57%)
Gallo1570	*S. equinus*[1] (94%; 48%)	*S. iniae* (64%; 51%)	*Lactobacillus apodemi* (65%; 51%)
Gallo1675	*S. pasteurianus*[2] (100%; 99%) *S. equinus*[1] (97%; 68%)	*S. dysgalactiae* (67%; 27%)	*Parascardovia denticolens* (88%; 31%)
Gallo2018	*S. equinus*[1] (100%; 93%) *S. lutetiensis* (72%; 33%)	*S. cristatus* (43%; 64%)	*Lactobacillus acidophilus* (49%; 46%)
Gallo2178	*S. equinus*[1] (100%; 73%)	*S. anginosus* (100%; 58%)	Lachnospiraceae AC2014 (100%; 49%)
Gallo2179	*S. lutetiensis* (49%; 62%)	*S. equi* (99%; 55%) *S. agalactiae* (94%; 38%)	*Parascardovia denticolens* (97%; 45%)

% query: percentage of query coverage that overlaps the subject sequence; % identity: percentage identity between the query and subject sequences over the coverage area; [1]predominantly isolated from horses, rarely isolated from humans [66]; [2]strain HC-2909-2, sequence (RefSeq: WP_041973257) identified in whole genome shotgun sequencing (NCBI RefSeq: CDEY01000010), query 51% and identity 37% to homologue in reference strain *S. pasteurianus* ATCC43144

DNA sequences for selected amino acid sequences of *S. gallolyticus* and *F. nucleatum* proteins were synthesized by eurofins genomics (Ebersberg, Germany) after codon adaption for expression in *E. coli*. Synthesized genes were subcloned via 5'-end BamH1 and 3'-end SalI restriction sites into the pGEX4T3tag [126] vector (eurofins genomics (Ebersberg, Germany)). The resulting constructs encoded N-terminal GST, the protein of interest and C-terminal tag sequence (eleven C-terminal amino acids from the large T antigen of simian virus 40) (GST-X-tag fusion protein) (Figure 5). All resulting vector maps can be found in Appendix I.

Figure 5: GST-X-tag fusion protein as present on the plasmid and resulting recombinant protein (modified from Lena-Mareen Kranz)

2.2.2. Recombinant expression of selected *F. nucleatum* and *S. gallolyticus* proteins in *E. coli* BL21

Recombinant expression of selected proteins was performed in *E. coli* BL21. Electrocompetent *E. coli* BL21 were grown in LB medium and stored at -80°C (provided by Ute Koch). Plasmids encoding the GST-X-tag fusion proteins were delivered by eurofins genomics (Ebersberg, Germany) in amounts of up to 8 µg of lyophilized plasmids. DNA was dissolved in 20 µl ddH_2O. One µl of a further 1:20 dilution was combined with 50 µl of electrocompetent *E. coli* BL21 that were thawed on ice. Bacteria were transformed in precooled electroporation cuvettes with the electroporation devices Gene Pulser and Pulse Controller with the following settings: voltage of 2.3 kV, resistance of 200 Ω and capacity of 960 µF (Controller) and 25 µF (Gene Pulser). Two hundred µl of LB medium were added and transformed bacteria were incubated for 1 hour at 37°C on a shaker. Twenty µl of the incubated bacteria were plated on LB_{amp} agar plates and incubated overnight at 37°C. One colony was picked and combined with 5 ml LB_{amp} medium and incubated for 6 hours at 37°C on a shaker. The 5 ml cultures were transferred to 250 ml LB_{amp} medium and incubated overnight at 37°C on a shaker. After the incubation 50 ml of the 250 ml overnight culture were centrifuged (15 min at 5,000 rpm) and the pellet was used for plasmid purification and subsequent analytical digestion (2.2.5). Seven hundred µl of bacterial culture were added to

700 µl 50% glycerol to obtain a glycerol stock of transformed bacteria that was stored at -80°C. The remaining bacterial culture was transferred to 1 l $LB_{amp.}$ Protein expression was induced with 0.5 mM IPTG at an optical density of 0.5 measured at 600 nm. After 6 hours of incubation at room temperature cultures were centrifuged at 6,000 rpm for 10 min at 4°C (rotor ThermoScientic F12-6x500 LEX) and pellets were stored in 10 ml PBS at -20°C until bacterial lysis (2.2.3).

2.2.3. Bacterial lysis

Frozen pellets (2.2.2) were thawed in a water bath at 37°C. Thawed pellets were kept on ice and 20 µl per 10 ml resuspended bacterial pellet of 1 M DTT (final concentration 2 mM) as well as 0.5 ml protease inhibitor per 1 l of original expression culture were added. Bacterial lysis was performed with a precooled high pressure homogenizer according to the manufacturer's protocol. Pressure of 1,000-1,500 bar was applied once for 2 min to ensure disruption of bacterial cells. One hundred µl aliquots of the total lysate were taken for subsequent Western blot analyses (2.2.8). Lysates were cleared by centrifugation at 14,000 rpm for 1 hour at 4°C (rotor Sorvall SA-600). A second aliquot (100 µl) was taken for subsequent Western blot analyses (2.2.8). Cleared lysates were combined with 100% glycerol 1+1 to obtain the final lysate and stored at -20°C.

2.2.4. Quantification of lysate protein concentration by Bradford assay

Total lysate protein concentration was determined using the Bradford reagent. Eight hundred µl ddH_2O, 200 µl Bradford reagent and 0.5 µl of cleared lysate were mixed and incubated for 5 min. Absorption was measured at $OD_{595.}$ A sample containing water instead of lysate thereby served as reference. The protein concentration was calculated as follows based on a calibration curve prepared with BSA:

$$c_{protein}\ (\mu g/\mu l) = OD_{595} * 44\ \mu g/\mu l\ \text{lysate}$$

2.2.5. Analytical DNA digestion

Plasmid DNA was isolated from bacterial pellets harvested just before induction (2.2.2) using the QIAprep Spin Midiprep Kit according to the manufacturer's instructions. The concentration of plasmid DNA was determined with the NanoDrop ND-100 and adjusted to a final concentration of 0.5 µg/µl. One µl plasmid DNA (500 ng), 1 µl or 2 µl of restriction buffer (1x or 2x), 0.3 µl of each enzyme and ddH_2O up to a total volume of 10 µl were

incubated for 1.5 hours at 37°C. The digests were analyzed by agarose gel electrophoresis (2.2.6).

Table 8: Enzymes used for asymmetric analytical digestions

Antigen	Enzyme 1	Enzyme 2
F. nucleatum		
Fn0131	NdeI	BsaA1
Fn0253	MfeI	PstI
Fn0264	XhoI	PstI
Fn0387	NsiI	PstI
Fn1426	EcoRI	PstI
Fn1449	NdeI	AlwNI
Fn1526	MluI	
Fn1817_1	MluI	
Fn1817_2	PvuII	
Fn1859	BsmI	PstI
Fn1893	EcoRV	
S. gallolyticus		
Gallo0112A	NsiI	SalI
Gallo0112B	SphI	SalI
Gallo0272	BsmI	SalI
Gallo0577	BsmI	SalI
Gallo0748	EcoRI	SalI
Gallo0933	KpnI	SalI
Gallo1570	KpnI	SalI
Gallo1675	HindIII	SalI
Gallo2018	BsmI	SalI
Gallo2178	KpnI	SalI
Gallo2179	HindIII	SalI

2.2.6. Agarose gel electrophoresis

Agarose gel electrophoresis was applied to verify the correct lengths of digested DNA. 1.5 g agarose was dissolved in 150 ml 1x electrophoresis buffer by heating it up in a microwave oven. Five µl peqGreen DNA dye were added. The polymerized gel was run in an electrophoresis chamber filled with 1x electrophoresis buffer. Ten µl of sample were mixed with 2 µl of 6x sample buffer and loaded onto the gel. 5 µl of DNA marker (smart ladder) served as reference to determine the size of fragments. Gel electrophoresis was performed for 1 hour at a voltage of 100 V. Gel documentation was done with the Gel Doc^{TM} EZ Imager.

2.2.7. Polymerase chain reaction with subsequent sequencing

Plasmid DNA was purified from 200 µl of cleared lysate (2.2.3) by QIAprep Spin Miniprep Kit according to the manufacturer's instructions. PCR was performed using the QIAGEN multiplex PCR kit. Fifty ng of purified DNA were mixed with 12.5 µl QIAGEN Multiplex PCR master mix, 0.25 µl sense primer (pGEXs+T3, final concentration 0.1 µM), 0.25 µl antisense primer (pGEXas+T7, final concentration 0.1 µM), 2.5 µl Q-solution and DNase/RNase free H_2O in a total volume of 25 µl. One µl DNase/RNase free H_2O instead of template was used as negative control. The PCR-program was as follows:

1. 95°C – 15 min: initial denaturation
2. 94°C – 30 sec: denaturation
3. 68 °C – 90 sec: annealing
4. 72 °C – 1 min/ 1000 base pairs (bp): elongation
   45 cycles of step 2. – 4.
5. 72 °C – 10 min: final elongation

Ten µl of PCR product were mixed with 2 µl of 6x DNA sample buffer and analyzed by agarose gel electrophoresis (1% agarose gel, separation at 100 V for 1 hour) (2.2.6).

The primers (sense: pGEXs+T7 and antisense: pGEXas+T3) used for PCR and their location are shown in Figure 6. Due to primer length and location the product size is 98 bp larger than the respective insert size.

If the PCR product was identified at the expected size in the agarose gel the remaining 15 µl of PCR product were purified using the QiaQuick PCR purification kit according to manufacturer's instructions. The purified product was sent for sequencing to eurofins genomics (Ebersberg, Germany) and sequencing was performed via the T3 and T7 primer sequences attached to the PCR product. If sequencing via T3 and T7 primers was unsuccessful purified plasmids were sent for sequencing with primers pGEXfor and pGEXrev with the disadvantage of lower quality of the sequencing results. Obtained sequences were compared to the expected DNA sequences using "Clustal Omega" to verify presence of the expected insert in the respective lysate.

A

```
841  ccagcaagta tatagcatgg cctttgcagg gctggcaagc cacgtttggt ggtggcgacc atcctccaaa
     ggtcgttcat atatcgtacc ggaaacgtcc cgaccgttcg gtgcaaacca ccaccgctgg taggaggttt
     >.....................................GST.....................................>

                          BamHI              SalI
                        --+----             -+-----
911  atcggatctg gttccgcgtg gatccccgaa ttcccgggtc gacaaacctc ccacacctcc ccctgaacct
     tagcctagac caaggcgcac ctaggggctt aagggcccag ctgtttggag ggtgtggagg gggacttgga
     >.........GST.........>>
                                                  >>...........tag............>

981  gaaacataag cggccgcatc gtgactgact gacgatctgc ctcgcgcgtt tcggtgatga cggtgaaaac
     ctttgtattc gccggcgtag cactgactga ctgctagacg gagcgcgcaa agccactact gccacttttg
     >...>> tag

1051  ctctgacaca tgcagctccc ggagacggtc acagcttgtc tgtaagcgga tgccgggagc agacaagccc
      gagactgtgt acgtcgaggg cctctgccag tgtcgaacag acattcgcct acggccctcg tctgttcggg
```

B

pGEXs+T7: 5'TAATACGACTCACTATAGGGtccaaaatcggatctggttccgcgtgga3'

pGEXas+T3: 5'AATTAACCCTCACTAAAGGGgatgcggccgcttatgtttcaggttcaggg3'

Figure 6: Location of primers for quality control PCR and sequencing. A) Sequence details of the primer binding region. Highlighted in green are pGEXs and pGEXas. Highlighted in yellow are primer pGEXfor and pGEXrev for sequencing of mini-Prep DNA without preceding PCR. **B)** pGEXs and pGEXas primer with attached T3 and T7 primer (capital letters) sequences. Adapted from Martina Willhauck-Fleckenstein.

2.2.8. Western blot

SDS-PAGE using the Mini-PROTEAN II system was performed for Western blot analyses. SDS-polyacrylamide gels were prepared as depicted in Table 9. Samples from total and cleared lysates (2.2.3) were adjusted to 1 µg/µl total protein concentration with ddH_2O in a total volume of 100 µl including 25 µl 4x SDS sample buffer. Samples were heated at 95°C for 5 min before loading. GST-tag lysate served as positive control and was treated the same way. SDS-gels were placed into the gel chamber and covered with 1x SDS running buffer. Ten µl of samples and 5 µl of pre-stained protein ladder were loaded onto the gels and the gel was run at 200 V for 50 min.

Table 9: Protocol for two SDS-polyacrylamide gels. The resolving gel is poured first followed by the stacking gel.

Reagent	Resolving gel (13.5%)	Stacking gel (5%)
H_2O	1.6 ml	3.675 ml
30% (w/v) acrylamid)	3.75 ml	0.625 ml
1 M Tris-HCl (pH 8.8)/1 M Tris-HCl (pH 6.8)	4.5 ml	0.665 ml
10% (w/v) SDS	0.1 ml	0.05 ml
TEMED	0.005 ml	0.005 ml
10% (w/v) APS	0.05 ml	0.025 ml

A nitrocellulose membrane, two Whatman papers and two pads per gel were soaked in EMBL transfer buffer. Proteins were blotted onto the membrane with the Mini Trans-Blot Electrophoretic Transfer Cell in a "sandwich" format: pad, Whatman paper, membrane, gel, Whatman paper and pad were layered into the transfer cartridge. The transfer chamber was supplied with an ice block and filled with EMBL transfer buffer. The transfer was performed at 100 V for 1 hour.

After transfer, membranes were blocked with 10% Western blot blocking buffer for 1 hour on a shaker at room temperature. Membranes were washed three times for 10 min with PBS-T and subsequently incubated with primary antibodies diluted in 25 ml 5% Western blot blocking buffer. Primary antibodies were directed either against the N-terminal GST (polyclonal rabbit anti-GST antibody, 1:10,000) or against the C-terminal tag (monoclonal mouse anti-tag antibody, 1:5,000). Incubation with primary antibody was done at 4°C overnight on a shaker. Membranes were washed three times for 10 min with PBS-T. Secondary antibodies (goat anti-rabbit IgG HRP conjugate and goat anti-mouse IgG HRP conjugate) were diluted 1:10,000 in 5% Western blot blocking buffer and incubated with the membranes for 1 hour at room temperature on a shaker. Membranes were washed three times for 10 min with PBS-T and incubated with ECL™ Western Blotting Detection reagents. Enhanced luminescence signals were visualized with x-ray films. Exposure time varied between 30 sec and 2 min.

2.2.9. Anti-tag ELISA

96-well microtiter plates were coated with 100 µl ELISA coating buffer per well and incubated overnight. Lysates were diluted in polystyrene plates with a 1:3 dilution series in ELISA blocking buffer starting at a concentration of 2 µg/µl in 300 µl volume. Lysate dilution series were prepared in duplicates. GST-tag lysate served as a reference and was

diluted the same way. ELISA coating buffer was discarded and coated 96-well microtiter plates were incubated with 180 µl ELISA blocking buffer for 1 hour at 37°C. Hundred µl of lysate dilution were added to the blocked plates after discarding the ELISA blocking buffer and incubated for 1 hour at room temperature on a shaker. Antigen dilutions were discarded and plates were washed five times with PBS-T. Plates were dried by knocking them on paper towels and 100 µl of antibody directed against the C-terminal tag were added (mouse anti-tag antibody, 1:5,000 in ELISA blocking buffer). After incubation for 1 hour at room temperature on a shaker, plates were washed 5 times with PBS-T and dried as described above. Plates were incubated with 100 µl of goat anti-mouse IgG HRP conjugate per well (1:10,000 in ELISA blocking buffer) for 1 hour at room temperature. Plates were washed five times with PBS-T and dried. Hundred µl of ELISA substrate solution were added per well and incubated for 2-8 min until the reaction was stopped with 50 µl of ELISA stop solution. Absorption was measured at 450 nm with the Multiskan PLUS MKII.

2.2.10. Multiplex serology

Recombinantly expressed GST-X-tag fusion proteins were affinity-purified on fluorescent polystyrene beads with coupled glutathione-casein (provided by Monika Oppenländer, [114]). Bacterial lysates were diluted to 1 mg/ml total protein with multiplex serology blocking buffer in a volume of 1 ml. Beads were added, mixed thoroughly and incubated for 1 hour at room temperature in the dark on a shaker. Beads were washed three times by intermitting centrifugation at 13,000 rpm for 2 min, discarding the supernatant and adding 1 ml of multiplex serology blocking buffer. Loaded beads were stored in storage buffer until further use.

Serum samples were diluted in polystyrene flat-bottom plates as indicated for the individual experiments in a total volume of 100 µl with serum pre-incubation buffer. The pre-incubation buffer contained 1 mg/ml casein and 2 mg/ml of GST-tag lysate to block unspecific binding of antibodies directed against residual bacterial proteins, and the N-terminal GST and C-terminal tag. Further, PVX was added mimicking the bead surface to suppress unspecific binding of antibodies to the beads [127]. Sera were incubated for 1 hour at room temperature on a shaker.

The multiplex serology assay was performed as described in [114]. Beads loaded with antigen were resuspended by four times intermitting sonification for 30 sec and vortexing. 96-well filter plates were incubated with ddH_2O for 10 min and dried using a vacuum manifold at 10

inHg of negative pressure and subsequent knocking on the lid with a hammer to mobilize residual liquid at the membrane. Beads were mixed and 50 µl of the bead mix were incubated with 50 µl of pre-incubated sera in filter plates for 1 hour at room temperature on a shaker in the dark. Serum was removed from the plates and plates were washed three times with 100 µl per well of multiplex serology blocking buffer and then dried as described above. 100 µl of biotinylated secondary antibody (goat anti-human IgA, IgM and IgG or goat anti-mouse IgG (1:1,000 in multiplex serology blocking buffer) or goat anti-bovine IgA, IgM and IgG (1:2,000 in multiplex serology blocking buffer)) were added to each well and incubated for 1 hour at room temperature on a shaker in the dark. One well containing beads but no serum was incubated with biotinylated mouse anti-tag (1:100 in multiplex serology blocking buffer) as a bead-loading control. Plates were washed and dried as described above and 100 µl of Strep-PE (1:750 in blocking buffer) was added to each well. Plates were incubated on a shaker for 30 min at room temperature in the dark. After washing 100 µl of storage buffer were added per well and plates were stored at 4°C overnight.

Serum antibodies bound to affinity-purified antigens on beads were quantified using the Luminex 100 or 200 analyzer. The output was given as the median fluorescence intensity (MFI) of at least 100 beads per bead set measured. Net MFI values were generated by subtraction of bead-background (one well per plate without serum but beads and secondary reagents) and GST-background (one bead set loaded with GST-tag lysate). Net values below one were set to one. Sera with GST-background higher than 150 MFI were excluded from analyses. Plate controls, standard sera pipetted on each plate within one assay, served for control of technical inter-plate variation. Bead-loading was controlled with the biotinylated mouse anti-tag antibody directed against the C-terminal tag of recombinantly expressed proteins. A positive pipetting control was given by the highly seroprevalent antigen VP1 of Polyomavirus JC [128] loaded onto one bead-type.

Cut-offs for antibody positivity were set arbitrarily as indicated.

2.2.11. Study designs

2.2.11.1. BliTz and DACHSplus

Study samples and study data were kindly provided by Hermann Brenner. A detailed description of the design of this case-control study was described elsewhere [129]. Briefly, serum samples included were part of the BliTz study (“Begleitende Evaluierung innovativer

Testverfahren zur Darmkrebsfrüherkennung") or DACHSplus study, a substudy of the DACHS study ("Darmkrebs: Chancen der Verhütung durch Screening") [130-132]. BliTz participants were recruited at screening colonoscopy in gastroenterology practices in southern Germany between 2005 and 2013. A random sample of 228 subjects with no colorectal neoplasm identified during colonoscopy served as controls. DACHSplus CRC cases were recruited after diagnosis but before treatment at four hospitals in Southern Germany. The subset analyzed here included 318 prevalent CRC cases (International Classification of Diseases (ICD-) codes C18.0-C18.7, C18.9, C19 and C20 as by the 10th Revision of the International Statistical Classification of Diseases, Injury and Causes of Death). Since the initial design aimed to resemble a true screening setting, DACHSplus CRC cases were not matched to BliTz controls [129]. Thus, DACHSplus CRC cases differed significantly from BliTz controls with cases being more frequently males, older and ever smokers (Table 10). A slight majority of DACHSplus CRC cases presented with UICC stages I and II (55%) compared to 45% with UICC stage III and IV.

I further analyzed a set of gastric cancer (GC) cases that was compared to BliTz controls to assess whether potentially observed associations were specific for CRC. These gastric cancer cases were also part of the DACHSplus study [133] and compared to BliTz controls significantly more often males, older and ever smokers (Table 10).

Table 10: Baseline Characteristics of BliTz controls and DACHSplus CRC and GC cases

		BliTz Controls (n=228) n (%)	DACHSplus CRC (n=318) n (%)	p-value*	DACHSplus GC (n=129) n (%)	p-value*
Sex	female	124 (54)	133 (42)		44 (34)	
	male	104 (46)	185 (58)	**0.004**	85 (66)	**0.0002**
Age (years)	31-59	96 (42)	74 (23)		41 (32)	
	60-65	66 (29)	56 (18)		19 (15)	
	66-94	66 (29)	187 (59)	**<0.0001**	69 (53)	**<0.0001**
	missing	0	1		0	
	mean (range)	62 (40-85)	68 (31-94)		64 (30-89)	
School education	< 10 years	125 (56)	197 (64)		57 (50)	
	≥ 10 years	98 (44)	111 (36)	0.066	57 (50)	0.291
	missing	5	10		15	
Smoking status	never	132 (58)	152 (49)		72 (62)	
	ever	95 (42)	158 (51)	**0.037**	45 (38)	**0.0005**
	missing	1	8		12	
BMI	< 25	89 (40)	118 (38)		50 (43)	
	25-29.9	95 (42)	134 (44)		52 (45)	
	≥ 30	40 (18)	56 (18)	0.946	13 (11)	0.265
	missing	4	10		14	
Family history of CRC	no	200 (88)	274 (89)		105 (90)	
	yes	28 (12)	35 (11)	0.734	12 (10)	0.578
	missing	0	9		12	

*Pearson's Chi-Square-test as compared to controls; significant associations are marked in bold font

Apart from colonoscopy-negative individuals (controls) there were also colorectal neoplasms identified in the BliTz study that were sub-grouped in non-advanced adenoma, advanced adenoma (high-grade dysplasia, villous architecture without high-grade dysplasia or large adenoma (> 10 mm)) and CRC. A separate approach of analyses included samples of these three groups (non-advanced adenoma n=30, advanced adenoma n=100, and CRC n=50) that were compared to the BliTz controls described above (n=228). BliTz CRC cases were significantly more males, older and ever smokers. BliTz advanced adenoma cases had a lower educational level than BliTz controls. BliTz non-advanced adenoma cases did not differ significantly from BliTz controls (Table 11).

All studies had been approved by the ethics committees of the University of Heidelberg and of the respective state medical boards. Informed consent had been obtained from each participant.

Table 11: Baseline Characteristics of BliTz controls, precursors and CRC cases

		Controls (n=228)	Non-advanced adenoma[1] (n=30)		Advanced adenoma[2] (n=100)		CRC (n=50)	
		n (%)	n (%)	p-value[3]	n (%)	p-value[3]	n (%)	p-value[3]
Sex	female	124 (54)	13 (43)		50 (50)		16 (32)	
	male	104 (46)	17 (57)	0.254	50 (50)	0.464	34 (68)	**0.004**
Age (years)	31-59	96 (42)	11 (37)		30 (30)		9 (18)	
	60-65	66 (29)	7 (23)		31 (31)		15 (30)	
	66-94	66 (29)	12 (40)	0.459	39 (39)	0.084	26 (52)	**0.002**
	mean (range)	62 (40-85)	64 (55-78)		63 (50-86)		67 (55-81)	
School education	< 10 years	125 (56)	16 (53)		67 (68)		30 (63)	
	≥ 10 years	98 (44)	14 (47)	0.778	32 (32)	**0.050**	18 (37)	0.413
	missing	5	0		1		2	
Smoking status	never	132 (58)	16 (53)		48 (48)		20 (40)	
	ever	95 (42)	14 (47)	0.616	52 (52)	0.089	30 (60)	**0.020**
	missing	1	0		0		0	
BMI	< 25	89 (40)	8 (27)		37 (38)		12 (24)	
	25-29.9	95 (42)	15 (50)		40 (41)		25 (52)	
	≥ 30	40 (18)	7 (23)	0.375	20 (21)	0.843	12 (24)	0.127
	missing	4	0		3		1	
Family history of colorectal cancer	no	200 (88)	26 (87)		82 (82)		41 (82)	
	yes	28 (12)	4 (13)	0.869	18 (18)	0.170	9 (18)	0.281
	missing	0	0		0		0	

[1]tubular adenoma, adenoma <10mm; [2]High grade dysplasia, villous adenoma with high-grade dysplasia, large adenoma (>10mm) with neither high-grade dysplasia nor villous architecture; [3]Pearson's Chi-square test in comparison to controls; significant associations are marked in bold font

2.2.11.2. Southern community cohort study (SCCS)

Study samples and study data were kindly provided by Meira Epplein. The nested case-control design presented here has been published elsewhere [134]. Briefly, study samples included were part of the SCCS, which is a large prospective cohort that enrolled approximately 86,000 men and women, aged 40-79, between 2002 and 2009 from 12 southeastern US states [135]. Participants were recruited from community health care centers (CHC) (86%) or via mail (14%) and represent a low-income and mainly uninsured population of Caucasian-Americans and a substantial number of African-Americans [135]. Participants completed a questionnaire (in-person, comprehensive computer-assisted for CHC recruited individuals or paper version for individuals recruited via mail) that addressed information about demographic and lifestyle factors including regular diet, personal and family medical history and health services utilization. Participants recruited at CHC provided a venous blood sample [134].

Briefly, among participants that donated a blood sample 188 incident CRC cases (International Classification of Diseases for Oncology, Third Edition (ICD-O-3) codes C18.0-C18.9, C19.9, C20.9) were identified via state cancer registries and/or the National Death Index mortality records until the end of 2011. For the nested case-control design each case was matched with two controls on age (± 2 years), ethnicity (African-American, Caucasian-American, or other), sex, menopausal status (women), CHC site and date of sample collection (± 6 months) [134]. Due to missing baseline data, lack of serum sample, duplicates, laboratory reasons and the missing matched counterparts 7 cases and 28 controls had to be excluded resulting in a final number of 181 cases and 348 controls.

At baseline incident CRC cases differed significantly from SCCS controls in smoking status with more cases being never smokers (Table 12). None of the other characteristics analyzed here was significantly different. The incident CRC cases were predominantly females (54%) and of African-American ethnicity (82%). The average age at diagnosis was 59 years with a range of 40 to 81 years. The average time between blood draw and diagnosis was 3.2 years ranging from 0.4 to 8.1 years. Tumor stages according to SEER staging [136] were predominantly localized (39%) followed by metastatic (36%) and distant (25%). The majority of incident CRC cases were located in the colon (75%) (Table 13).

The SCCS had been reviewed and approved by the institutional review boards at Vanderbilt University and Meharry Medical College. Written informed consent had been obtained from all participants.

Table 12: Baseline characteristics of the CRC case-control study nested within SCCS

		Controls (n=348) n (%)	Cases (n=181) n (%)	p-value*
Sex	female	191 (55)	98 (54)	
	male	157 (45)	83 (46)	0.871
Age at blood draw (years)	40-55	164 (47)	86 (48)	
	56-60	76 (22)	37 (20)	
	61-77	108 (31)	58 (32)	0.927
	mean (range)	56 (40-77)	57 (40-77)	
Ethnicity	Caucasian-American	60 (18)	32 (18)	
	African-American	276 (82)	142 (82)	0.882
	missing	12	7	
Education	< high school	135 (40)	78 (44)	
	≥ high school	206 (60)	100 (56)	0.352
	missing	7	3	
Smoking	never	95 (28)	68 (38)	
	former	108 (32)	58 (33)	
	current	138 (40)	52 (29)	**0.018**
	missing	7	3	
BMI	< 25	84 (25)	35 (20)	
	25-29.9	96 (28)	62 (36)	
	≥ 30	160 (47)	76 (44)	0.184
	missing	8	8	
Family history of CRC	no	127 (86)	55 (77)	
	yes	20 (14)	16 (23)	0.096
	missing	201	110	

*Pearson's Chi-Square-test; significant associations are marked in bold font

Table 13: Characteristics of cases in the CRC case-control study nested within SCCS at time of diagnosis

		Cases (n=181) n (%)
Age at diagnosis (years)	38-59	93 (51)
	60-65	40 (22)
	66-81	48 (27)
	mean (range)	59 (40-81)
Time between blood draw and diagnosis (years)	< 2	57 (31)
	2-4	61 (34)
	> 4	63 (35)
	mean (range)	3.2 (0.5-8.3)
SEER Stage	localized	64 (39)
	regional	59 (36)
	distant	41 (25)
	missing	17
Site	colon	129 (75)
	rectum	44 (25)
	missing	8

2.2.11.3. European prospective investigation into nutrition and cancer (EPIC)

The case-control study presented here is nested within EPIC [137], a large multi-national prospective cohort with 23 centers in 10 Western European countries (Denmark, France, Greece, Germany, Italy, Netherlands, Norway, Spain, Sweden, United Kingdom). Access to biological samples and data from the EPIC cohort was authorized by the EPIC steering Committee. A detailed description of the EPIC study design is published elsewhere [138]. Briefly, 521,468 participants, aged 35 to 70 years, were enrolled between 1992 and 2000. Dietary and lifestyle data as well as biological samples, including blood, were collected at enrollment.

The nested CRC case-control study analyzed here included pre-diagnostic serum samples of 485 incident CRC cases (primary tumors, ICD codes C18.0-C18.9, C19 and C20 as by the 10th Revision of the International Statistical Classification of Diseases, Injury and Causes of Death). These cases were a random subset of the larger total CRC cases accrued within the cohort, identified until 2004 and with existing sufficient volume of serum for the laboratory analyses [139]. The average age at diagnosis was 63 years (range 38 to 81 years) and the average time between blood draw and diagnosis was 3.4 years (range 0.4 to 8.5 years). The majority of cases (89%) were diagnosed with cancer in the colon and 49% of cases were of localized stage according to the EPIC staging classification [140] (Table 15).

485 controls were matched by age at blood collection (± 6 month to ± 2 years), sex, study center, time of the day at blood collection (± 2 to 4 hours interval), fasting status at blood collection (< 3/3-6/6 hours); among women by menopausal status, and among premenopausal women, by phase of menstrual cycle and hormone replacement therapy use at time of blood collection. Controls were free of cancer by time of matching (except for non-melanoma skin cancer). There was no significant difference between controls and cases in the baseline characteristics assessed here (Table 14).

Table 14: Baseline characteristics of the CRC case-control study nested within EPIC

		Controls (n=485) n (%)	Cases (n=485) n (%)	p-value*
Sex	female	247 (51)	247 (51)	
	male	238 (49)	238 (49)	1.000
Age at blood draw (years)	37-55	120 (25)	121 (25)	
	56-60	124 (25)	122 (25)	
	61-77	241 (50)	242 (50)	0.989
	mean (range)	60 (37-77)	59 (37-77)	
Country	Italy/Greece	112 (23)	112 (23)	
	France/Spain	93 (19)	93 (19)	
	UK	134 (28)	134 (28)	
	Germany/Netherlands	146 (30)	146 (30)	1.000
Education	none/primary school completed	212 (45)	215 (46)	
	technical/professional	115 (25)	95 (21)	
	secondary school/longer education	142 (30)	153 (33)	0.317
	missing	16	22	
Smoking status	never	234 (48)	202 (42)	
	former	154 (32)	183 (38)	
	current	95 (20)	96 (20)	0.089
	missing	2	4	
Alcohol intake	never	48 (10)	28 (6)	
	former	40 (9)	39 (9)	
	current 0-6 g/day	133 (29)	141 (31)	
	6.01-20 g/day	130 (28)	126 (28)	
	> 20 g/day	109 (24)	120 (26)	0.195
	missing	25	31	
BMI	< 25	167 (34)	160 (33)	
	25-29.9	238 (49)	220 (45)	
	≥ 30	80 (16)	105 (22)	0.120
Family history of CRC	no	176 (93)	173 (90)	
	yes	13 (7)	19 (10)	0.288
	missing	296	293	

*Pearson's Chi-Square-test

Table 15: Characteristics of cases in the CRC case-control study nested within EPIC at time of diagnosis

		Cases (n=485) n (%)
Age at diagnosis (years)	38-59	158 (33)
	60-65	123 (25)
	66-81	204 (42)
	mean (range)	63 (38-81)
Time between blood draw and diagnosis (years)	< 2	130 (27)
	2-< 3.5	127 (26)
	3.5-< 5	119 (25)
	≥ 5	109 (22)
	mean (range)	3.4 (0.01-8.5)
EPIC Stage	localized	154 (49)
	metastasic	162 (51)
	missing	169
Site	colon	432 (89)
	rectum	53 (11)

2.2.11.4. Irish CRC case-control study

Study samples and study data were kindly provided by David Hughes. Participants of the Irish CRC case-control study were recruited at the Departments of Gastroenterology and Surgery at the Adelaide and Meath Hospital (AMNCH) in Dublin, Ireland between 2008 and 2011. The majority (n=235, 80%) of the in total 292 participants in this sub-study were part of the AMNCH immunochemical FOBT CRC screening pilot program [141]. In this screening program approximately 10,000 individuals aged between 50 and 75 years in the AMNCH catchment area were invited to have an immunochemical fecal occult blood test (FIT) performed. FIT-positive individuals were further invited for colonoscopy. The remaining 57 participants were recruited at the gastroenterology and surgery department of AMNCH and presented with positive FIT, rectal bleeding, control colonoscopy or other reasons. Of the in total 292 participants 37 were found in colonoscopy to be normal (controls) and 63 were found with minor diagnoses including 39 participants with hemorrhoids, 27 with diverticulosis, 2 with mucosal inflammation/ulceration, 2 with erythema, and each 1 participant with melanosis coli, menorrhagia/disordered proliferative endometrium, mild active colitis and diarrhea. 192 participants were diagnosed with colorectal neoplasm including polyps (n=85, hyperplastic polyps and small adenoma less than 10 mm in diameter), adenoma (n=60, more than 10 mm in diameter, including tubular, villous or tubulovillous adenoma), adenoma with high-grade dysplasia (n=22) and CRC (n=25). Lifestyle information

was limited for these patients, however, information on age and sex revealed that individuals with minor diagnoses and colorectal neoplasm were more frequently of male sex and that especially CRC cases were of older age compared to controls.

Table 16: Baseline characteristics of the Irish CRC case-control study

		Controls (n=37)	Minor diagnoses[1] (n=63)	p-value[4]	Polyp[2] (n=85)	p-value[4]	Adenoma[3] (n=60)	p-value[4]	High grade dysplasia (n=22)	p-value[4]	CRC (n=25)	p-value[4]
Sex	female	23 (62)	30 (48)		34 (40)		28 (47)		8 (36)		13 (52)	
	male	14 (38)	33 (52)	0.160	51 (60)	**0.024**	32 (53)	0.138	14 (64)	0.055	12 (48)	0.426
Age at blood draw (years)	36-59	18 (49)	29 (46)		32 (38)		17 (28)		8 (36)		10 (40)	
	60-65	11 (30)	16 (30)		26 (31)		20 (33)		7 (32)		2 (8)	
	66-109	8 (22)	8 (22)	0.730	27 (32)	0.429	23 (38)	0.095	7 (32)	0.590	13 (52)	**0.022**
	mean (range)	59 (42-71)	61 (51-71)		62 (44-75)		64 (50-109)		62 (44-84)		66 (36-89)	

[1]includes hemorrhoids, diverticulosis, mucosal ulceration/inflammation, melanosis coli, menorrhagia/disordered proliferative endometrium, mild active colitis, diarrhea, erythema; [2]hyperplastic polyp or small tubular adenoma (< 10mm); [3]tubular adenoma, tubulovillous adenoma, villous adenoma; [4]Pearson's Chi-square test in comparison to controls;

2.2.12. Statistical analyses

Differences between cases and controls in baseline characteristics as well as risk factors for *F. nucleatum* and *S. gallolyticus* positivity were analyzed by Pearson's Chi-Square test. Correlations between positivity to individual proteins in *F. nucleatum* or *S. gallolyticus* multiplex serology were also analyzed by Pearson's Chi-Square test.

Association of antibody responses to *F. nucleatum* and *S. gallolyticus* with prevalent CRC in the BliTz-DACHSplus study were assessed by logistic regression models to compute OR and 95% CI. The respective associations with CRC risk in the SCCS and EPIC studies were assessed by conditional logistic regression models, since cases were matched to controls. The following variables were considered to potentially confound the associations: age and sex (only in case of BliTz-DACHSplus), BMI, education, smoking and alcohol status (data only available in EPIC). In BliTz-DACHSplus sex was identified as significantly related to the outcome CRC as well as to *F. nucleatum* and *S. gallolyticus* positivity and was therefore considered as important confounder in the model. None of the other potential confounders was identified in any other study as being simultaneously related to the outcome (CRC) and the exposure (*F. nucleatum* / *S. gallolyticus* positivity) and would therefore not be considered to substantially influence the estimate. Indeed comparison of unadjusted models to models adjusting for age and sex (BliTz-DACHSplus only), BMI, education, smoking and alcohol status (EPIC only) did not alter the estimate by more than 15%, which is only slightly higher than the often in epidemiology applied 10% change mark as criterion for adjustment (Appendices V and VI). Although adjustment for potential confounders BMI, education, smoking and alcohol status did not alter the estimate substantially, it was decided in personal communication with David Hughes and Mazda Jenab to include these variables in the final model for the EPIC study in order to be consistent with statistical analyses models applied to EPIC data and with known information about CRC confounders. To make the individual studies more comparable, models in BliTz-DACHSplus and SCCS were also applied under adjustment for BMI, education and smoking status (data on alcohol not available). Missings in the individual variables thereby decreased the final number of samples included in the estimate calculation and were therefore included as individual categories to save statistical power. Sensitivity analyses comparing both, exclusion of missings and inclusion as individual category, did not exhibit substantial differences in the estimates (Appendices V and VI).

A sensitivity analyses was carried out in the EPIC and SCCS studies to exclude reverse causation in the potential prospective associations. Associations were estimated under exclusion of cases diagnosed within 2 years after blood draw and their respective matched controls.

Sex, age and country of residence are considered major risk factors for CRC [142]. To assess whether these variables affect the association of antibody responses to *F. nucleatum* and *S. gallolyticus* proteins with CRC I further estimated the association stratified by the baseline characteristics age at blood draw, sex, country of residence (EPIC) or ethnicity (SCCS). To potentially identify subgroups at increased risk for developing CRC with antibody responses to *S. gallolyticus* and *F. nucleatum* I estimated the association for certain case subgroups, i.e. tumor stage, tumor site and age at diagnosis in separate (conditional) logistic regression models under adjustment for above mentioned confounders.

The sample numbers for precancerous lesions and CRC in the BliTz and Irish case-control study were rather small (n <100) and important baseline data in the Irish case-control study was missing. Therefore analyses of associations of antibody responses to *S. gallolyticus* proteins and precancerous lesions in these two studies should be regarded as exploratory and were carried out in a crude model comparing the frequencies of antibody positivity between the groups by Pearson's Chi-Square test.

Significance of the associations was further assessed with Bonferroni corrections to address multiple testing. The p-value indicating significance thereby decreased to 0.004 with 12 possibilities for *F. nucleatum* positivity (eleven individual proteins and positivity to any *F. nucleatum* protein) and to 0.0036 with 14 possibilities for *S. gallolyticus* positivity (eleven individual proteins, positivity to any *S. gallolyticus* protein, double-positivity to Gallo2178 and Gallo2179 and positivity to two or more proteins of the 6-marker panel).

All statistical analyses were carried out using the SAS 9.4 software, all graphical representations using GraphPad Prism 6.

3. Results

3.1. Expression and quality control of *F. nucleatum* and *S. gallolyticus* proteins

3.1.1. Cloning and expression

DNA sequences coding for selected *F. nucleatum* and *S. gallolyticus* proteins (2.2.1) were codon optimized for expression in *E. coli* and resulting genes were synthesized by eurofins genomics (Ebersberg). Synthesized genes were subcloned into the pGEX4T3tag vector resulting in a construct encoding the protein of interest flanked by an N-Terminal GST and a C-terminal tag (GST-X-tag fusion proteins).

Electrocompetent *E. coli* BL21 were transformed with plasmids encoding the recombinant *F. nucleatum* and *S. gallolyticus* GST-X-tag fusion proteins. Expression of proteins was induced with IPTG and bacterial cells were mechanically lysed after 6 hours of induction. Protein concentrations of cleared lysates were determined and ranged from 8.8 to 33.5 mg/ml (Table 17).

Table 17: Total protein concentration of *F. nucleatum* and *S. gallolyticus* fusion protein containing bacterial lysates

Antigen	Concentration (mg/ml)
F. nucleatum	
Fn0131	21.0
Fn0253	11.8
Fn0264	13.2
Fn0387	12.3
Fn1426	11.4
Fn1449	11.4
Fn1526	15.8
Fn1817_1	23.8
Fn1817_2	25.1
Fn1859	10.1
Fn1893	17.2
S. gallolyticus	
Gallo0112A	9.3
Gallo0112B	17.1
Gallo0272	33.5
Gallo0577	18.3
Gallo0748	24.6
Gallo0933	28.1
Gallo1570	26.9
Gallo1675	11.5
Gallo2018	8.8
Gallo2178	11.1
Gallo2179	19.4

3.1.2. Verification of the expression constructs by analytical digestion and PCR with subsequent sequencing

The expression constructs were verified at an intermediate step (by analytical plasmid DNA digestion) and in the final lysate (by PCR and subsequent sequencing of inserted DNA). Before induction, plasmids were isolated from a sample of the transformed bacterial culture and analyzed for the correct insert by analytical restriction digestion. The plasmids were linearized, digested symmetrically at the restriction sites used for cloning (BamHI and SalI) as well as digested asymmetrically using a restriction site inside the insert (Figure 7). Small fragments (< 230 nt) expected in the symmetric digest of Fn1817_1 and Fn1817_2 as well as the asymmetric digest of Fn1817_2 could not be detected with the applied gel electrophoresis. All other digests showed fragments with the expected sizes. Extra bands visible in digestions of Gallo0112A, Gallo1570, Gallo1675 and Gallo2179 most probably resulted from an excess amount of DNA loaded on the gel.

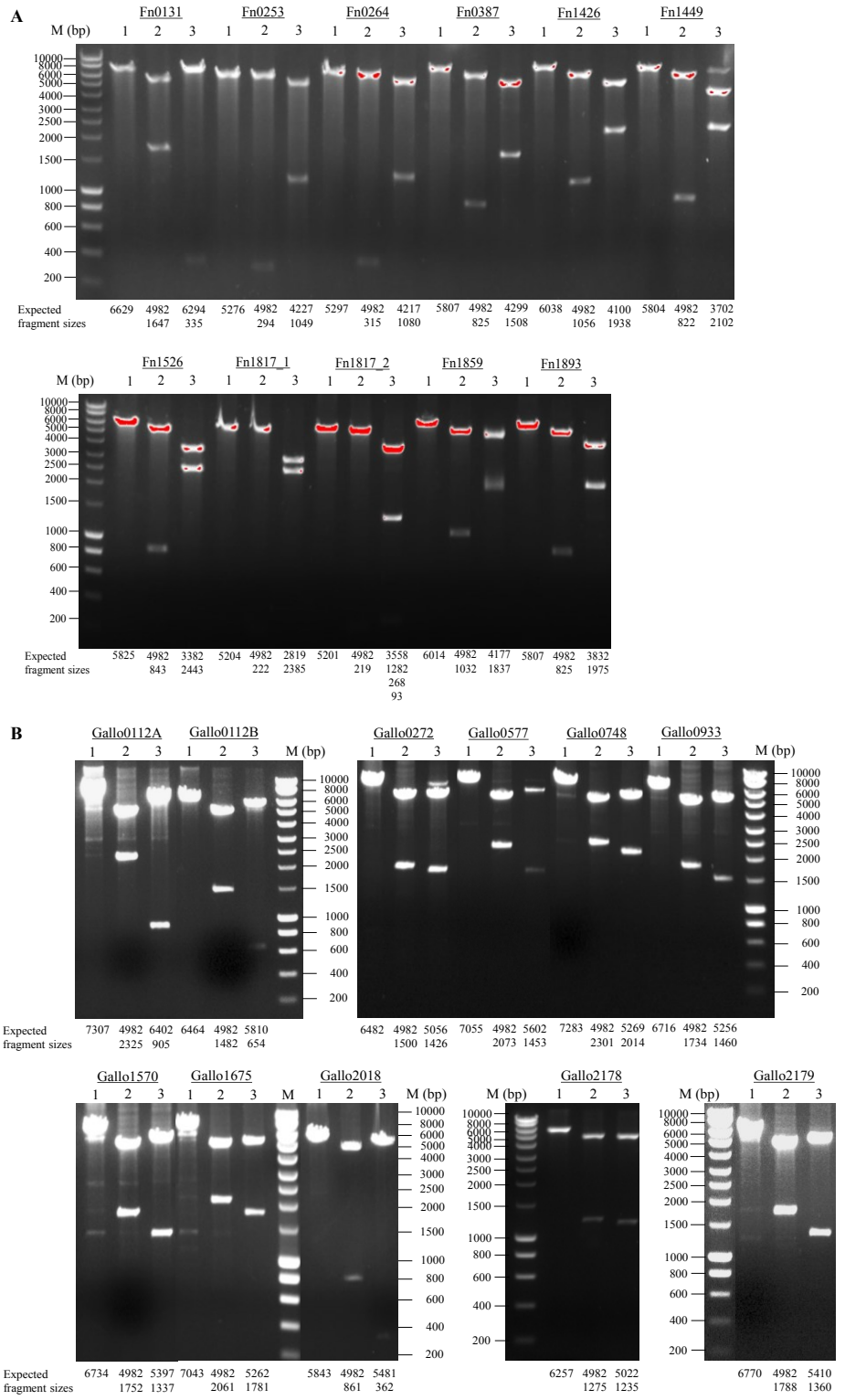


Figure 7: Analytical restriction digests of plasmids isolated from transformed *E. coli* BL21. A) Analytical digests of *F. nucleatum* constructs. B) Analytical digests of *S. gallolyticus* constructs. Plasmids were purified from bacterial cultures before induction and linearized (1), digested symmetrically (2) and digested asymmetrically (3). Fragments were separated by agarose gel electrophoresis. Expected fragment sizes (in base pairs, bp) are given below the respective lanes. M = marker.

Plasmids were isolated from the final lysate and analyzed for the presence of the correct insert via PCR and sequencing of the PCR products. Applied primers were located 5' and 3' outside the insert and attached sequences for T7 and T3 were used for subsequent sequencing of the PCR product. Gel electrophoresis showed all amplicon fragments with the expected sizes (Figure 8). Sequencing of the PCR products, either reversely by T3 primer alone or additionally with forward T7 primer, was successful for all amplicons except Fn0131, Gallo0112A, Gallo0577, Gallo0748, Gallo1570, Gallo1675 and Gallo2179 (Table 18). Amplicons of Gallo0577, Gallo0748, Gallo1570 and Gallo1675 were too long to be sequenced completely, however, partially sequenced fragments clearly showed the presence of the expected insert sequence.

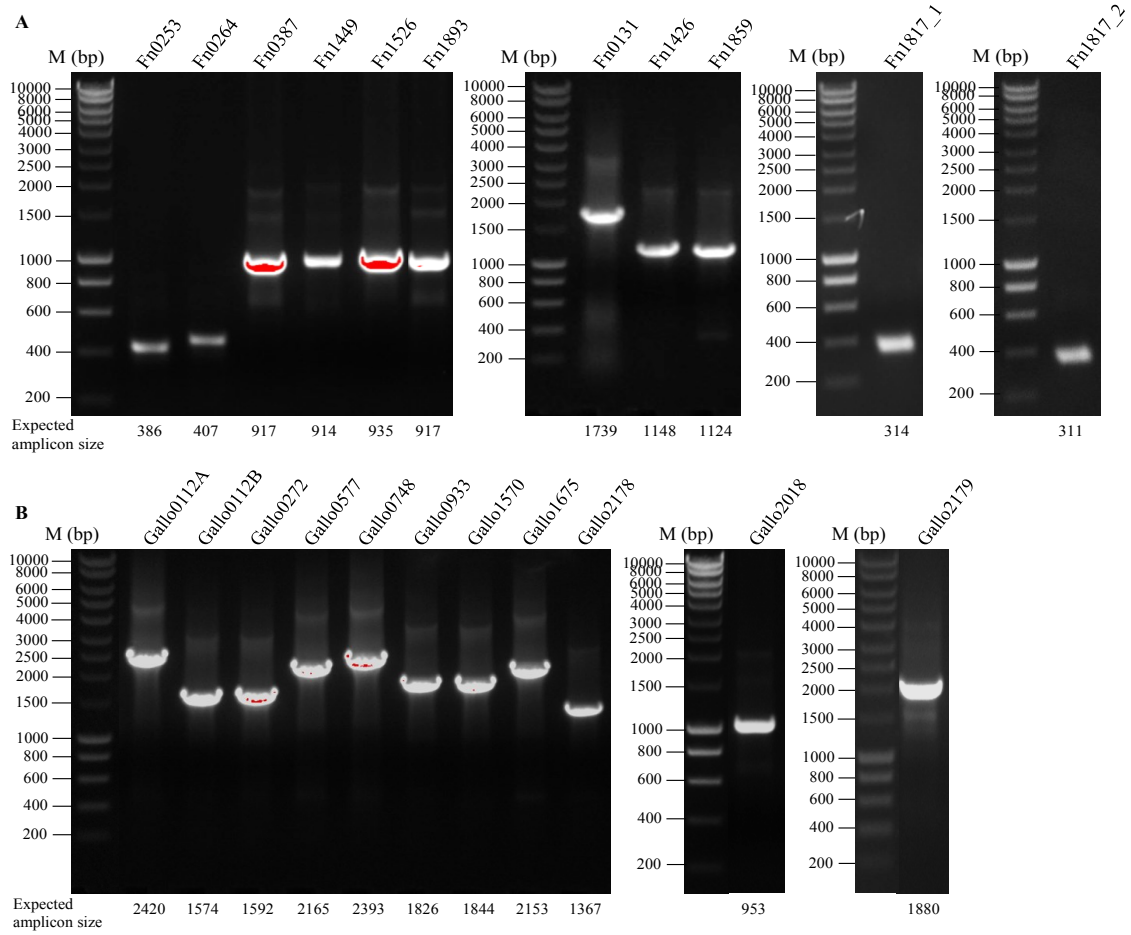


Figure 8: Amplification of insert sequence in plasmids isolated from final lysate. A) PCR of *F. nucleatum* constructs. B) PCR of *S. gallolyticus* constructs. 50 ng of purified plasmid DNA were applied to PCR with primer pGEXs+T3 and pGEXas+T7 and products were separated by agarose gel electrophoresis. Expected amplicon sizes in base pairs (bp) are given below the respective lanes. M = marker.

Table 18: Summary of sequencing results for amplified inserts or expression plasmids in final lysates

Antigen	Sequencing primer	Result
F. nucleatum		
Fn0131	Forward: pGEXfor[2]; reverse: T3[1]	Incomplete sequencing of the insert; nt 1-610 and 742-1641 100% match to ref
Fn0253	Reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Fn0264	Reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Fn0387	Reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Fn1426	Forward: T7; reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Fn1449	Reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Fn1526	Reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Fn1817_1	Reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Fn1817_2	Reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Fn1859	Forward: T7; reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Fn1893	Forward: T7; reverse: T3[1]	Complete sequencing of insert; 100% match to ref
S. gallolyticus		
Gallo0112A	Forward: pGEXfor[2]; reverse: T3[1]	Incomplete sequencing of the insert; nt 1-545 and 1373-2319 100% match to ref
Gallo0112B	Forward: T7; reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Gallo0272	Forward: T7; reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Gallo0577	Forward: T7; reverse: T3[1]	Incomplete sequencing of the insert; nt 1-941 and 1060-2067 100% match to ref
Gallo0748	Forward: T7; reverse: T3[1]	Incomplete sequencing of the insert; nt 1-982 and 1377-2295 100% match to ref
Gallo0933	Forward: T7; reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Gallo1570	Forward: pGEXfor[2]; reverse: T3[1]	Incomplete sequencing of the insert; nt 1-432 and 763-1746 100% match to ref
Gallo1675	Forward: T7; reverse: T3[1]	Incomplete sequencing of the insert; nt 1-1011 and 1079-2055 100% match to ref
Gallo2018	Forward: T7; reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Gallo2178	Forward: T7; reverse: T3[1]	Complete sequencing of insert; 100% match to ref
Gallo2179	Forward: pGEXfor[2]; reverse: pGEXrev[2];	Incomplete sequencing of the insert; nt 1-398 and 1440-1782 100% match to ref

[1]Preceding PCR performed with pGEXs+T7 and pGEXas+T3 primer; [2]Sequencing of plasmid DNA without preceding PCR; nt = nucleotide; ref = reference.

Forward sequencing of Fn0131 and Gallo0112A as well as forward and reverse sequencing of Gallo2179 was not possible with the T3 and T7 primers. Therefore extracted plasmids were directly sent for sequencing without preceding PCR and sequenced using primers located further upstream and downstream of the insert. Although the lower quality of DNA compared to the purified PCR products led to shorter sequence read, the expected insert sequence could be identified. Thus, the presence of the expected insert could be verified for all expression

constructs before induction as well as in the final lysates. Complete alignments are presented in Appendix II.

3.1.3. Verification of full-length expression by Western blot and anti-tag ELISA

Expression of the selected proteins was characterized by Western blot and semi-quantitative anti-tag ELISA.

In Western blot analyses antibodies were applied that can detect either the N-terminal GST or the C-terminal tag-peptide of the GST-X-tag fusion protein sequence. This allowed detecting full-length proteins as well as N-terminal and/or C-terminal fusion protein fragments. Total and cleared lysates were compared to address the solubility of the proteins. Lysate containing GST-tag served as positive control for antibody reactivity as well as indicator of unspecific binding.

In general, with both, anti-GST and anti-tag blot, in addition to full-length protein of the expected size minor bands were also detectable (Figure 9) that mostly migrated faster than the full-length protein indicating protein degradation or premature abortion of translation. The anti-GST antibody is polyclonal and therefore more sensitive than the monoclonal anti-tag antibody. The detection of several epitopes in the N-terminal GST-tag might explain the higher frequency of unexpected bands in the anti-GST blot. Furthermore, there were faint bands appearing in the GST-tag lane with the anti-GST antibody at approximately 20 kD and 50 kD indicating unspecific binding of the antibody to endogenous *E. coli* proteins. The monoclonal anti-tag antibody specific for the C-terminal tag sequence detects only proteins and their N-terminally degraded fragments with full-length expression.

For all *S. gallolyticus* proteins as well as *F. nucleatum* proteins Fn0253, Fn0264, Fn0387, Fn1449, Fn1817_1, Fn1817_2 and Fn1893 proteins of expected sizes were detected with anti-tag antibody in the cleared lysate indicating the presence of full-length proteins of these constructs in the final lysates. Total lysates of Fn0131, Fn1426 and Fn1526 showed signals in the anti-GST and the anti-tag blots, however, signals were weaker or even absent in the cleared lysates, indicating strong insolubility of these proteins. For Fn1859 there is a faint band in the anti-tag blot of the cleared lysate detectable, however, at a smaller size than expected (approximately 56kD instead of 64kD). To exclude that this was due to a gel artifact the Western blots were repeated three times and gave the same result. Since PCR control of

the final lysate verified the presence of the correct plasmid in the lysate this might indicate N-terminal degradation of Fn1859.

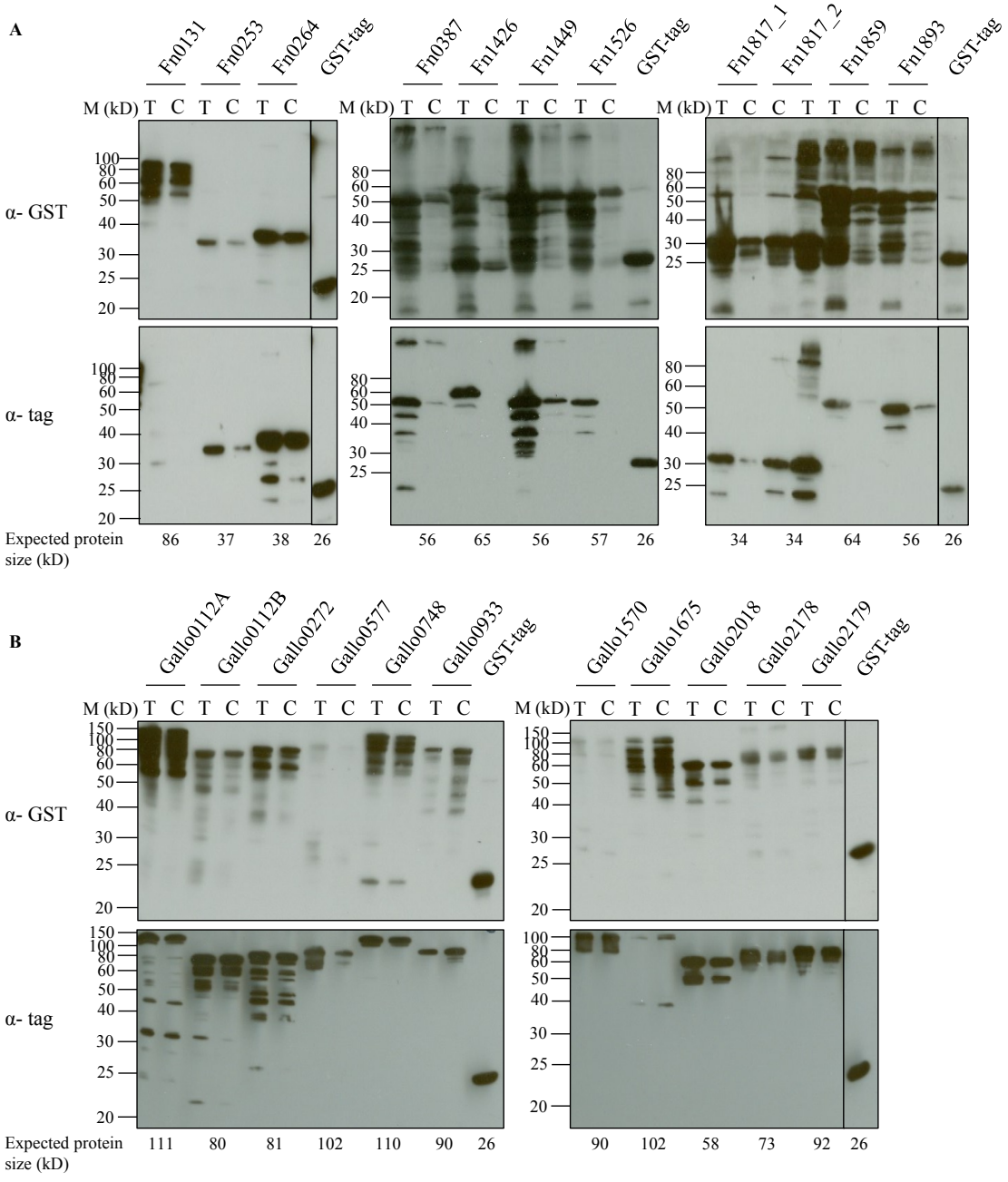


Figure 9: Quality control of protein expression by anti-GST and anti-tag Western blot of (A) *F. nucleatum* and (B) *S. gallolyticus* lysates. *E. coli* BL21 expressing recombinant *F. nucleatum* and *S. gallolyticus* proteins were lysed and 10 µg of total (T) and cleared (C) lysate proteins were analyzed in Western blot. Upper panel anti- (α-) GST antibody, lower panel anti-tag antibody. M = marker.

Anti-tag ELISA allowed relative quantitation of full-length fusion protein compared to the reference GST-tag lysate. Fusion proteins in bacterial lysates were bound to the glutathione-casein coated ELISA plate via the N-terminal GST and full length protein was detected with the anti-tag antibody directed against the C-terminal tag. Bacterial lysates were titrated in a 1:3 dilution series and OD was plotted against the total lysate protein concentration (Figure 10).

Total protein concentrations in GST-X-tag lysates necessary to reach half maximal absorption of GST-tag ($OD50_{(GST-tag)}$) were set in relation to the total protein concentration of the GST-tag lysate at $OD50_{(GST-tag)}$ (Table 19). *S. gallolyticus* proteins Gallo0112A, Gallo0112B, Gallo0933, Gallo2018 and Gallo2178 as well as *F. nucleatum* proteins Fn0253, Fn0264, Fn0387, Fn1449, Fn1817_1, Fn1817_2 and Fn1893 reached $OD50_{(GST-tag)}$ with concentrations less than 30-fold compared to the GST-tag lysate itself. For all other lysates the concentration necessary to reach $OD50_{(GST-tag)}$ ranged from 33- to 143-fold indicating lower relative full-length protein concentration in the lysates. Saturation levels of Gallo0577 were found to be even below $OD50_{(GST-tag)}$. However, all lysates reached saturation at or below 1 mg/ml total lysate protein concentration.

Anti-tag ELISA results were concordant to Western blot results for *F. nucleatum* proteins and indicated in summary lower quantities of full-length expressed protein in the lysates of Fn0131, Fn1426, Fn1526 and Fn1859.

Most of the *S. gallolyticus* proteins showed lower relative full-length protein concentration in anti-tag ELISA, while all proteins had bands at the expected sizes in the anti-tag blot of cleared lysate. A major difference in the two assays is the linearization of the proteins in the Western blot while proteins in anti-tag ELISA should keep their conformation. The anti-tag epitope might be hidden in the folded protein and therefore not accessible for anti-tag antibody in the ELISA leading to lower quantities relative to GST-tag lysate alone.

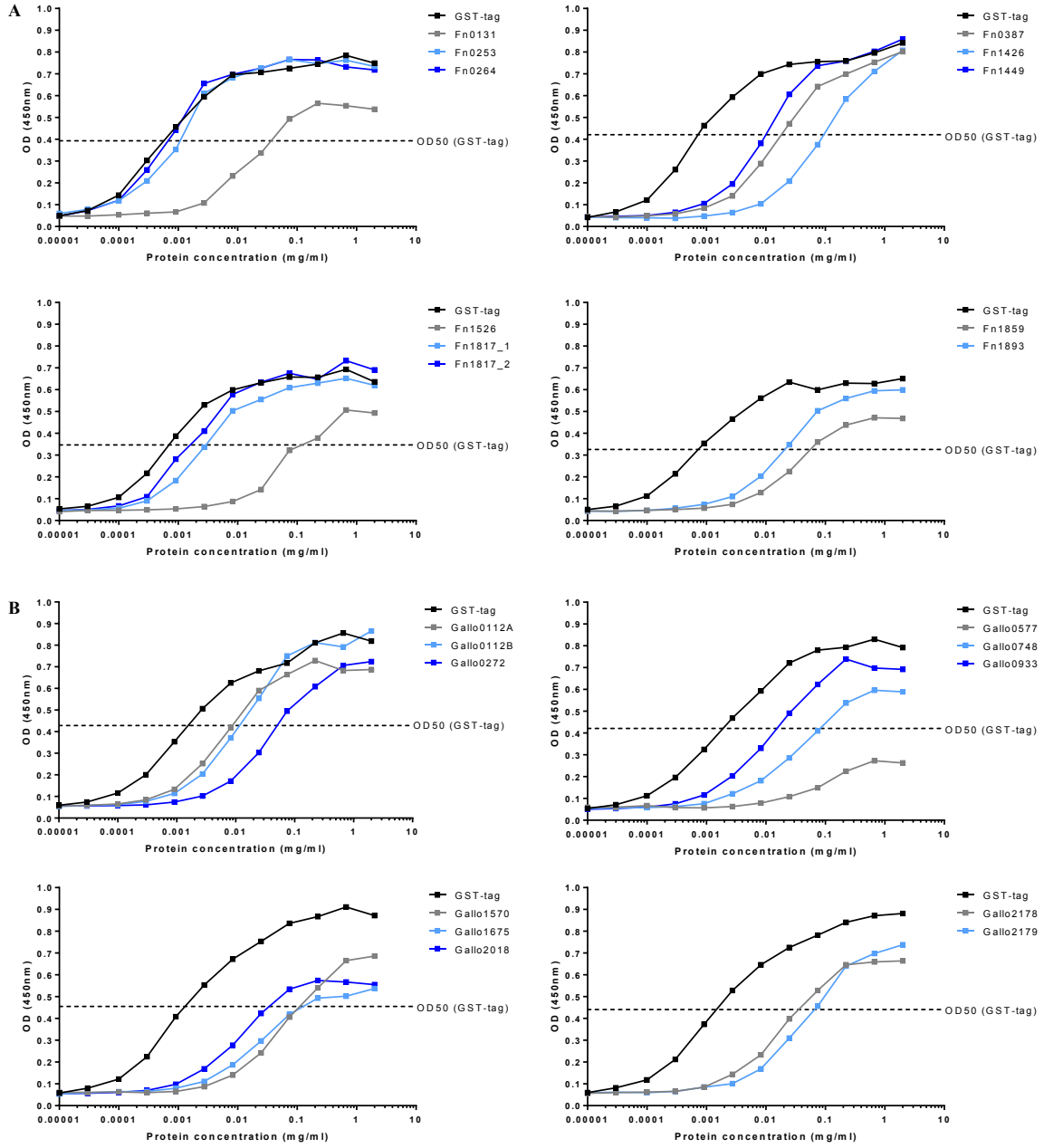


Figure 10: Relative quantitation of full-length protein in cleared lysates by anti-tag ELISA. A) *F. nucleatum* fusion proteins. B) *S. gallolyticus* fusion proteins. Absorption at 450 nm (OD(450nm)) was plotted against total lysate protein concentration. Full-length fusion proteins were detected by mouse anti-tag antibody. The dashed line indicates the half-maximum absorption reached by GST-tag lysate (OD50(GST-tag)).

Adjustment of the protein lysate volume applied in bead-loading to achieve higher concentrations of full-length protein bound to the beads is undesirable since increasing glycerol content would interfere with bead sedimentation and the washing procedure of the

beads. All recombinantly expressed *F. nucleatum* and *S. gallolyticus* proteins, despite lower relative full-length protein concentration for some of the proteins, were applied in multiplex serology at 1 mg/ml total protein concentration. Immunogenic epitopes might be found also on non-full length expressed proteins. However, results from Western blot and anti-tag ELISA should be kept in mind for interpretation of observed findings.

Table 19: Total lysate protein concentrations at half-maximum OD of GST-tag (OD50).

Fusion protein	Total lysate protein concentration at $OD50_{(GST\text{-}tag)}$ (mg/ml)	x-fold to GST-tag lysate concentration at OD50
F. nucleatum		
Fn0131	0.03	60
Fn0253	0.001	2
Fn0264	0.0006	1
Fn0387	0.02	29
Fn1426	0.09	129
Fn1449	0.01	14
Fn1526	0.1	143
Fn1817_1	0.003	4
Fn1817_2	0.0015	2
Fn1859	0.06	86
Fn1893	0.02	29
S. gallolyticus		
Gallo0112A	0.008	5
Gallo0112B	0.009	6
Gallo0272	0.05	33
Gallo0577*	-	-
Gallo0748	0.08	44
Gallo0933	0.02	8
Gallo1570	0.1	67
Gallo1675	0.1	67
Gallo2018	0.03	20
Gallo2178	0.03	20
Gallo2179	0.06	40

*OD of respective lysate did not reach $OD50_{(GST\text{-}tag)}$

3.1.4. Loading of GST-X-tag fusion proteins onto beads

GST-X-tag fusion proteins from bacterial lysates were affinity-purified in situ by binding onto glutathione-casein coated beads. A biotinylated antibody directed against the C-terminal tag was applied as a technical bead-loading control for presence of full-length protein on the beads. Anti-tag antibody bound to all *F. nucleatum* and *S. gallolyticus* proteins as well as GST-tag as a reference was quantified (MFI) in three independent bead-loadings and mean, standard deviation and the coefficient of variation in percent (%CV) were calculated (Table 20). In general, there was little variation observable between the three bead-loadings. The

%CV was below 25% for 20 out of the 23 proteins indicating a good reproducibility and slightly higher but still acceptable for Fn1859, Fn1893 and Gallo2179 (32%, 29.6% and 31.3%, respectively). The mean MFI to none of the GST-X-tag fusion proteins was more than 1.5-fold lower, e.g. 2846 MFI to Fn1526, compared to the GST-tag lysate (3652 MFI). This indicated a comparable amount of full-length proteins loaded onto the beads relative to GST-tag. In addition, mean anti-tag signals from three independent bead-loading reactions correlated only weakly (R^2 = 0.2204) with the total protein lysate concentration necessary to reach $OD50_{(GST\text{-}tag)}$ in anti-tag ELISA (3.1.3) (Figure 11). Thus, although Western blot and anti-tag ELISA results showed lower concentrations of full-length protein for some of the lysates, sufficient amounts of full-length protein loaded onto the beads were detected probably resulting from an excess of protein applied in the bead-loading. In conclusion, even lowly concentrated full-length fusion proteins were at saturating levels when beads were loaded with GST-X-tag fusion proteins at 1 mg/ml total lysate concentration.

Table 20: Anti-tag signal (MFI) on antigen-loaded beads in three independent bead-loading reactions

Antigen	MFI (anti-tag)					
	Load 1	Load 2	Load 3	Mean	Standard deviation	%CV
F. nucleatum						
Fn0131	4881	4184	2951	4005	977	24.40
Fn0253	5754	5349	4545	5216	615	11.80
Fn0264	6163	5188	4870	5407	674	12.46
Fn0387	4919	3662	3698	4093	715	17.48
Fn1426	5475	5945	4444	5288	768	14.52
Fn1449	4027	5182	3774	4328	751	17.35
Fn1526	3005	2831	2701	2846	153	5.36
Fn1817_1	3401	3747	3899	3682	255	6.93
Fn1817_2	5132	4967	4428	4842	368	7.60
Fn1859	3337	4511	2340	3396	1087	32.00
Fn1893	4785	5544	2986	4438	1314	29.60
S. gallolyticus						
Gallo0112A	3826	3790	3761	3792	33	0.86
Gallo0112B	4091	2945	4135	3724	675	18.12
Gallo0272	3306	2629	3652	3196	520	16.28
Gallo0577	3539	3682	3195	3472	250	7.21
Gallo0748	3567	3037	3464	3356	281	8.37
Gallo0933	3270	3598	3548	3472	177	5.09
Gallo1570	3839	3539	4046	3808	255	6.69
Gallo1675	2802	2926	3018	2915	108	3.72
Gallo2018	4113	3174	3665	3651	470	12.87
Gallo2178	3526	3936	3637	3700	212	5.73
Gallo2179	2424	3200	4513	3379	1056	31.26
GST-tag	3283	4162	3512	3652	456	12.49

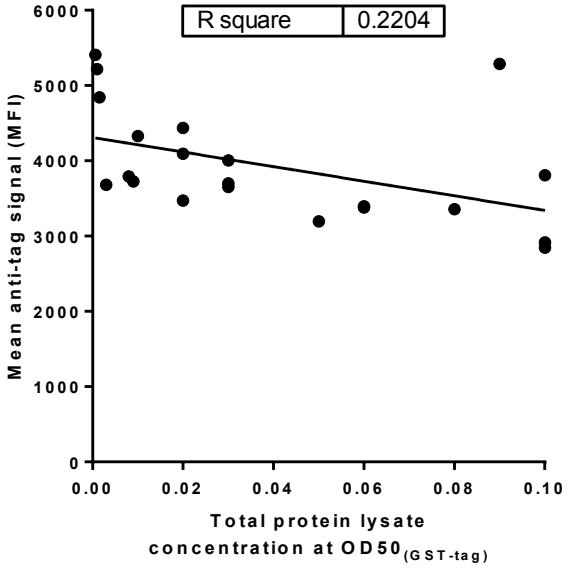


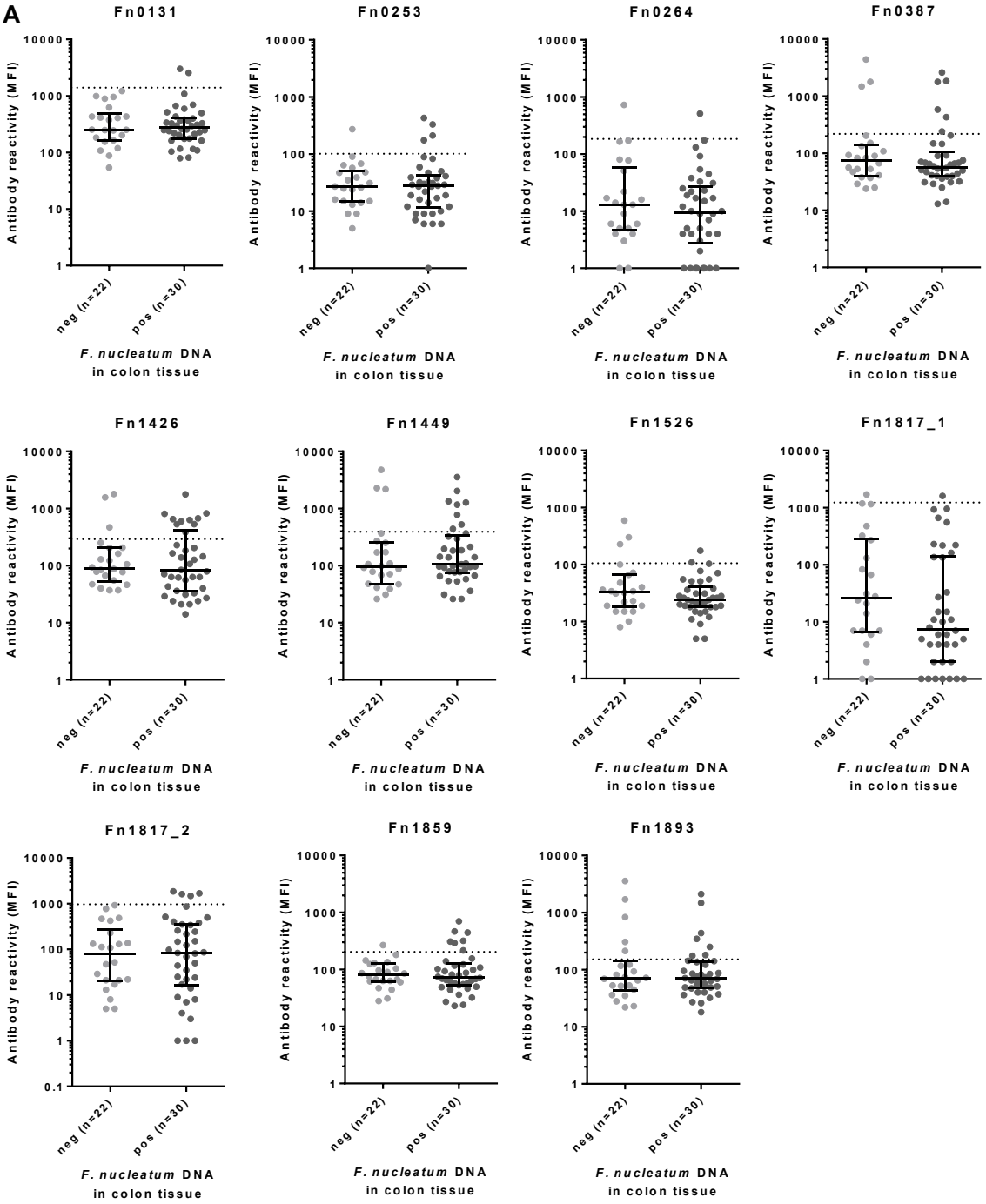
Figure 11: Comparison of mean anti-tag signal (MFI) from three independent bead-loading reactions to total protein lysate concentration at $OD50_{(GST\text{-}tag)}$ in anti-tag ELISA (3.1.3).

3.2. Validation of *F. nucleatum* and *S. gallolyticus* multiplex serology

There is no serological gold standard test available for *F. nucleatum* and *S. gallolyticus* antibody detection making a proper antibody-based validation of the multiplex assays difficult. However, to validate the assays by alternative means I explored whether positive serological results are correlated with presence of *F. nucleatum* DNA in normal or tumorous colonic tissue, with experimental *S. gallolyticus* infection in mice, with presence of *S. gallolyticus* DNA in feces of cattle and with *S. gallolyticus* bacteremia in human individuals.

3.2.1. Antibody responses to *F. nucleatum* in comparison to *F. nucleatum* DNA in colon tissue

Tumor and adjacent normal colon tissues of participants in the Irish CRC case-control study had been analyzed with qPCR for *F. nucleatum* DNA [44]. Corresponding sera (n=52) and data on *F. nucleatum* DNA status were kindly provided by David Hughes. Sera were analyzed with *F. nucleatum* multiplex serology in a 1:100 and 1:1000 dilution. For each individual *F. nucleatum* antigen MFI values were plotted and compared between DNA-negative and -positive individuals (Figure 12).


A
Fn0131
Fn0253
Fn0264
Fn0387
Fn1426
Fn1449
Fn1526
Fn1817_1
Fn1817_2
Fn1859
Fn1893
Antibody reactivity (MFI)
neg (n=22)
pos (n=30)
F. nucleatum DNA in colon tissue

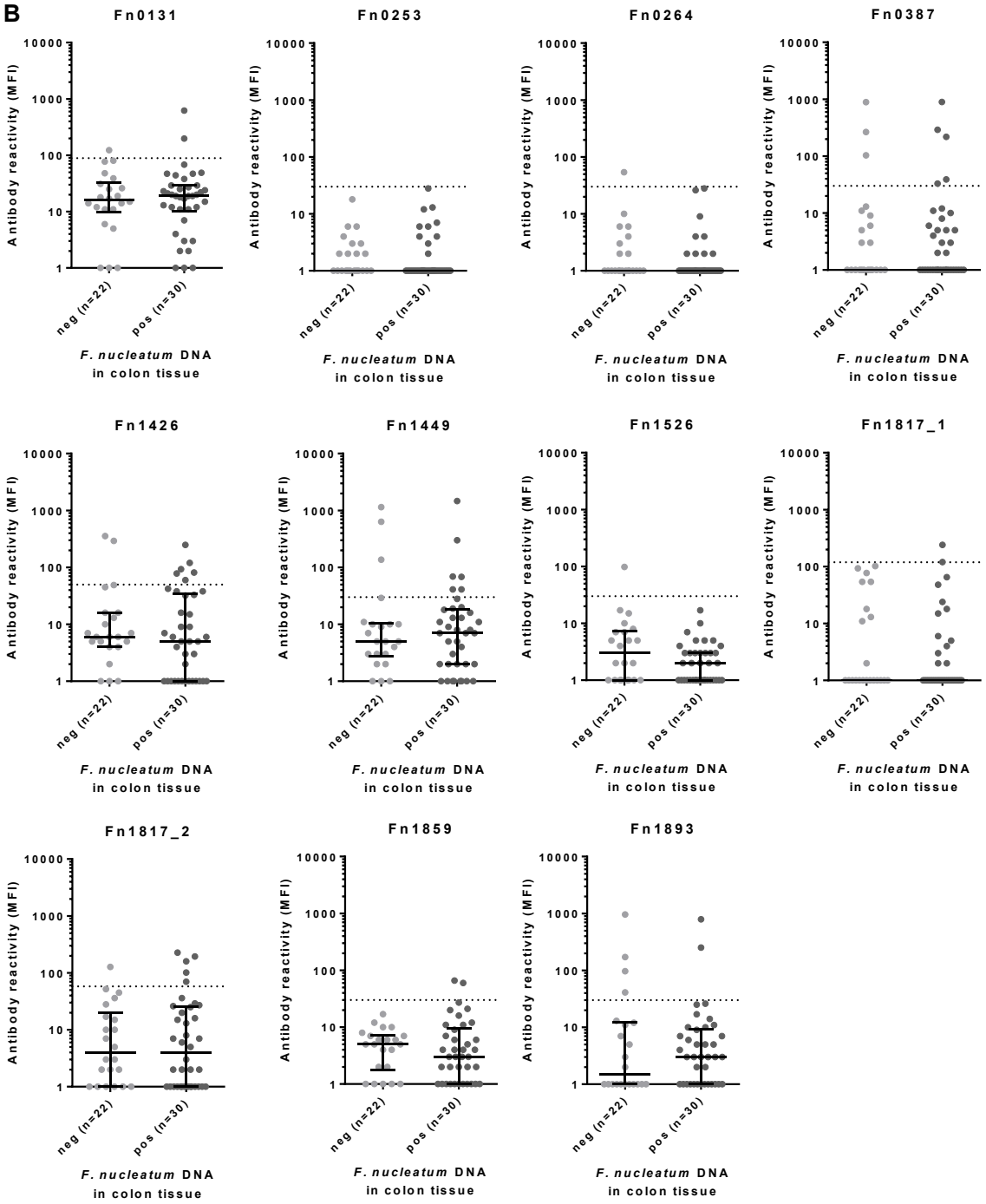


Figure 12: Antibody responses to individual *F. nucleatum* proteins in colon tissue DNA-negative and –positive individuals. Sera of 52 individuals (kindly provided by David Hughes) were analyzed in *F. nucleatum* multiplex serology in A) 1:100 and B) 1:1000 dilution. Antibody reactivity to individual *F. nucleatum* proteins is given in MFI. Horizontal bars represent the 25th, 50th and 75th percentile. The dashed line indicates antigen-specific cut-offs derived from mean MFI in controls + 3 * standard deviation excluding outliers. neg = DNA-negative, pos = DNA-positive.

Overall, responses to all *F. nucleatum* proteins were low in both reference groups. The maximum MFI value achieved in the 1:100 dilution was with Fn1449 (4763 MFI). However, such high MFI values were rare. The median MFI to any of the antigens in the two groups ranged from 11 (Fn1817_1) to 257 MFI (Fn0131).

In the 1:1000 dilution the median antibody responses were approximately 10-fold lower and ranged from 1 MFI (Fn0253, Fn0264, Fn0387 and Fn1817_1) to 20 MFI (Fn0131). Maximum MFI values were approximately 3-fold lower and with 1464 MFI highest again to Fn1449. There was no significant difference between *F. nucleatum* DNA-negative and -positive individuals in median antibody response to any of the eleven *F. nucleatum* antigens given by Wilcoxon rank-sum test, neither in the 1:100 nor in the 1:1000 dilution.

Application of a cut-off derived from the mean MFI plus three times the standard deviation among controls under exclusion of outliers was used to analyze whether positivity to individual and multiple proteins was more frequent in *F. nucleatum* DNA- positive compared to –negative individuals (Table 21). Positivity to proteins Fn0131, Fn0387, Fn1426, Fn1449, Fn1817_1, Fn1817_2 as well as positivity to any and to more than one of the proteins was consistently more frequent in both dilutions in DNA-positives compared to DNA-negatives. However, due to the small sample size there was no significant difference.

Table 21: Percentage antibody-positives to individual and multiple *F. nucleatum* proteins in colon tissue DNA-negative and –positive individuals

Antigen (-combination)	n (%) antibody-positives[1] in serum dilution 1:100 DNA-negative[2]	DNA-positive[2]	p-value[3]	n (%) antibody-positives[1] in serum dilution 1:1000 DNA-negative[2]	DNA-positive[2]	p-value[3]
Fn0131	0 (0)	2 (7)	0.217	1 (5)	2 (9)	0.746
Fn0253	1 (5)	3 (10)	0.466	0 (0)	0 (0)	1.000
Fn0264	1 (5)	1 (3)	0.822	1 (5)	0 (0)	0.238
Fn0387	3 (14)	6 (20)	0.549	3 (14)	5 (23)	0.764
Fn1426	3 (14)	8 (27)	0.256	2 (9)	4 (18)	0.695
Fn1449	3 (14)	7 (23)	0.381	3 (14)	6 (27)	0.549
Fn1526	3 (14)	1 (3)	0.168	1 (5)	0 (0)	0.238
Fn1859	1 (5)	1 (3)	0.822	0 (0)	2 (9)	0.217
Fn1817_1	0 (0)	4 (13)	0.075	1 (5)	4 (18)	0.081
Fn1817_2	1 (5)	6 (20)	0.107	0 (0)	2 (9)	0.217
Fn1893	5 (23)	6 (20)	0.812	4 (18)	2 (9)	0.199
= 1 protein	11 (50)	21 (70)	0.143	8 (36)	16 (53)	0.225
> 1 protein	5 (23)	13 (43)	0.123	4 (18)	7 (23)	0.653
> 2 proteins	4 (18)	7 (23)	0.653	4 (18)	3 (10)	0.393
> 3 proteins	1 (5)	3 (10)	0.465	0 (0)	1 (3)	0.387

[1]cut-off was determined by the mean MFI + 3 * standard deviation in controls under exclusion of outliers; [2]DNA status in colon tissue was kindly provided by David Hughes [44], n(DNA-negative) = 22, n(DNA-positive) = 30; [3]Pearson's Chi-square test

In summary, there was no correlation between *F. nucleatum* DNA-positivity in colonic tissue and continuous antibody responses in *F. nucleatum* multiplex serology. However, application of a cut-off found a trend for a higher number of antibody-positives among DNA-positives as compared to DNA-negative individuals. This finding was independent of the serum dilution.

3.2.2. *S. gallolyticus* multiplex serology in sera from experimentally infected mice

Serum samples of three mice, two inoculated by gavage with $4.5*10^8$ cells of *S. gallolyticus* UCN34 (mice 91 and 98) and the third (mouse 68) with PBS, were kindly provided by Annemarie Boleij. Starting two weeks post inoculation blood was taken every week up to week seven and serum samples were analyzed by *S. gallolyticus* multiplex serology in 1:100 serum dilution (Figure 13). The PBS-inoculated mouse constantly showed antibody responses below 100 MFI for all antigens except for a slight increase with Gallo0272 to up to 316 MFI in week 6. At this time point the mouse had to be sacrificed due to an unidentified illness. Mouse 91 showed seroconversion to two *S. gallolyticus* antigens. Beginning in week 3 antibody response to Gallo2179 increased, and reached 4335 MFI at week 4 and declined to 2244 MFI in week 6. Antibody response to Gallo0577 also increased in week 3 reaching a maximum of 527 MFI in week 5. Mouse 98 also showed seroconversion but only weakly and to a single and different protein (Gallo0272) reaching 814 MFI in week 7.

These results corresponded to data obtained with a newly developed PCR (Indra-Jasmin Gierse and Daniela Höfler, unpublished data): *S. gallolyticus* DNA was found in feces of mice sampled at week 7 post inoculation in mouse 91 but not 68 and 98. Similar to that Annemarie Boleij measured colonization of the bacteria and found mouse 91 still colonized with *S. gallolyticus* at week 7 while mouse 98 lost colonization with *S. gallolyticus* already at week 3 post inoculation.

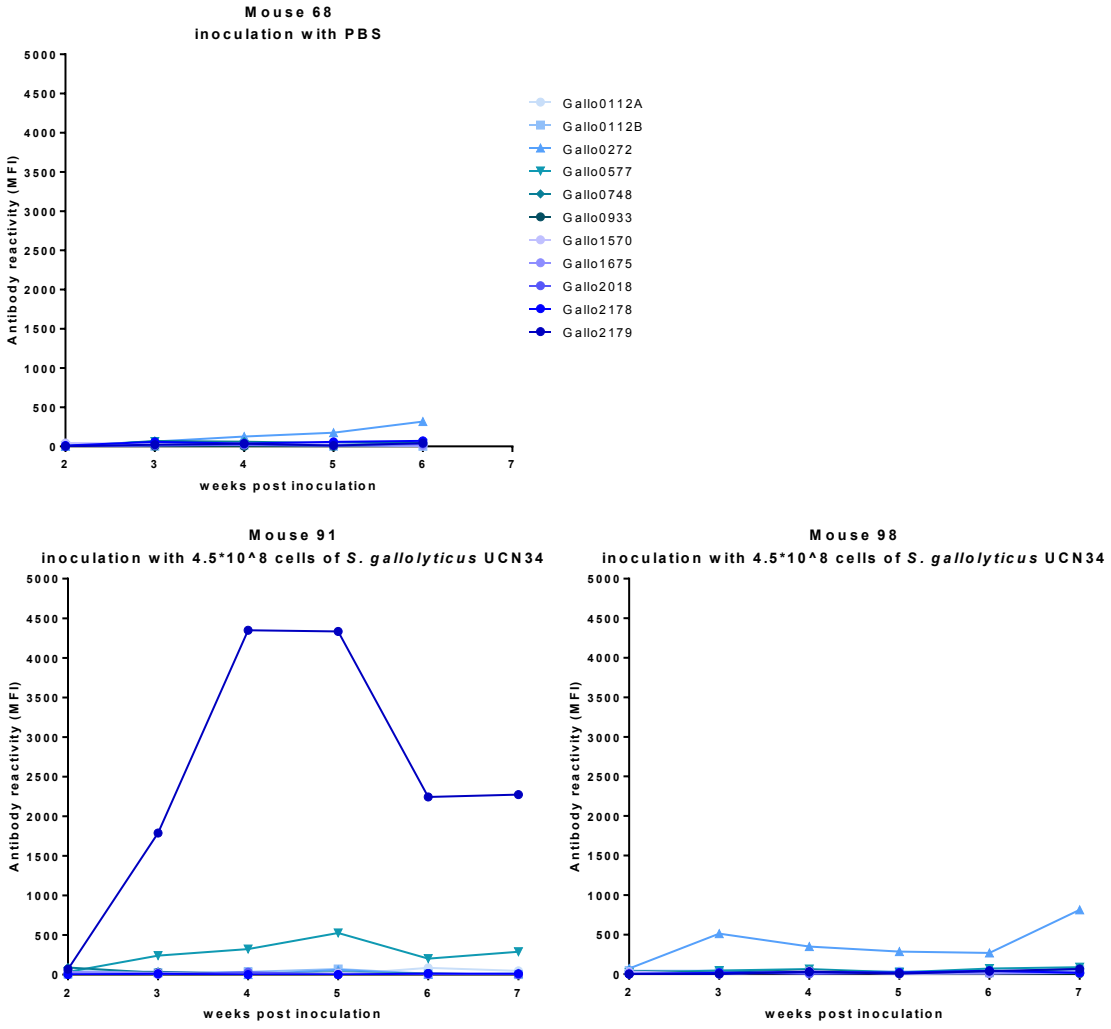


Figure 13: Antibody responses to *S. gallolyticus* proteins in experimentally infected mice. Antibody reactivity (MFI) against individual *S. gallolyticus* proteins is plotted over time. Two mice had been inoculated by gavage with *S. gallolyticus* UCN34 and one negative control mouse with PBS. Blood samples were taken weekly after inoculation up to week 7. Sera were analyzed in a 1:100 dilution. Sera were provided by Annemarie Boleij.

3.2.3. *S. gallolyticus* multiplex serology in bovine sera

S. gallolyticus is highly prevalent in the rumen of cattle [62]. Serum and fecal samples had been collected from 51 dairy cows (kindly provided by Indra-Jasmin Gierse) and analyzed for *S. gallolyticus* DNA in fecal samples by *S. gallolyticus* PCR (Indra-Jasmin Gierse et al., unpublished data). Twenty-eight of the 51 cows were positive for *S. gallolyticus* DNA in their feces, 13 were negative and 10 had invalid PCR results, since neither *S. gallolyticus* DNA nor DNA of a bovine housekeeping gene could be identified. Antibody responses to all eleven

S. gallolyticus proteins in *S. gallolyticus* fecal-DNA-negative cows were compared to those of –positive cows (Figure 14). Overall, antibody responses to *S. gallolyticus* proteins were low. At 1:100 dilution the majority of serum samples showed an antibody response higher than 1000 MFI only with Gallo0748. With Gallo1570, Gallo2018 and Gallo2178 only few serum samples exceeded 1000 MFI. Lowest median MFI were achieved with Gallo0112B, Gallo1675 and Gallo0933 (14, 26 and 19 MFI, respectively) followed by Gallo0577 and Gallo0272 (38 and 40 MFI). For the remaining antigens the median MFI ranged around 100 MFI with the exception of Gallo0748 that achieved a median MFI of above 1000 MFI in both groups. There was no significant difference between *S. gallolyticus* DNA-negative and -positive individuals in median antibody response to any of the eleven *S. gallolyticus* antigens given by Wilcoxon rank-sum test.

Application of a cut-off derived from controls as the mean MFI plus three times the standard deviation under exclusion of positive outliers allowed analyzing further whether positivity to individual and multiple proteins differed between the two groups. Indeed positivity to Gallo0112A, Gallo0272, Gallo0933, Gallo1570, Gallo1675, Gallo2018, Gallo2179, any of the proteins and multiple proteins was more frequent among DNA-positive compared to -negative cows. Positivity to any of the eleven proteins among DNA-negative cows (23%) was significantly less frequent than among DNA-positive cows (71%). None of the DNA-negative cows was positive to more than two proteins compared to 18% of the DNA-positive cows indicating that correlation of the proteins strengthens the specificity of *S. gallolyticus* antibody detection in comparison to PCR.

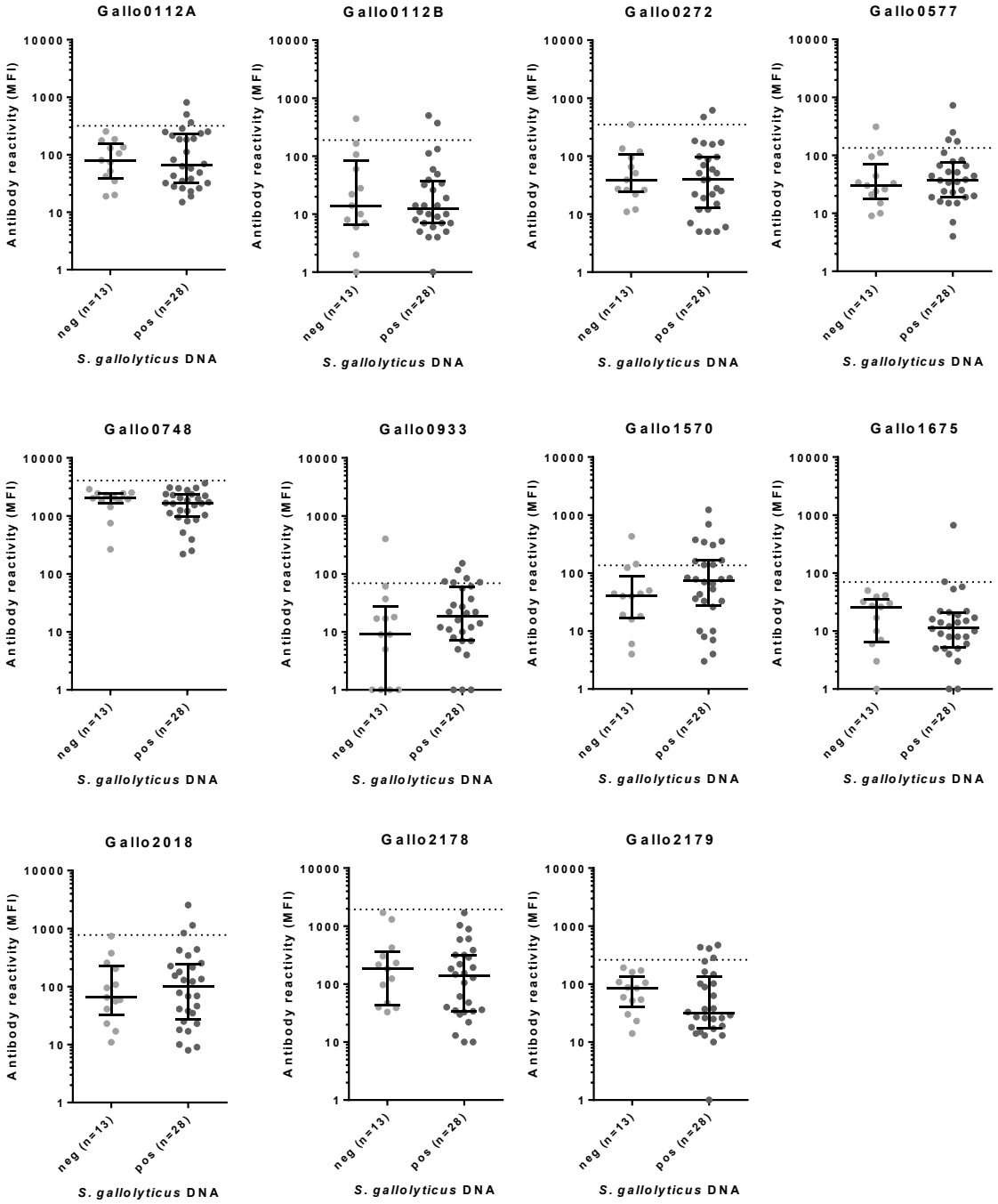


Figure 14: Antibody responses to individual *S. gallolyticus* proteins in cows. Sera of 41 cows with known fecal *S. gallolyticus* DNA status (DNA data and serum samples provided by Indra-Jasmin Gierse) were analyzed with *S. gallolyticus* multiplex serology in 1:100 dilution. Antibody reactivity to individual *S. gallolyticus* proteins is given in MFI. Horizontal bars represent the 25th, 50th and 75th percentile. The dashed line indicates antigen-specific cut-offs derived from mean MFI in controls + 3 * standard deviation excluding positive outliers. neg = DNA-negative, pos = DNA-positive.

Table 22: Percentage antibody-positives to individual and multiple *S. gallolyticus* proteins in fecal DNA-negative and -positive cows

Antigen (-combination)	n (%) antibody-positives[1] in serum dilution 1:100 DNA-negative[2]	DNA-positive[2]	p-value[3]
Gallo0112A	0 (0)	3 (11)	0.220
Gallo0112B	1 (8)	2 (7)	0.950
Gallo0272	0 (0)	2 (7)	0.323
Gallo0577	1 (8)	14 (4)	0.548
Gallo0748	0 (0)	0 (0)	1.000
Gallo0933	1 (8)	6 (21)	0.276
Gallo1570	2 (15)	10 (36)	0.183
Gallo1675	0 (0)	2 (7)	0.323
Gallo2018	0 (0)	3 (11)	0.220
Gallo2178	0 (0)	0 (0)	1.000
Gallo2179	0 (0)	4 (14)	0.151
= 1 protein	3 (23)	20 (71)	**0.004**
> 1 protein	2 (15)	10 (36)	0.183
> 2 proteins	0 (0)	5 (18)	0.104
> 3 proteins	0 (0)	1 (4)	0.490

[1]cut-off was determined by the mean MFI + 3* standard deviation in controls under exclusion of outliers; [2]DNA status in feces was kindly provided by Indra-Jasmin Gierse, n(DNA-negative) = 13, n(DNA-positive) = 28; [3]Pearson's Chi-square test, significant associations are marked in bold font.

3.2.4. *S. gallolyticus* multiplex serology in humans with *S. gallolyticus* bacteremia

In a previous study we directly had coupled four pilus proteins, including Gallo2178 and Gallo2179, as recombinantly expressed and affinity-purified his-tagged proteins to Luminex beads [81]. In the attempt to validate these proteins we had analyzed in total six sera, three from healthy controls (HC 1-3) and three from patients diagnosed with *S. gallolyticus* bacteremia (SGG-pos 1-3) for antibody responses to the his-tagged pilus proteins (both, proteins and sera had been provided by Harold Tjalsma) (Figure 15B). Antibody responses to his-tagged Gallo2178 had been higher in all positive serum controls compared to healthy controls (SGG-pos1: 15039, SGG-pos2: 12297 and Sgg-pos3: 1907 MFI versus HC: 188 and 1 MFI in 1:100 dilution, respectively). Antibody responses to his-tagged Gallo2179 were higher in two out of three positive serum controls (SGG-pos1: 8413, SGG-pos2:8053 vs HC1: 889 MFI in 1:100 dilution).

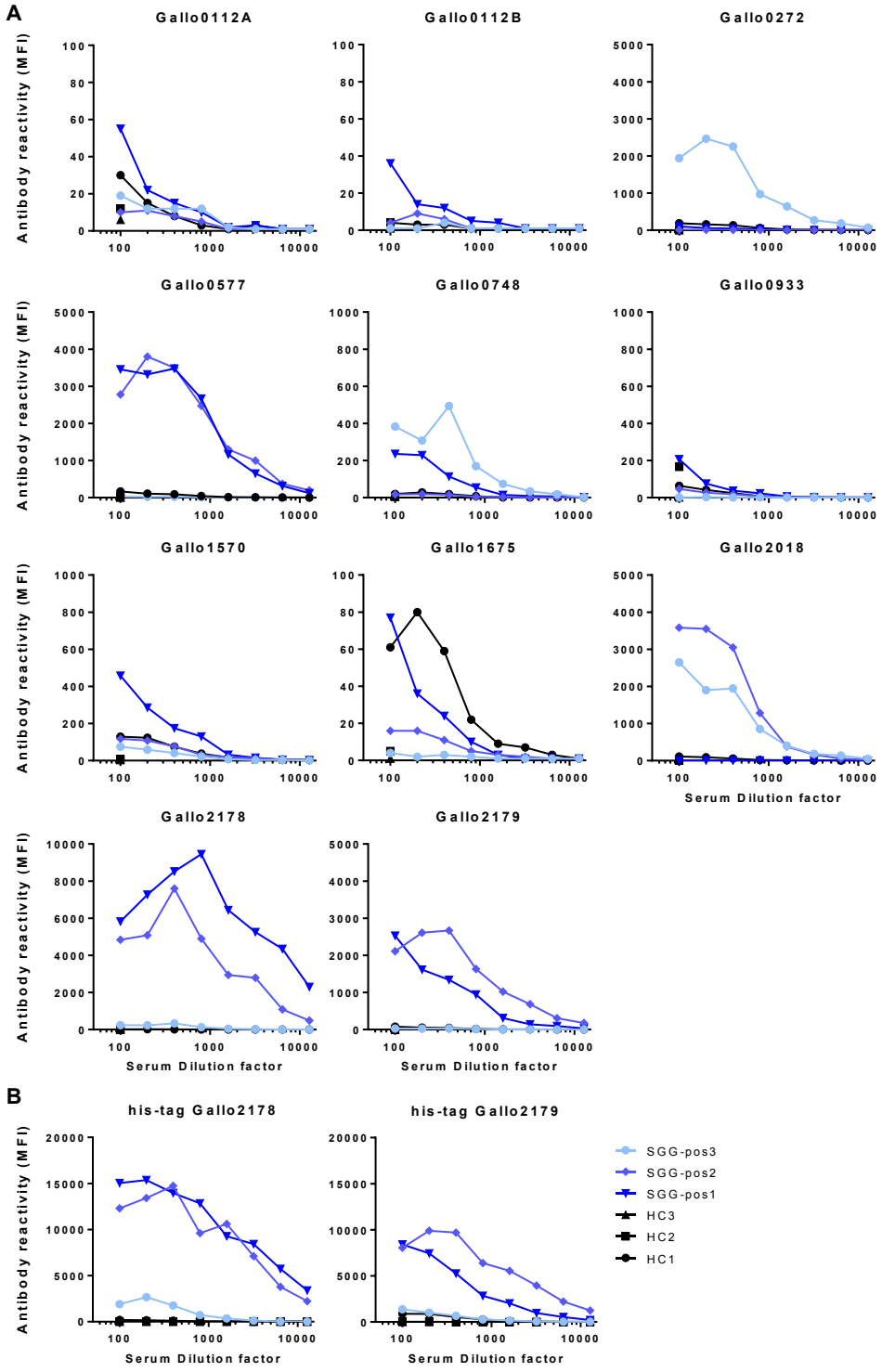


Figure 15: Antibody responses to *S. gallolyticus* proteins in individuals with *S. gallolyticus* bacteremia and healthy controls. A) Multiplex serology with *S. gallolyticus* GST-X-tag fusion proteins and B) with his-tagged Gallo2178 and Gallo2179 directly coupled to beads [81]. Antibody reactivity (MFI) to individual *S. gallolyticus* proteins was measured in a 1:2 dilution series of serum samples from three *S. gallolyticus* bacteremia patients (SGG-pos 1-3) and healthy controls (HC 1-3). Serum samples and his-tagged Gallo2178 and Gallo2179 had been kindly provided by Harold Tjalsma.

These serum samples were analyzed with the newly developed *S. gallolyticus* proteins in a similar multiplex serology setup with the exception that two of the healthy controls were analyzed at 1:100 serum dilution only (Figure 15A). First of all, I was able to replicate results from his-tagged Gallo2178 and Gallo2179 with the GST-X-tag fusion proteins. SGG-pos1 (5821 MFI) and SGG-pos2 (4835 MFI) reacted strongly with GST-tagged Gallo2178 followed by SGG-pos3 (242 MFI) still having higher responses than the highest healthy control (16 MFI at 1:100 dilution). The increase in MFI at higher dilutions as seen with Gallo2178 in dilution 1:800 (9456 MFI) compared to dilution 1:100 (5821 MFI) can be explained by a so-called "hook-effect": Saturation with serum antibody in lower dilution blocks binding of secondary antibodies/reagents, which is enabled in higher serum dilutions by the liberation of more potential binding sites. SGG-pos1 (2529 MFI) and SGG-pos2 (2111) also reacted strongly to GST-tagged Gallo2179, whereas SGG-pos3 (27 MFI) did not exceed the antibody response of HC1 (78 MFI).

Overall, in these human serum samples lowest antibody responses were present with Gallo0112A (6-55 MFI in 1:100 dilution), Gallo0112B (1-36 MFI), Gallo1675 (1-77 MFI) and Gallo0933 (1-207 MFI) irrespective of serum type. There was also no difference in response to these proteins between positive serum controls and healthy controls. Positive serum controls had higher MFI compared to healthy controls to the remaining antigens, however, in different patterns (Table 23):

SGG-pos1 to Gallo0577, Gallo0748, Gallo1570, Gallo2178 and Gallo2179;

SGG-pos2 to Gallo0577, Gallo2018, Gallo2178 and Gallo2179;

SGG-pos3 to Gallo0272, Gallo0748, Gallo2018 and Gallo2178.

Table 23: Antibody responses (MFI) to *S. gallolyticus* proteins in individuals with *S. gallolyticus* bacteremia and healthy controls analyzed at 1:100 serum dilution.

Antigen	HC-1	HC-2	HC-3	SGG-pos1	SGG-pos2	SGG-pos3
Gallo0112A	30	12	6	55	10	19
Gallo0112B	4	4	1	36	4	1
Gallo0272	188	5	1	10	11	**1943**
Gallo0577	166	11	1	**3458**	**2780**	27
Gallo0748	20	6	5	**236**	17	**383**
Gallo0933	64	166	1	207	47	1
Gallo1570	129	7	3	**458**	117	75
Gallo1675	61	5	1	77	16	4
Gallo2018	114	13	3	3	**3584**	**2648**
Gallo2178	16	3	1	**5821**	**4835**	**242**
Gallo2179	78	3	1	**2529**	**2111**	27

MFI > 100 and at least 2-fold higher than in HC are marked in bold font; SGG-pos = *S. gallolyticus* bacteremia patients 1 to 3; HC = Healthy controls 1 to 3.

In summary, I observed similar patterns of antibody responses to *S. gallolyticus* in three different hosts. One out of two experimentally infected mice was successfully colonized with *S. gallolyticus* for 7 weeks and showed strong antibody responses to Gallo2179 and Gallo0577 only. The second mouse inoculated with *S. gallolyticus* raised antibody responses only to Gallo0272. The PBS-inoculated mouse for unidentified reasons also developed two-fold lower antibody responses to Gallo0272. In bovine serum samples, application of a cut-off derived from the mean plus three times the standard deviation of MFI in controls showed that fecal *S. gallolyticus* DNA-positive cows were more frequently antibody-positive than DNA–negative cows also to Gallo0272 and Gallo2179 and additionally to Gallo0112A, Gallo0933, Gallo1570, Gallo1675, Gallo2018, any of the proteins and multiple proteins. Due to the small sample sizes none of the differences was significant except for positivity to any of the proteins. Gallo0272 and Gallo2179 were consistently distinguishing defined negatives from positives among all three types of hosts, also humans. Additionally, in human samples, which will be the type of samples analyzed for an association with CRC, I observed that Gallo2178 is recognized specifically by sera of patients diagnosed with *S. gallolyticus* bacteremia together with Gallo0272 and Gallo2179 but also Gallo0577, Gallo0748, Gallo1570 and Gallo2018. Also these proteins mostly overlap with the proteins that best distinguished *S. gallolyticus* DNA-positive from -negative cows as well as experimentally infected from non-infected mice (Table 24). In addition, the finding that patients with *S. gallolyticus* bacteremia raised antibody responses to several proteins simultaneously was concordant to the observation in bovine sera, where positivity to several proteins strengthened the specificity in comparison to DNA.

Table 24: Proteins distinguishing *S. gallolyticus* infected from non-infected individuals in three different host species

Antigen	Murine[1]	Bovine[2]	Human[1]
Gallo0112A		x	
Gallo0112B			
Gallo0272	x	x	x
Gallo0577	x		x
Gallo0748			x
Gallo0933		x	
Gallo1570		x	x
Gallo1675		x	
Gallo2018		x	x
Gallo2178			x
Gallo2179	x	x	x

[1]x = MFI at least 2-fold higher in defined positives compared to negatives; [2]x = % antibody-positives at least 2-fold higher in DNA-negatives compared to -positives

An ideal agreement of different assays, e.g. PCR and serology, would imply a high sensitivity and high specificity. However, here both were only moderate. Serology is a cumulative measure of systemic past and present infections, whereas DNA detection only measures present infection at the sampled site. I could not exclude that DNA-negative cows never had a *S. gallolyticus* infection neither could I exclude that *S. gallolyticus* DNA is under the detection limit of the PCR nor that seroconversion could have occurred in the past and/or from infection of sites not resulting in *S. gallolyticus* DNA-positive feces. The number of artificially infected mice and human individuals with known *S. gallolyticus* infection was not sufficient to define robust cut-offs for individual proteins neither to define an algorithm for *S. gallolyticus* overall positivity.

Thus, also the *S. gallolyticus* multiplex serology could not be completely validated. However, the different attempts indicated that *S. gallolyticus* multiplex serology developed here is able to measure antibody responses to *S. gallolyticus* but sensitivity and specificity remain to be determined further.

3.3. Search for serological associations of *F. nucleatum* with CRC by multiplex serology in a retrospective case-control study and prospective nested case-control studies

3.3.1. Cut-off definition

Since infection-based validation of *F. nucleatum* multiplex serology by DNA-status in the intestine failed I had to arbitrarily define cut-offs for *F. nucleatum* antibody positivity. The BliTz study included samples from colonoscopy negative individuals that served as control group. The distribution of antibody responses in these CRC- and precursor lesion- free study participants was overall skewed towards low MFI for all eleven *F. nucleatum* antigens (Figure 16). Fn1426 showed strongest responses but still with only 12% of sera reaching > 500 MFI, while with Fn1526 only 1% reached this antibody level. For each *F. nucleatum* antigen a cut-off defining 10% of controls as antibody-positive was close to the approximate point of inflection in the antibody distribution curve. Visual inspection of these percentile plots was previously used in our laboratory to define cut-offs in the absence of gold-standard references since it is assumed that a sudden rise in the antibody response over percentile of serum indicates the cut-off for antibody-positivity ([128] and personal communication with Michael Pawlita). Therefore, I arbitrarily defined cut-offs for antibody-positivity to *F. nucleatum*

proteins allowing 10% of positive control individuals (Figure 17). The technical minimum cut-off was 30 MFI. This definition will be applied separately for the individual studies, i.e. cut-offs will be based on the distribution of MFI among the respective control population.

BliTz and SCCS samples were analyzed in the same experimental setup and could therefore be directly compared. Cut-off values in the two studies were similar for the majority of antigens. An exception, however, was Fn1817_1, for which the cut-off in BliTz was 133 MFI compared to 2039 MFI in the SCCS indicating a strong difference in the antibody response to this protein among the two study populations. EPIC samples were analyzed in a different experimental setup and in a higher serum dilution (1:1000 compared to 1:100 in BliTz and SCCS) and absolute MFI values could not be directly compared to those in BliTz and SCCS. However, in EPIC 10% of controls exceeded only a rather low MFI of 42 to Fn1817_1 similar to the low cut-off in BliTz. Antibody positivity to Fn1817_1 strongly correlated with being African-American in SCCS controls (data not shown). The difference between ethnicities will be elucidated later also in the context of *S. gallolyticus*.

In summary, I had to arbitrarily define cut-offs for positivity to *F. nucleatum* proteins due to a lack of a serological gold standard assay. This cut-off was based on the antibody distribution among control subjects in the separate studies and defined 10% of controls as being positive to the individual proteins. Overall *F. nucleatum* positivity will be arbitrarily defined in a first step as being positive to any of the eleven proteins allowing for inter-individual differences in the immune response but also infection with different bacterial strains.

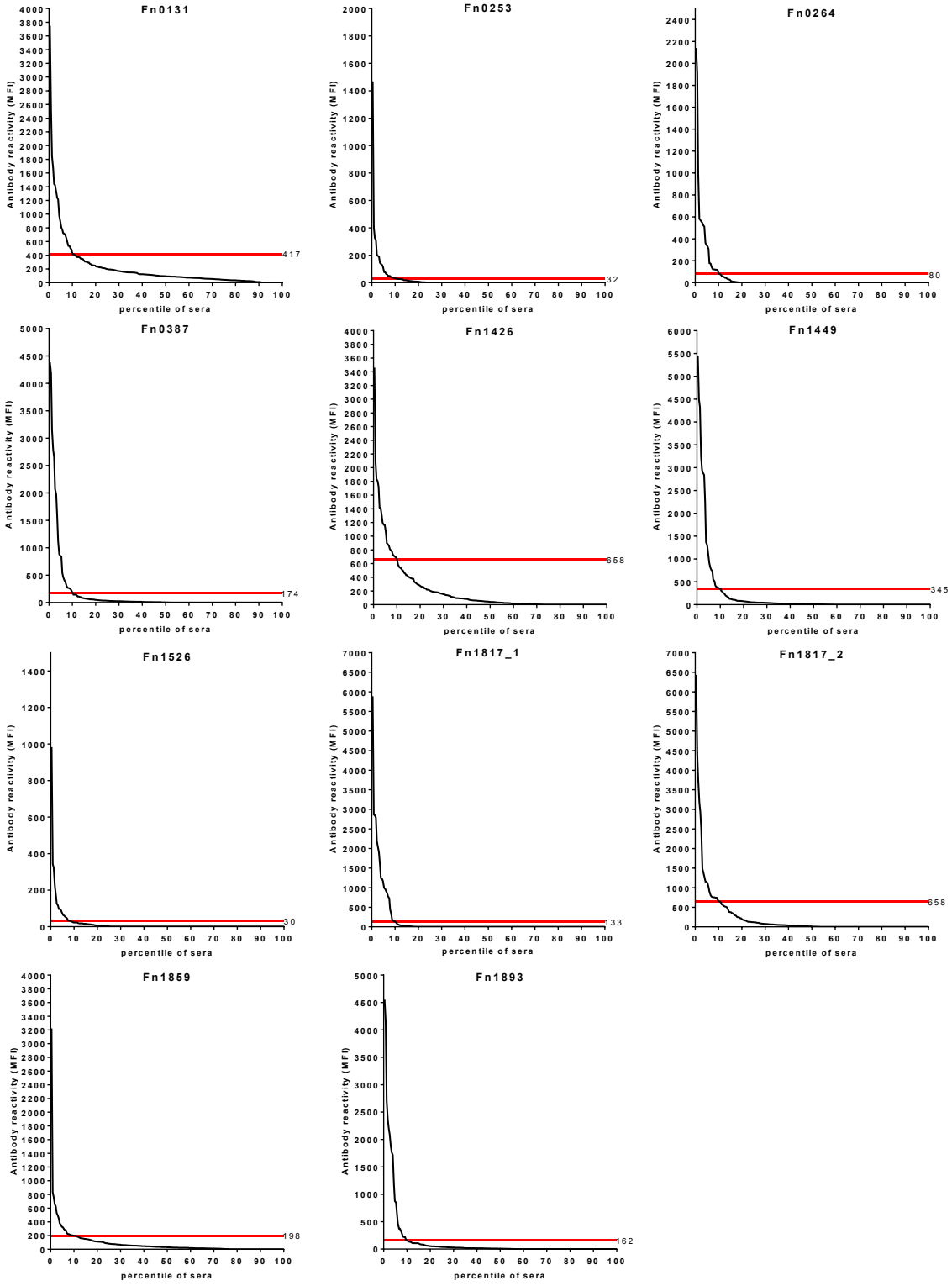


Figure 16: Distribution of antibody responses to *F. nucleatum* proteins in BliTz controls. Antibody reactivity (MFI) is plotted against the percentile of sera. The red line indicates arbitrary antigen-specific cut-offs defining 10% of controls as sero-positive.

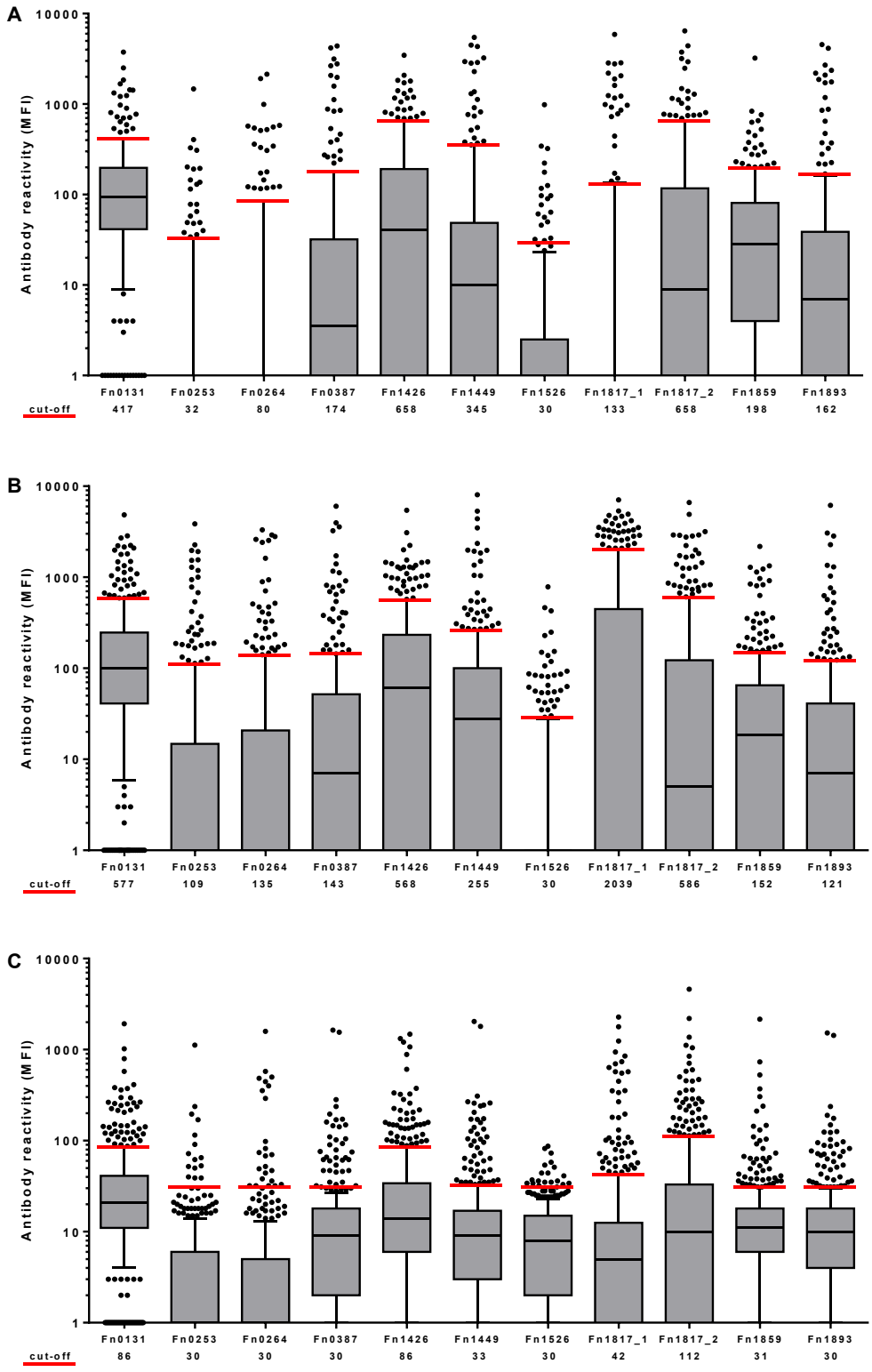


Figure 17: Distribution of antibody responses to individual *F. nucleatum* proteins and antigen-specific cut-offs in (A) BliTz, (B) SCCS and (C) EPIC controls. Boxes represent the 25th to 75th percentile and the solid line represents the median. Whiskers include the 10th to 90th percentile. The red line indicates the arbitrary antigen-specific cut-offs defining 10% of controls as antibody-positive. The technical minimum cut-off was 30 MFI applied in A) for Fn1526 and in C) for Fn0253, Fn0264, Fn0387, Fn1526, Fn1893.

3.3.2. Risk factors for antibody-positivity to any *F. nucleatum* protein among controls

I first assessed whether controls positive to any *F. nucleatum* protein differed from negative control individuals in any of the given baseline characteristics to assess potential risk factors for antibody responses to *F. nucleatum*. These risk factors might affect and/or confound a potential association with CRC.

Significantly more males were positive to any *F. nucleatum* protein among BliTz controls (Table 25). This difference was less pronounced and consequently non-significant in SCCS (Table 26) and EPIC controls (Table 27). None of the other variables was significantly associated with positivity to any *F. nucleatum* protein in BliTz, SCCS and EPIC controls.

Table 25: Comparison of individuals antibody-positive or -negative to any *F. nucleatum* protein for demographic and other risk factors among BliTz controls.

		Any *F. nucleatum* protein		
		neg (n=99)	pos (n=129)	
		n (%)	n (%)	p-value*
Sex	female	62 (63)	62 (48)	
	male	37 (37)	67 (52)	**0.029**
Age (years)	40-59	42 (42)	54 (42)	
	60-65	28 (28)	38 (29)	
	66-85	29 (29)	37 (27)	0.981
	mean (range)	62 (40-85)	62 (50-80)	
School education	< 10 years	55 (57)	70 (55)	
	≥ 10 years	41 (43)	57 (45)	0.746
	missing	3	2	
Smoking status	never	56 (57)	76 (59)	
	ever	43 (43)	52 (41)	0.671
	missing	0	1	
BMI	< 25	42 (43)	47 (37)	
	25-29.9	39 (40)	56 (44)	
	≥ 30	17 (17)	23 (18)	0.693
	missing	1	3	
Family history of CRC	no	86 (87)	114 (88)	
	yes	13 (13)	15 (12)	0.732

*Pearson's Chi-square-test; significant associations are marked in bold font

Table 26: Comparison of individuals antibody-positive or -negative to any *F. nucleatum* protein for demographic and other risk factors among SCCS controls

		Any *F. nucleatum* protein		
		neg (n=134)	neg (n=205)	p-value*
Sex	female	85 (59)	106 (52)	
	male	58 (41)	99 (48)	0.154
Age at blood draw (years)	40-55	67 (47)	97 (47)	
	56-60	28 (20)	48 (23)	
	61-77	48 (34)	60 (29)	0.585
	mean (range)	57 (40-77)	56 (40-76)	
Ethnicity	Caucasian-American	31 (23)	29 (15)	
	African-American	106 (77)	170 (85)	0.058
	missing	6	6	
Education	< high school	60 (44)	75 (37)	
	≥ high school	77 (56)	129 (63)	0.193
	missing	6	1	
Smoking	never	41 (30)	54 (26)	
	former	40 (29)	68 (33)	
	current	56 (41)	82 (40)	0.669
	missing	6	1	
BMI	< 25	31 (23)	53 (26)	
	25-29.9	39 (28)	57 (28)	
	≥ 30	67 (49)	93 (46)	0.751
	missing	6	2	
Family history of CRC	no	54 (87)	73 (86)	
	yes	8 (13)	12 (14)	0.832
	Missing	81	120	

*Pearson's Chi-square-test

Table 27: Comparison of individuals antibody-positive or -negative to any *F. nucleatum* protein for demographic and other risk factors among EPIC controls.

		Any *F. nucleatum* protein		
		neg (n=230)	pos (n=255)	p-value*
Sex	female	122 (53)	125 (49)	
	male	108 (47)	130 (51)	0.376
Age at blood draw, years	37-55	61 (27)	59 (23)	
	56-60	56 (24)	68 (27)	
	61-77	113 (49)	128 (50)	0.657
	mean (range)	60 (37-76)	60 (37-76)	
Country	Italy/Greece	43 (19)	69 (27)	
	France/Spain	42 (18)	51 (20)	
	UK	64 (28)	70 (27)	
	Germany/Netherlands	81 (35)	65 (25)	0.054
Education	≤ primary school	98 (45)	114 (46)	
	technical/professional	63 (29)	52 (21)	
	≥ secondary school	58 (26)	84 (34)	0.082
	missing	11	5	
Smoking status	never	106 (46)	128 (51)	
	former	73 (32)	81 (32)	
	current	51 (22)	44 (17)	0.385
	missing	0	2	
Alcohol	never	27 (13)	21 (9)	
	former	21 (10)	19 (8)	
	current > 0-6 g/day	61 (28)	72 (30)	
	6.01-20 g/day	58 (27)	72 (30)	
	> 20 g/day	49 (23)	60 (25)	0.612
	missing	14	11	
BMI	< 25	79 (34)	88 (35)	
	25-29.9	112 (49)	126 (49)	
	≥ 30	39 (17)	41 (16)	0.966
Family history of CRC	no	82 (94)	94 (92)	
	yes	5 (6)	8 (8)	0.570
	missing	143	153	

*Pearson's Chi-Square-test

3.3.3. Antibody responses to *F. nucleatum* and prevalent CRC in the BliTz-DACHSplus study

The association of antibody responses to *F. nucleatum* was first assessed in prevalent CRC cases from the DACHSplus study compared to BliTz controls as reference. Overall, I neither identified a positive association of CRC with positivity to any of the eleven *F. nucleatum* proteins nor with positivity to individual *F. nucleatum* proteins (Figure 18). Odds for CRC were even decreased with positivity to Fn1859 though this was borderline not significant (OR: 0.53, 95% CI: 0.27-1.04).

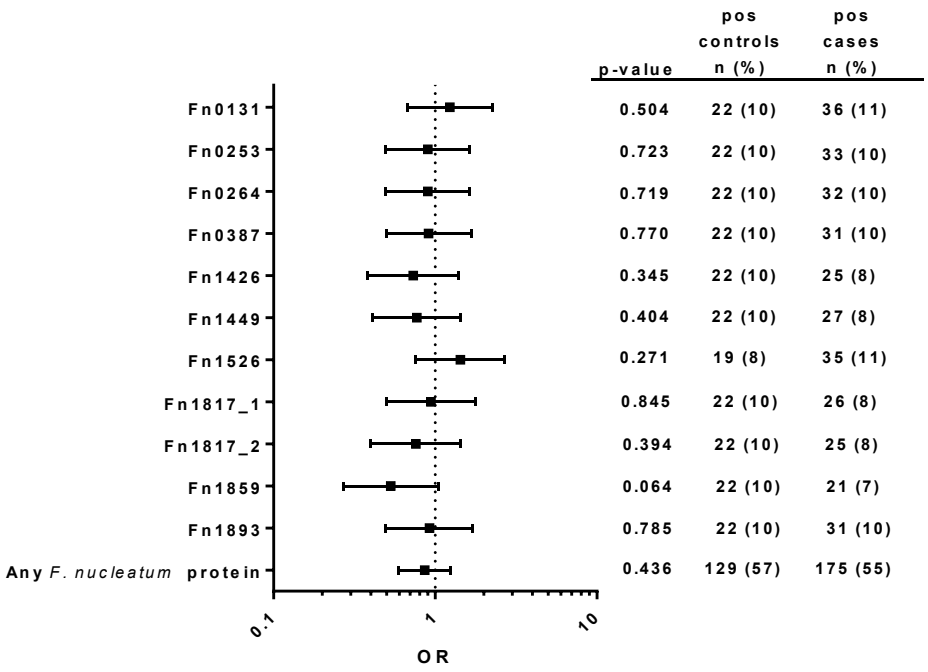


Figure 18: Antibody-positivity to individual *F. nucleatum* proteins and to any *F. nucleatum* protein in relation to CRC in DACHSplus. BliTz controls (n=228) served as reference. OR and 95% CI were estimated using logistic regression models with adjustment for age, sex, smoking, BMI and education. The dashed line indicates null association. pos = antibody-positive.

It was shown in *H. pylori* multiplex serology that correlation among antibody responses to several bacterial proteins increased specificity for detection of the infection [143]. I analyzed whether there were correlations between positivity to individual *F. nucleatum* proteins and whether positivity to two or more proteins occurred preferentially among cases compared to controls and thus in a disease-specific manner (Table 28). I indeed found several protein pairs, where positivity significantly correlated. Strongest pairwise correlations were seen among three proteins: Fn0387, Fn1449 and Fn1893. The fraction of double-positive individuals was thereby similar for controls and cases. However, these three proteins share an 80% amino acid homology (Appendix IV) and the observed correlations most probably represent cross-reactions. Only one pair of sequence-unrelated proteins was significantly correlated and had a more than 2-fold higher fraction of double-positives among cases than among controls: Fn0264-Fn1859. However, double-positivity was present only in 1.6% of cases compared to 0.4% of controls leading to very small group sizes. For most of the other correlated protein pairs, correlation was preferentially among controls and the fraction of double-positives was even lower among cases than among controls, for some even equal to or more than 0.5-fold: Fn0264-Fn1426 (0.5-fold), Fn0264-Fn1817_1 (0.5-fold), Fn1426-Fn1526 (0.4-fold), Fn1526-

FN1817_1 (0.4-fold), Fn1817_1-Fn1893 (0.5-fold) indicating an overall inconsistent and possibly even an opposite trend to what was hypothesized.

In summary, positivity to *F. nucleatum* proteins generally correlated, however, not preferentially among cases. Thus, I was not able to identify a panel of protein pairs that correlated in a higher fraction among cases than among controls.

Since age and sex are main risk factors for CRC development I assessed whether any of the two characteristics might affect the estimate for positivity to any *F. nucleatum* protein with CRC (Figure 19A). As already seen with the risk factor analyses positivity to any *F. nucleatum* protein was more frequent in male controls compared to female controls. This relation was also seen with CRC cases: The fraction of cases positive to any *F. nucleatum* protein and being of female sex was 44% compared to 64% in males. The resulting estimate was an inverse, even significant association of positivity to any *F. nucleatum* protein with CRC cases among women (OR: 0.59 95% CI: 0.37-0.95) in contrast to a null association in men (OR: 1.30, 95% CI: 0.84-1.99). There was no effect of age observed on the estimate.

It was previously reported that *F. nucleatum* abundance was higher in cases with advanced stage compared to lower tumor stage [47]. I here analyzed the association of positivity to any *F. nucleatum* protein separately for UICC stages I/II and III/IV and did not observe a strong difference in the estimate between the two groups. The estimate for higher stages (OR: 0.69, 95% CI: 0.44-1.09) was even lower than that for lower stages (OR: 1.11, 95% CI: 0.71-1.74) (Figure 19B).

Table 28: Percentage *F. nucleatum* antibody double-positives (DP) among BliTz controls (n=228) and DACHSplus cases (n=318)

	0131	0253	0264	0387	1426	1449	1526	1817_1	1817_2	1859	1893
0131		1.3 1.9 (1.5x)[1]	1.8 **2.2** (1.2x)	**2.2** 1.3 (0.6x)	0.4 0.9 (2.3x)	1.3 1.3 (1x)	1.3 **2.5** (1.9x)	0.9 1.3 (1.4x)	1.3 1.6 (1.2x)	1.3 1.6 (1.2x)	1.8 1.3 (0.7x)
0253			0.9 1.6 (1.8x)	**2.6** 1.6 (0.6x)	1.3 1.6 (1.2x)	**3.1** 1.9 (0.6x)	1.8 **2.2** (1.2x)	1.8 0.9 (0.5x)	0.9 1.6 (1.8x)	1.8 **1.6** (0.9x)	**2.2** 1.9 (0.9x)
0264				0.4 0.6 (1.5x)	1.8 **0.9** (0.5x)	0.4 0.9 (2.3x)	1.8 1.6 (0.9x)	1.8 **0.9** (0.5x)	1.3 0.9 (0.7x)	0.4 **1.6** (4x)	0.4 0.6 (1.25x)
0387					1.3 0.9 (0.7x)	**6.6** **6.3** (0.9x)	0.4 1.3 (3.3x)	**2.2** 1.3 (0.6x)	0.4 0.9 (2.3x)	1.3 0.3 (0.2x)	**7.0** **6.3** (0.9x)
1426						**2.2** 1.3 (0.6x)	**2.2** 0.9 (0.4x)	0.4 0.3 (0.8x)	0 0.6 (>999x)	0.4 0.6 (1.5x)	1.3 1.6 (1.2x)
1449							0.9 1.9 (2.1x)	1.8 1.3 (0.7x)	0.4 **0.3** (0.8x)	0.9 0.6 (0.7x)	**6.6** **6.3** (0.9x)
1526								**2.2** 0.9 (0.4x)	0.4 0.6 (1.5x)	0.9 1.3 (1.4x)	0.4 1.3 (3.3x)
1817_1									1.3 0.6 (0.5x)	0.9 0 (0)	**2.6** 1.3 (0.5x)
1817_2										0.4 0.6 (1.5x)	0.9 0.3 (0.3x)
1859											0.9 0.3 (0.3x)
1893											

Grey: % DP among controls; red: % DP among cases; Significant correlations (Pearson's Chi-square test, p-value <0.05) are marked in bold font; [1]%DP among cases relative to %DP among controls; blue field: %DP cases >2-fold than %DP controls and correlation significant

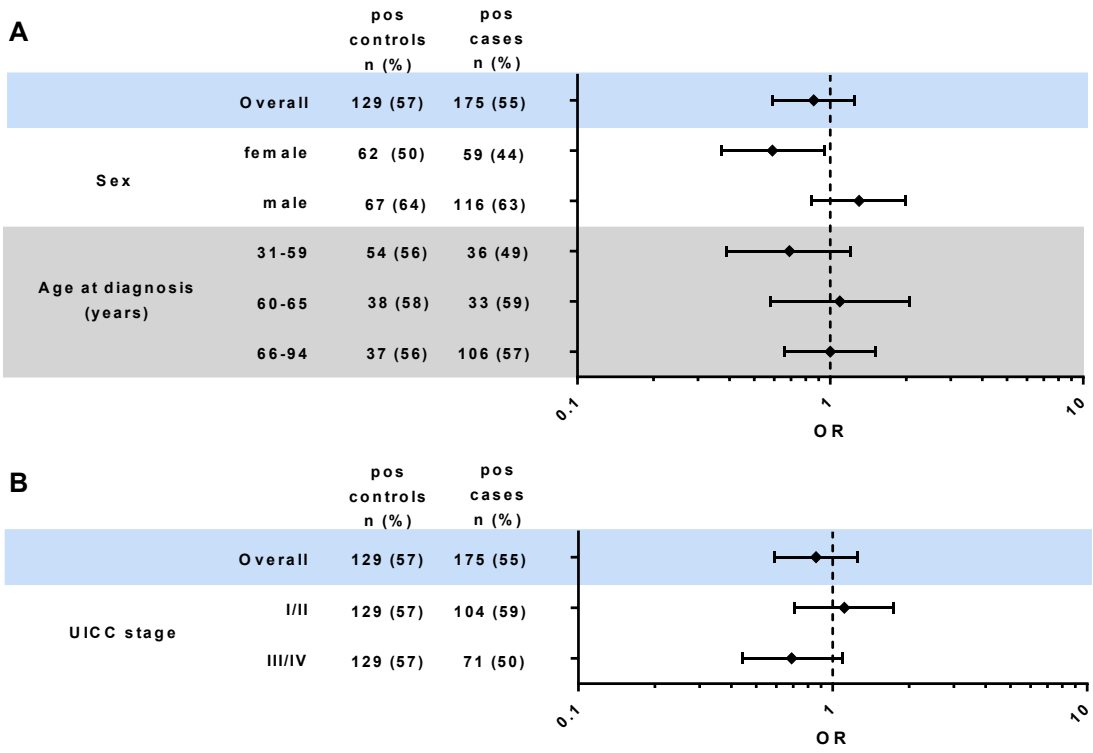


Figure 19: OR and 95% CI for antibody-positivity to any *F. nucleatum* protein in relation to CRC in BliTz-DACHSplus, overall and in separate logistic regression models for (A) sex and age at diagnosis and (B) UICC stage under adjustment for age, sex, BMI, education, and smoking where applicable. The dashed line indicates null association. pos = antibody-positive.

3.3.4. Antibody responses to *F. nucleatum* and incident CRC in the SCCS study

I did not observe any positive association of antibody positivity to individual proteins or any *F. nucleatum* protein with prevalent CRC in BliTz-DACHSplus. A potential association with risk of developing CRC was analyzed with a case-control study nested within SCCS. Overall, there was also no positive association of antibody responses to *F. nucleatum* with CRC risk (Figure 20). However, positivity to Fn1426 individually showed a significant association with CRC (OR: 1.85, 95%CI: 1.04-3.29) with 17% of cases positive compared to 10% of controls. This association was not significant anymore (p-value 0.035) after Bonferroni-correction (p-value < 0.004). In a sensitivity analysis excluding samples with CRC diagnosed within 2 years after blood draw the association with Fn1426 positivity was not significant anymore as well.

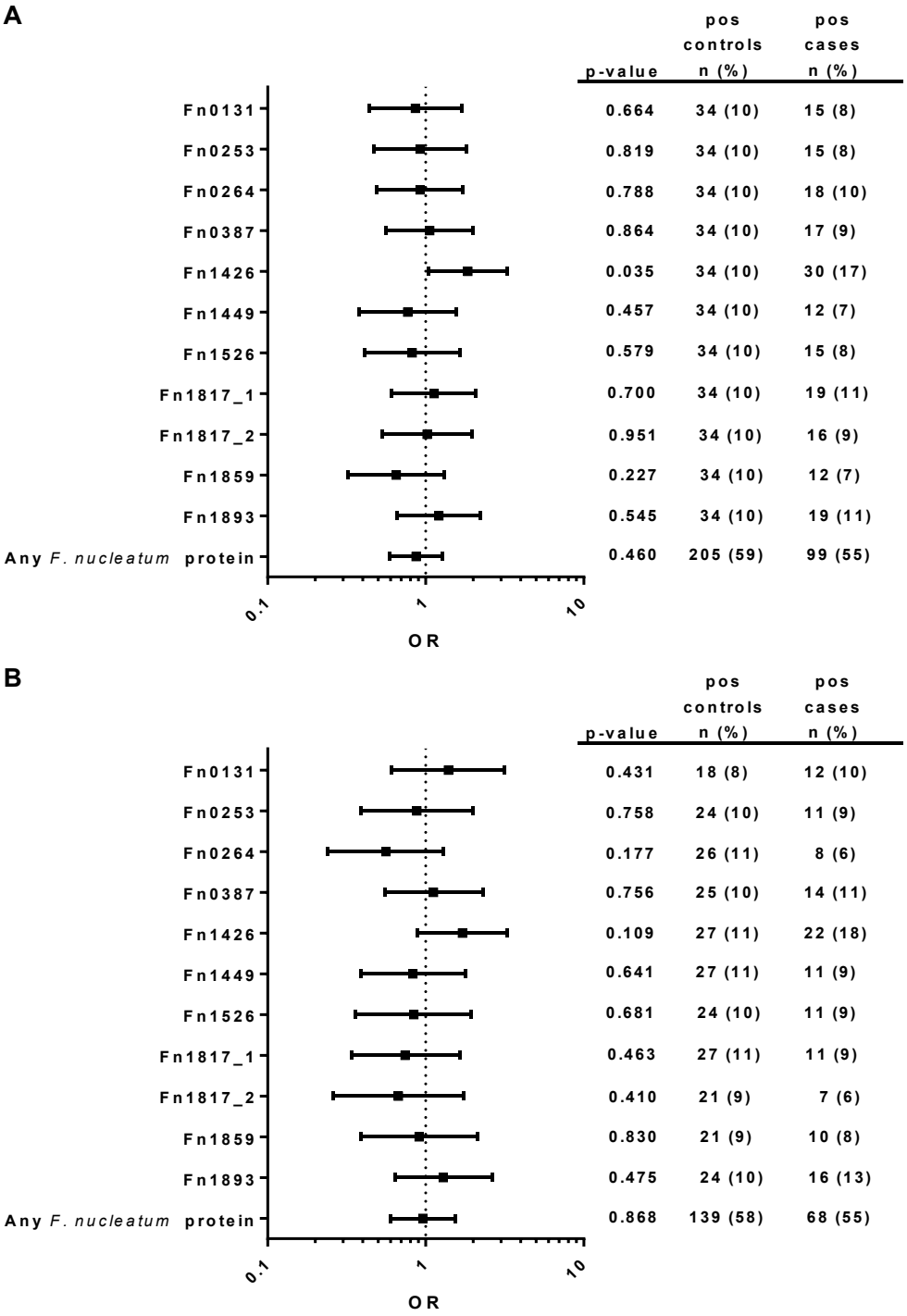


Figure 20: Antibody-positivity to individual *F. nucleatum* proteins and to any *F. nucleatum* protein in relation to CRC risk in SCCS. A) Analyses with all individuals. B) Analyses of individuals diagnosed more than two years after blood draw. OR and 95% CI were estimated using conditional logistic regression models with adjustment for smoking, BMI and education. The dashed line indicates null association. pos = antibody-positive.

Analyses stratified by sex showed a lower OR for males (OR: 0.60, 95% CI: 0.32-1.13) compared to females (OR: 1.02, 95% CI: 0.63-1.66) opposite to what was observed in BliTz-DACHSplus (Figure 21A). However, similar to BliTz-DACHSplus there were more male controls being positive to any *F. nucleatum* protein (63%) compared to female controls (56%). The fraction of positive cases was similar between males and females (53% and 56%, respectively). Stratification by age and ethnicity did not reveal differences in the estimate between the different groups.

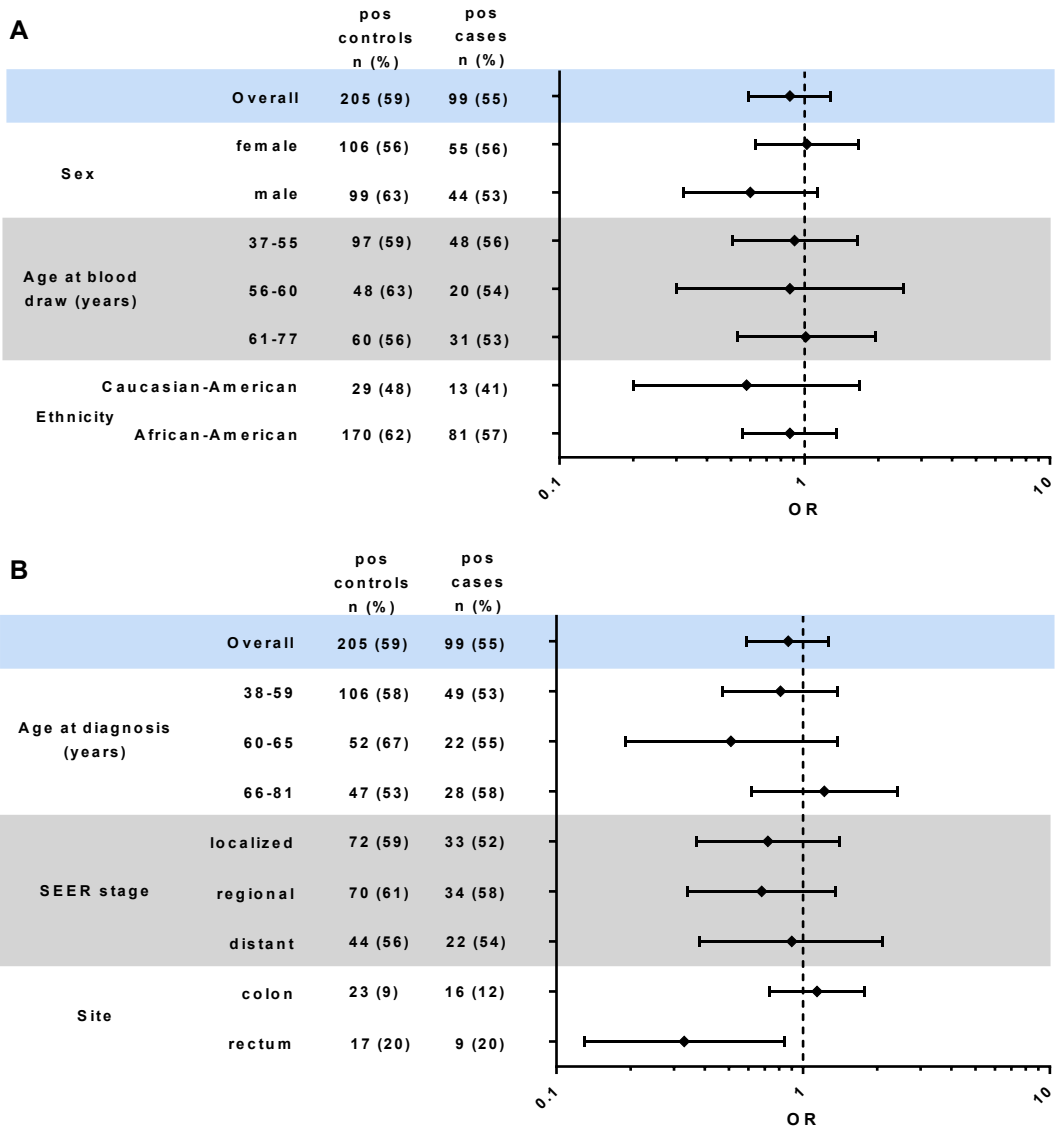


Figure 21: OR and 95% CI for antibody-positivity to any *F. nucleatum* protein in relation to CRC risk in SCCS, overall and in separate conditional logistic regression models for (A) sex, age at blood draw and ethnicity as well as (B) for age at diagnosis, SEER stage and tumor site under adjustment for BMI, education, and smoking. The dashed line indicates null association. pos = antibody-positive.

Analyses separated by case characteristics age at diagnosis, stage and tumor site revealed a significant inverse association of positivity to any *F. nucleatum* protein with rectal cancer only (OR: 0.33, 95% CI: 0.13-0.84) (Figure 21B).

3.3.5. Antibody responses to *F. nucleatum* and incident CRC in the EPIC study

The association of antibody responses to *F. nucleatum* proteins with risk of developing CRC was finally assessed in a second case-control study nested within the prospective study EPIC. Also in this study I did not observe a positive association of positivity to any *F. nucleatum* protein nor with individual *F. nucleatum* proteins with CRC risk. Positivity to Fn0131 was even significantly inversely associated (OR: 0.59, 95% CI: 0.36-0.95) with CRC risk. However, this inverse association (p-value 0.030) was not significant after Bonferroni-correction (p-value < 0.004) and also not in samples that were diagnosed more than two years after blood draw.

Analyses stratified by baseline characteristics showed a difference in the estimate between male and female sex. The difference observed in EPIC was similar but less pronounced to that observed in BliTz-DACHSplus with females having an inverse and even significant association of positivity to any *F. nucleatum* protein with CRC (OR: 0.64, 95% CI: 0.43-0.96). In EPIC there was also a difference between the age groups observable with a significant inverse association for age 56-60 (OR: 0.43, 95% CI: 0.22-0.85) compared to younger and older individuals. Also I observed a difference in the association with CRC risk between the different countries of residence with individuals from Italy/Greece being significantly inversely associated (OR: 0.55, 95% CI: 0.31-0.98) and individuals from Germany/Netherlands having non-significantly increased odds (OR: 1.21, 95%CI: 0.70-2.11).

There was no difference observable with separate analyses by age at diagnosis, stage or tumor site.

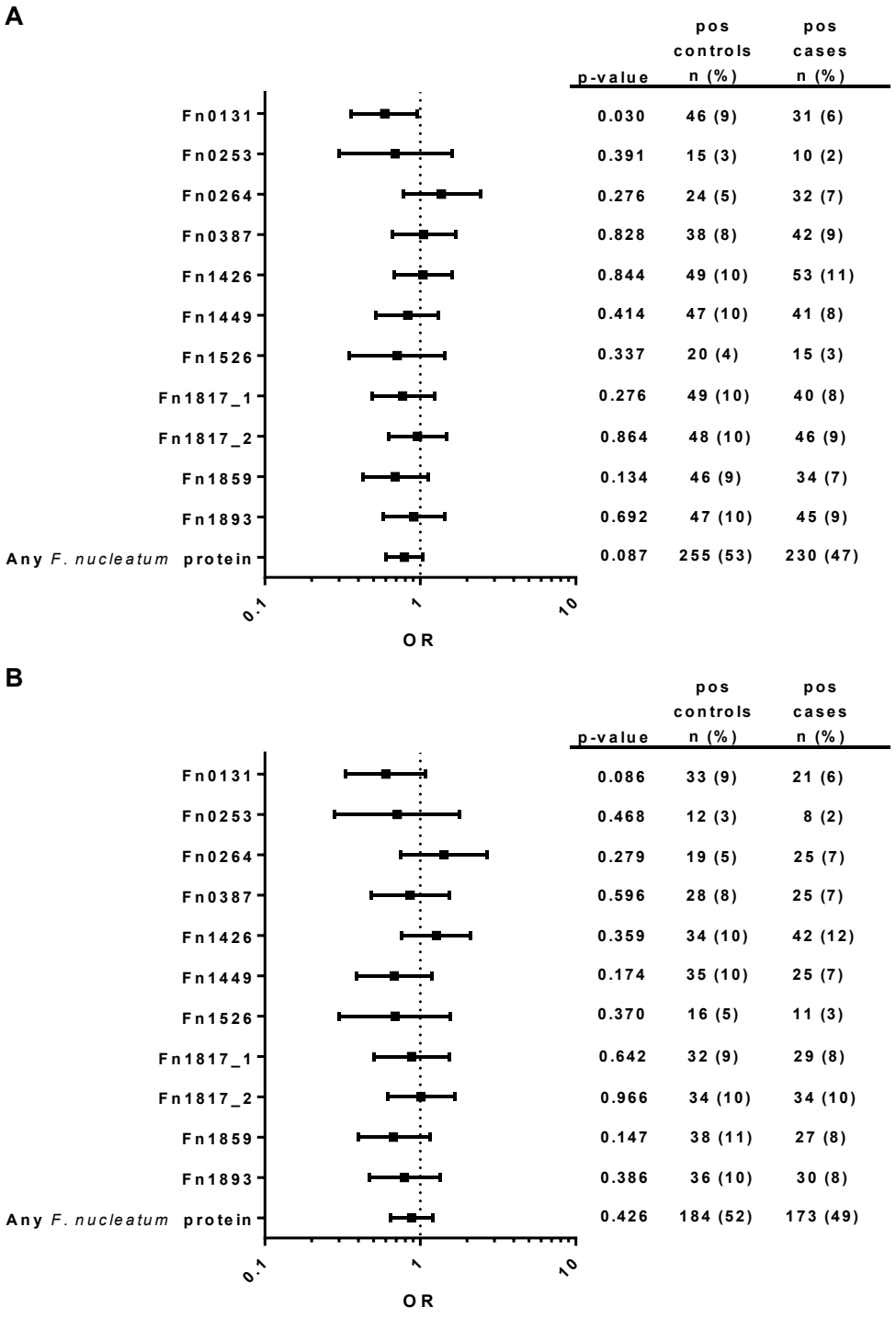


Figure 22: Antibody-positivity to individual *F. nucleatum* proteins and to any *F. nucleatum* protein in relation to CRC risk in EPIC. A) Analyses with all individuals. B) Analyses of individuals diagnosed more than two years after blood draw. OR and 95% CI were estimated using conditional logistic regression models with adjustment for smoking, alcohol status, BMI and education. The dashed line indicates null association. pos = antibody-positive.

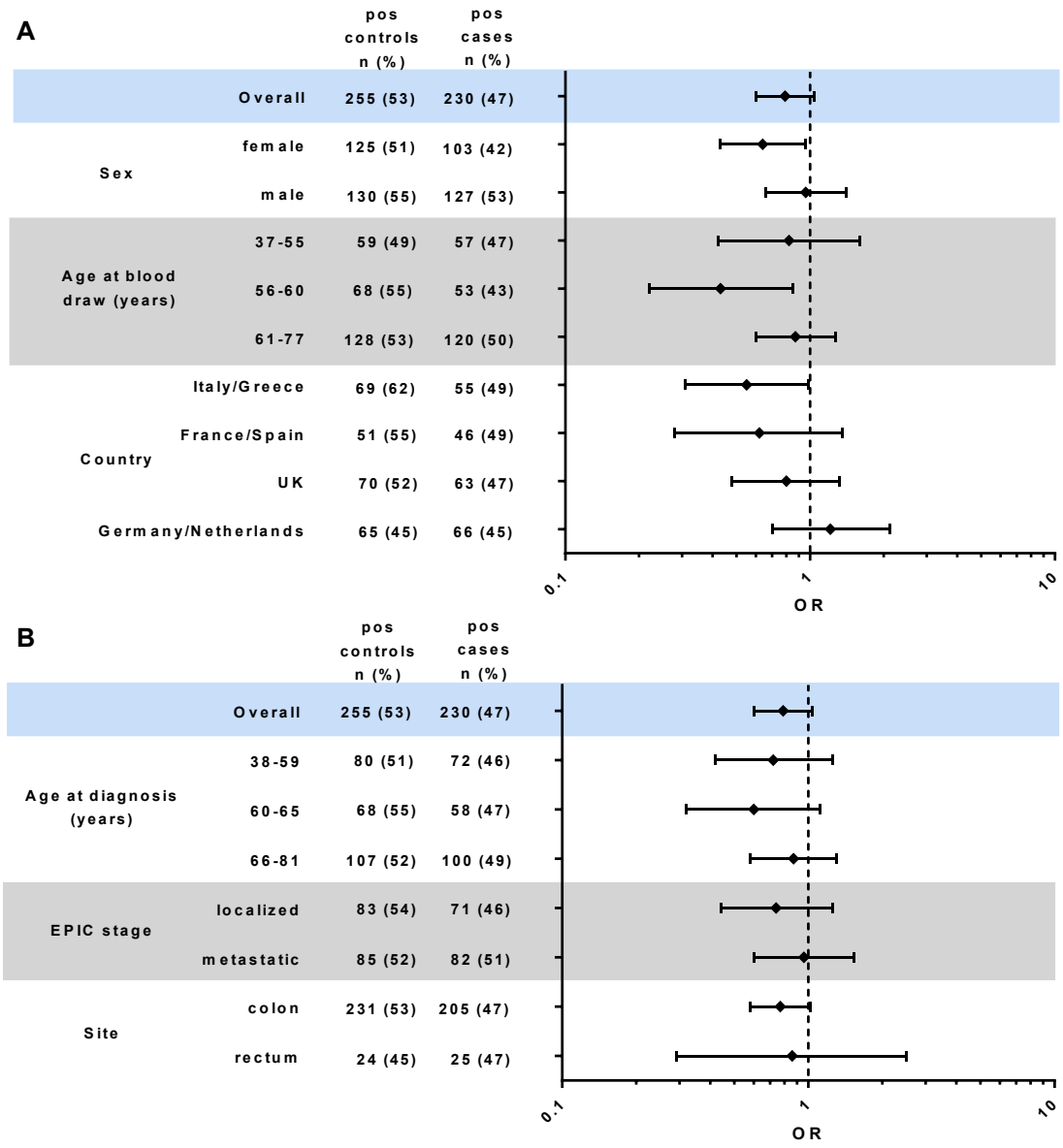


Figure 23: OR and 95% CI for antibody-positivity to any *F. nucleatum* protein in relation to CRC risk in EPIC, overall and in separate conditional logistic regression models for (A) sex, age at blood draw and country of residence as well as (B) for age at diagnosis, EPIC stage and tumor site under adjustment for BMI, education, alcohol and smoking status. The dashed line indicates null association. pos = antibody-positive.

3.4. Search for serological associations of *S. gallolyticus* with CRC by multiplex serology in a retrospective case-control study and prospective nested case-control studies

3.4.1. Cut-off definition

The different attempts to validate *S. gallolyticus* multiplex serology based on *S. gallolyticus* positivity in animal but also human samples were not sufficient to define cut-offs for individual proteins as well as an algorithm for *S. gallolyticus* antibody-positivity. Therefore, I arbitrarily defined cut-off values for each individual *S. gallolyticus* protein that allowed 10% of controls to be antibody-positive to this protein. The distribution of antibody responses in BliTz controls to the individual *S. gallolyticus* antigens is shown in Figure 24. The arbitrarily defined cut-offs were mostly more stringent than cut-offs that would have been defined by the inflection points of the curves identified by visual inspection of the percentile plots. For example the cut-off for Gallo0272 was 1780 MFI, whereas the approximate inflection point of the curve was at 750 MFI and corresponding to the 20^{th} percentile. In the attempt for higher specificity I opted for the more stringent cut-off. The cut-off values were defined separately for each individual study based on the respective control population. The technical minimum cut-off was 30 MFI.

The cut-offs defined for each *S. gallolyticus* antigen in the different studies are shown in Figure 25. BliTz and SCCS samples were analyzed in the same experiment and can therefore be directly compared. 10% of controls achieved slightly higher MFI in SCCS compared to colonoscopy-negative individuals in BliTz. The strongest differences were seen with Gallo0272 (cut-off BliTz: 1780 MFI vs. SCCS: 2272 MFI), Gallo0933 (cut-off BliTz: 1640 MFI vs. SCCS: 2566 MFI), Gallo2018 (cut-off BliTz: 984 MFI vs. SCCS: 1475 MFI), Gallo2178 (cut-off BliTz: 30 MFI (technical minimum cut-off, reached by only 5%) vs. SCCS: 140 MFI) and Gallo2179 (cut-off BliTz: 919 MFI vs. SCCS: 1425 MFI). This indicates a difference in the antibody response *S. gallolyticus* proteins among the two control groups. The analysis of EPIC serum samples was performed in a different experiment and also in a higher serum dilution (1:1000 compared to 1:100 in BliTz and SCCS) and was therefore not directly comparable to BliTz and SCCS for the absolute MFI.

In summary, I arbitrarily defined cut-off values for antibody-positivity to individual *S. gallolyticus* proteins based on the assumption that colonization with and seroconversion

against *S. gallolyticus* are rare (10%) events in healthy individuals. This cut-off was based on the antibody distribution among control subjects in the separate studies and defined 10% of controls as being positive to the individual proteins. An algorithm for overall *S. gallolyticus* antibody-positivity will be arbitrarily defined in a first step as being positive to any of the eleven proteins allowing for inter-individual differences in the immune response but also infection with different bacterial strains potentially expressing different sets of proteins.

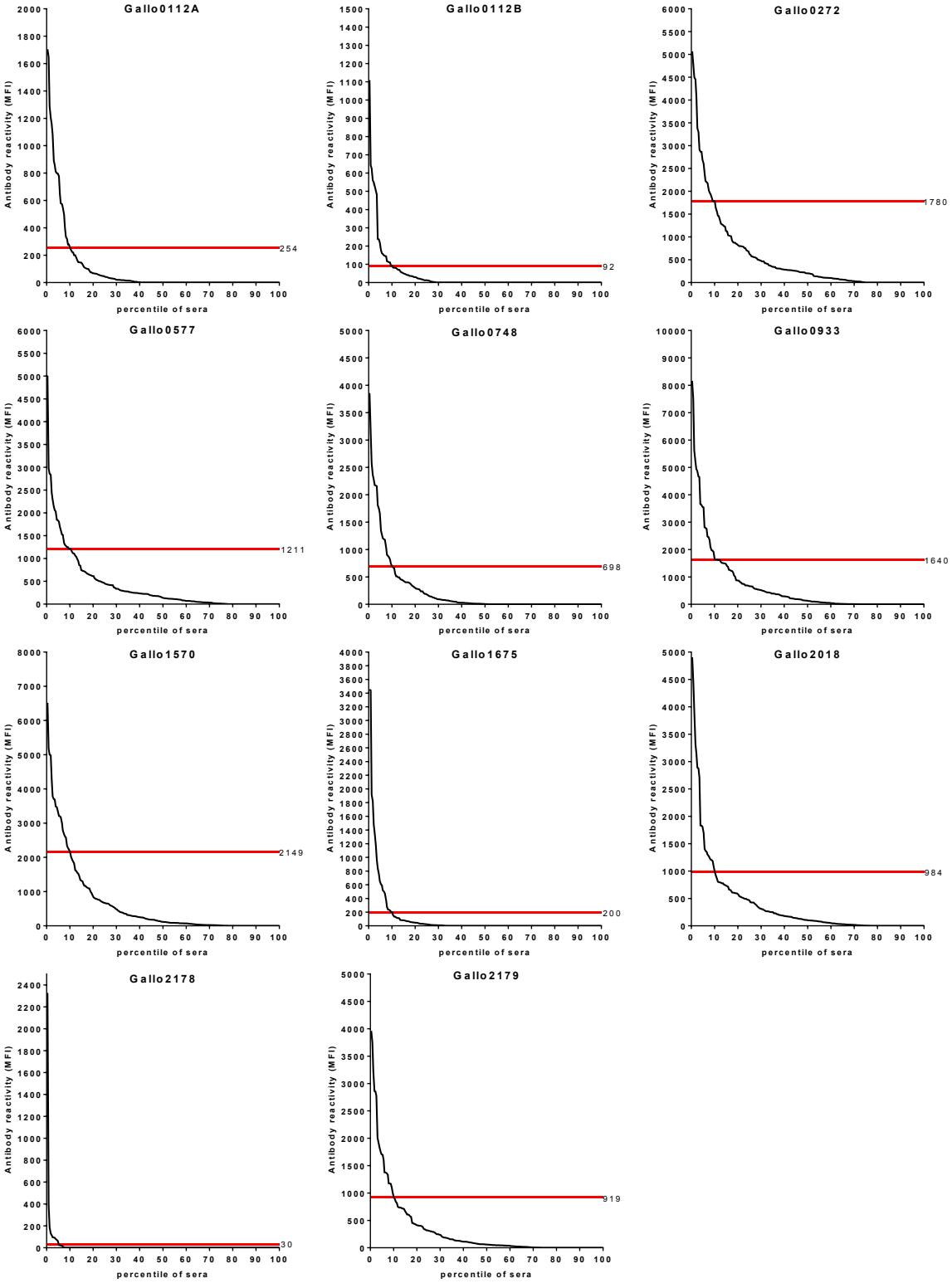


Figure 24: Distribution of antibody responses to *S. gallolyticus* proteins in BliTz controls. Antibody reactivity (MFI) is plotted over the percentile of sera. The red line indicates the arbitrarily chosen antigen-specific cut-offs defining 10% of controls as antibody-positive.

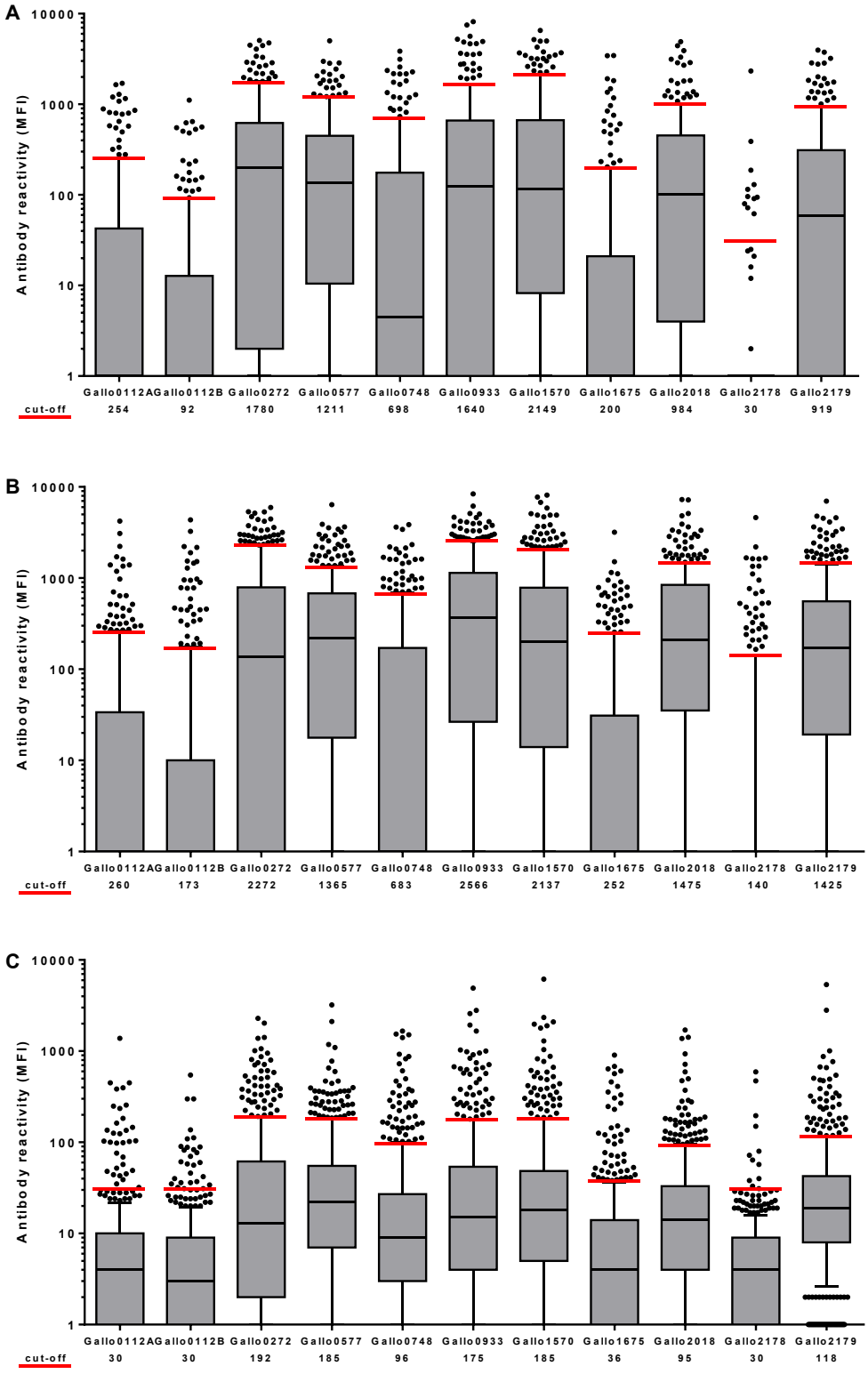


Figure 25: Distribution of antibody responses to individual *S. gallolyticus* proteins and antigen-specific cut-offs in (A) BliTz, (B) SCCS and (C) EPIC controls. Boxes represent the 25th to 75th percentile and the solid line represents the median. Whiskers include the 10th to 90th percentile. The red line indicates the arbitrary antigen-specific cut-offs defining 10% of controls as antibody-positive. The technical minimum cut-off was 30 MFI applied in A) for Gallo2178 and in C) for Gallo0112A, Gallo0112B and Gallo2178.

3.4.2. Risk factors for antibody-positivity to any *S. gallolyticus* protein among controls

I first addressed whether controls antibody-positive to any *S. gallolyticus* protein differed from negative controls in any of the given baseline characteristics to assess potential risk factors for antibody responses to *S. gallolyticus*. These risk factors might affect and/or confound a potential association with CRC. However, none of the variables was significantly associated with positivity to any *S. gallolyticus* protein in BliTz (Table 29), SCCS (Table 30) and EPIC controls (Table 31).

Table 29: Comparison of individuals antibody-positive or -negative to any *S. gallolyticus* protein for demographic and other risk factors among BliTz controls.

		Any *S. gallolyticus* protein		
		neg (n=97)	pos (n=131)	
		n (%)	n (%)	p-value*
Sex	female	60 (62)	64 (49)	
	male	37 (38)	67 (51)	0.051
Age (years)	40-60	39 (40)	57 (44)	
	60-65	31 (32)	35 (27)	
	66-85	27 (28)	39 (30)	0.689
	mean (range)	62 (40-79)	62 (48-85)	
School education	< 10 years	58 (61)	67 (52)	
	≥ 10 years	37 (39)	61 (48)	0.195
	missing	2	3	
Smoking status	never	55 (57)	77 (59)	
	ever	41 (43)	54 (41)	0.929
	missing	1	0	
BMI	< 25	44 (45)	45 (35)	
	25-29.9	37 (38)	58 (46)	
	≥ 30	16 (16)	24 (19)	0.320
	missing	0	4	
Family history of CRC	no	82 (85)	118 (90)	
	yes	15 (15)	13 (10)	0.208

*Pearson's Chi-square-test

Table 30: Comparison of individuals antibody-positive or -negative to any *S. gallolyticus* protein for demographic and other risk factors among SCCS controls

		Any *S. gallolyticus* protein		
		neg (n=142)	pos (n=206)	p-value*
Sex	female	80 (56)	111 (54)	
	male	62 (44)	95 (46)	0.651
Age at blood draw (years)	40-55	61 (43)	103 (50)	
	56-60	34 (24)	42 (20)	
	61-77	47 (33)	61 (30)	0.427
	mean (range)	57 (40-76)	56 (40-77)	
Ethnicity	Caucasian-American	28 (20)	32 (16)	
	African-American	111 (80)	165 (84)	0.358
	missing	3	9	
Education	< high school	48 (35)	87 (43)	
	≥ high school	91 (65)	115 (57)	0.113
	missing	3	4	
Smoking	never	39 (28)	56 (28)	
	former	43 (31)	65 (32)	
	current	57 (41)	81 (40)	0.970
	missing	3	4	
BMI	< 25	36 (26)	48 (24)	
	25-29.9	42 (30)	54 (27)	
	≥ 30	61 (44)	99 (49)	0.617
	missing	3	5	
Family history of CRC	no	52 (91)	75 (83)	
	yes	5 (9)	15 (17)	0.174
	Missing	85	116	

*Pearson's Chi-square-test

Table 31: Comparison of individuals antibody-positive or -negative to any *S. gallolyticus* protein for demographic and other risk factors among EPIC controls.

		Any *S. gallolyticus* protein		
		neg (n=212)	pos (n=273)	p-value*
Sex	female	103 (49)	144 (53)	
	male	109 (51)	129 (47)	0.363
Age at blood draw, years	37-55	44 (21)	76 (28)	
	56-60	55 (26)	69 (25)	
	61-77	113 (53)	128 (47)	0.180
	mean (range)	60 (39-77)	59 (37-75)	
Country	Italy/Greece	50 (24)	62 (23)	
	France/Spain	41 (19)	52 (19)	
	UK	60 (28)	74 (27)	
	Germany/Netherlands	61 (29)	85 (31)	0.955
Education	none/primary school completed	92 (45)	120 (46)	
	technical/professional	54 (26)	61 (23)	
	secondary school/longer education	60 (29)	82 (31)	0.736
	missing	6	10	
Smoking status	never	94 (45)	140 (51)	
	former	73 (35)	81 (30)	
	current	44 (21)	51 (19)	0.316
	missing	1	1	
Alcohol intake	never	20 (10)	28 (11)	
	former	17 (8)	23 (9)	
	current > 0-6 g/day	49 (24)	84 (33)	
	6.01-20 g/day	63 (31)	67 (26)	
	> 20 g/day	53 (26)	56 (22)	0.230
	missing	10	15	
BMI	< 25	76 (36)	91 (33)	
	25-29.9	95 (45)	143 (52)	
	≥ 30	41 (19)	39 (14)	0.177
Family history CRC	no	83 (95)	93 (91)	
	yes	4 (5)	9 (9)	0.253
	missing	125	171	

*Pearson's Chi-Square-test

3.4.3. Antibody responses to *S. gallolyticus* and prevalent CRC in the BliTz-DACHSplus study

The association of antibody responses to *S. gallolyticus* was first assessed with prevalent CRC in samples of the BliTz-DACHSplus study.

In a previous independent CRC case-control study conducted in Spain using different protein sources we had found a significant association of antibodies to Gallo2178 and Gallo2179, individually but also in combination, with prevalent CRC [81]. Therefore, I first aimed to

replicate this finding in the BliTz-DACHSplus study (Figure 26). Indeed, also in this study positivity to Gallo2178 was significantly associated with CRC (OR: 4.50, 95%CI: 2.22-9.11) with 17% positive cases compared to 5% positive controls. The association was even significant (p-value < 0.0001) after Bonferroni-correction (p-value < 0.0036). Gallo2179 alone was not significantly associated with CRC here (OR: 1.35, 95%CI: 0.73-2.51), however, double-positivity to Gallo2178-Gallo2179 showed a significant association with CRC. Since none of the controls was double-positive to Gallo2178-Gallo2179, it was not possible to calculate an estimate for the strength of the association, a crude analyses, however, showed a strong association (p-value 0.001) with 4% of DACHSplus cases being double-positive.

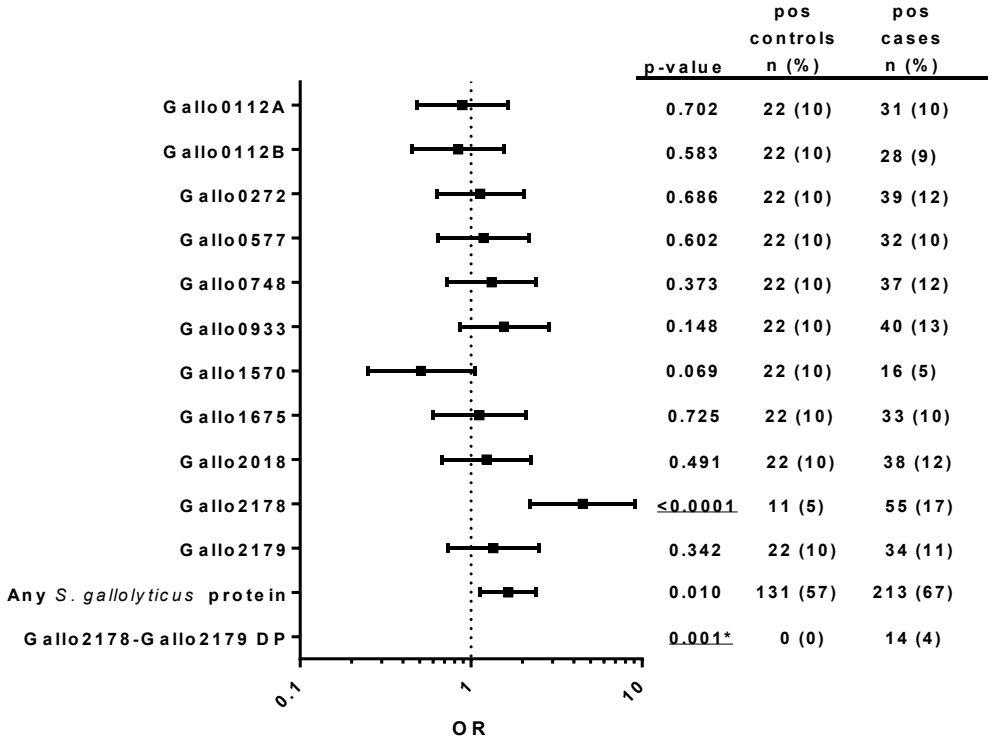


Figure 26: Antibody-positivity to individual *S. gallolyticus* proteins, double-positivity to Gallo2178 and Gallo2179 as well as positivity to any *S. gallolyticus* protein in relation to CRC in DACHSplus. BliTz controls (n=228) served as reference. OR and 95% CI were estimated using logistic regression models with adjustment for age, sex, smoking, BMI and education. The dashed line indicates null association. Significant associations after Bonferroni-correction (p-value < 0.0036) are underlined. *Crude analyses using Pearson's Chi-square test. pos = antibody-positive.

After the replication for Gallo2178 and Gallo2179 I explored whether any of the newly developed proteins was individually associated with CRC. However, this was not the case for any of the additional nine proteins individually. Positivity to any of the in total eleven proteins, however, was significantly associated with CRC (OR: 1.65, 95% CI: 1.13-2.41) with 67% positive cases compared to 57% positive controls. Considering multiple testing the association was not significant (p-value 0.01) (Figure 26). Sensitivity analyses with exclusion of the strongest individual marker Gallo2178 resulted in a non-significant OR (OR: 1.29; 95% CI: 0.89-1.87) indicating that the association with positivity to any *S. gallolyticus* protein was mostly driven by the strong impact of Gallo2178.

The analyses of bovine sera as well as of sera from humans with *S. gallolyticus* bacteremia indicated that *S. gallolyticus* antibody-positive individuals could be positive to several of the eleven proteins simultaneously (3.2.3 and 3.2.4). Therefore, I assessed whether positivity to the different proteins correlated. Under the assumption that the presence of CRC or precancerous lesions is a prerequisite for *S. gallolyticus* infection, I expected this correlation to be stronger among cases than among controls. Several correlating protein pairs were identified (Table 32). Two of them (Gallo0577-Gallo1570 and Gallo0577-Gallo2179) correlated significantly among controls and cases. Sequence alignment of these three proteins (Appendix III) identified homologous stretches especially at the C-terminus of the proteins. This indicated that correlations seen among these proteins were probably due to antibodies reacting to conserved epitopes in the C-terminus of the three proteins. Apart from Gallo2178-Gallo2179, which were correlating only among cases and where the double-positivity to both proteins was more than 2-fold higher among cases than among controls, three additional pairs with the same pattern were identified: Gallo0272-Gallo0748 (3.4-fold more double-positive cases compared to controls), Gallo0272-Gallo1675 (2.8-fold) and Gallo2018-Gallo2179 (2.2-fold). The in total 6 proteins (Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178 and Gallo2179) were combined in a 6-marker panel and positivity was defined as being positive to two or more proteins of this panel. 11% of controls were positive to two or more proteins of the 6-marker panel, compared to 19% of cases resulting in a significant association with CRC (OR: 1.99, 95%CI: 1.15-3.45) (Table 33). However, with Bonferroni-correction the association was not significant anymore (p-value 0.014).

I further addressed whether the association of CRC with positivity to two or more proteins of the 6-marker panel was affected by baseline characteristics age and sex and whether the

association differed between different stages of the tumor (Figure 27). However, stratification by age or sex did not affect the estimate. Also the estimate of the association did not differ between different stages of the tumor.

Table 32: Percentage *S. gallolyticus* antibody double-positives (DP) among Blitz controls (n=228) and DACHSplus cases (n=318)

	0112A	0112B	0272	0577	0748	0933	1570	1675	2018	2178	2179
0112A		1.3 0.9 (0.7x)[1]	0.4 1.6 (4x)	1.8 1.3 (0.7x)	0.9 1.3 (1.4x)	0 1.6 (>999x)	1.3 0.3 (0.2x)	1.3 1.3 (1x)	1.3 1.9 (1.5x)	0.4 1.9 (4.8x)	1.3 0.9 (0.7x)
0112B			0.4 1.9 (4.8x)	**2.2** 0.6 (0.3x)	1.3 0.9 (0.7x)	**2.2** 0.9 (0.4x)	0.9 **1.6** (1.8x)	0.4 0.9 (2.3x)	1.8 1.6 (0.9x)	**1.3** 1.3 (1x)	0 0.6 (>999x)
0272				1.3 1.3 (1x)	0.9 **3.1 (3.4x)**	0.9 1.3 (1.4x)	1.3 1.9 (1.5x)	0.9 **2.5 (2.8x)**	0.4 2.2 (5.5x)	0.4 2.5 (6.3x)	1.3 1.9 (1.5x)
0577					1.8 1.6 (0.9x)	1.3 0.6 (0.5x)	**4.0** **2.2** (0.6x)	1.3 1.6 (1.2x)	1.3 1.9 (1.5x)	0.4 2.5 (6.3x)	**3.5** **4.4** (1.3x)
0748						1.8 2.5 (1.4x)	1.3 **1.6** (1.2x)	1.8 1.9 (1.1x)	**2.2** 1.3 (0.6x)	0.9 **0.3** (0.3x)	0.9 0.9 (1x)
0933							0.9 0.6 (0.7x)	0.9 0.9 (1x)	1.3 **0.3** (0.2x)	0 1.3 (>999x)	0.4 0.6 (1.5x)
1570								0.4 0.6 (1.5x)	1.3 0.3 (0.2x)	0.4 1.3 (3.3x)	**2.6** 1.3 (0.5x)
1675									0.9 1.9 (2.1x)	0.4 1.9 (4.8x)	1.3 **2.2** (1.7x)
2018										0.4 3.1 (7.8x)	1.3 **2.8 (2.2x)**
2178											0 **4.4** (>999x)
2179											

Grey: %DP among controls; red: %DP among cases; Significant correlations (Pearson's Chi-square test, p-value <0.05) are marked in bold font; [1]%DP among cases relative to %DP among controls; blue: %DP cases >2-fold than %DP controls and correlation significant

Table 33: Antibody responses to the *S. gallolyticus* 6-marker panel in relation to prevalence of CRC in the DACHSplus study compared to BliTz controls

	Positive n (%)				
	Controls n=228	Cases n=318	OR[1]	95% CI	p-value
≥2 of 6-marker panel	24 (11)	60 (19)	**1.99**	**1.15-3.45**	**0.014**

6-marker panel includes Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178 and Gallo2179; [1]Logistic regression model with adjustment for age (continuous variable), sex, BMI, education and smoking; Significant associations are marked in bold font

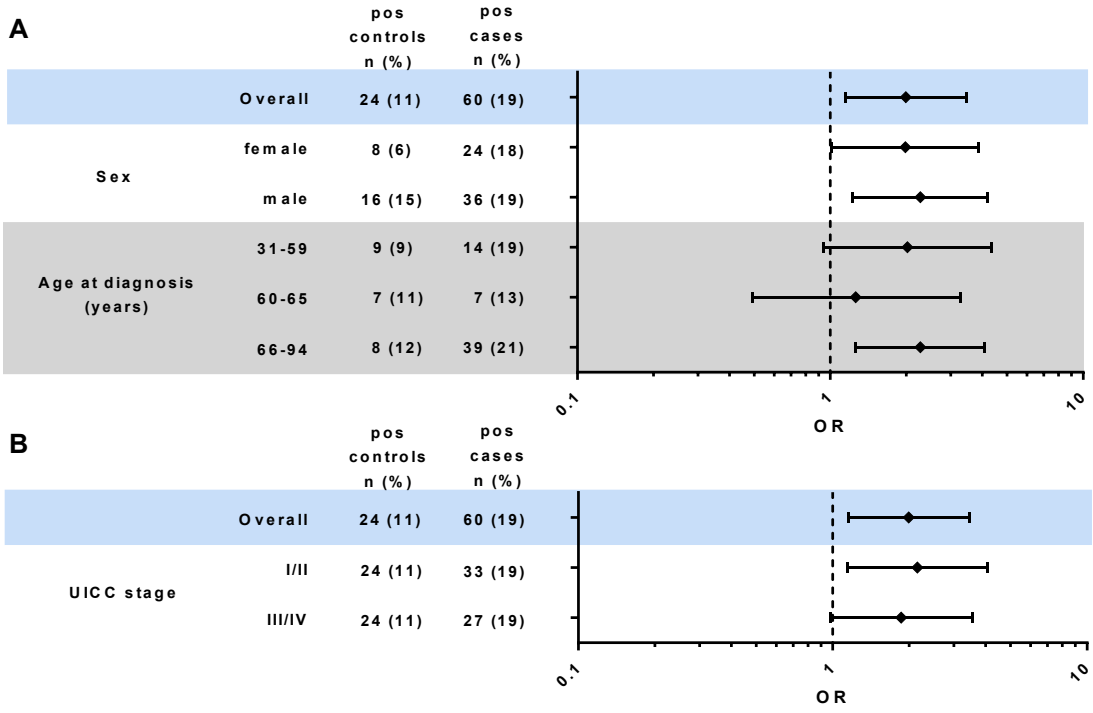


Figure 27: OR and 95% CI for antibody-positivity to two or more proteins of the *S. gallolyticus* 6-marker panel in relation to CRC in BliTz-DACHSplus, overall and in separate logistic regression models for (A) sex, age at diagnosis and (B) UICC stage under adjustment for age, sex BMI, education, and smoking where applicable. The dashed line indicates null association. pos = antibody-positive.

A specificity analyses was carried out to assess whether observed associations of antibody responses to *S. gallolyticus* proteins with CRC were disease-specific. It was estimated whether positivity to individual *S. gallolyticus* proteins, any *S. gallolyticus* protein, to two or more proteins of the 6-marker panel or double-positivity to Gallo2178-Gallo2179 was associated with gastric cancer (GC) cases of the DACHSplus study in reference to BliTz controls (Figure 28). Odds for GC were not significantly increased with any of the above mentioned possibilities for *S. gallolyticus* positivity indicating a disease-specific association of antibody responses to *S. gallolyticus* with prevalent CRC in the BliTz-DACHSplus study.

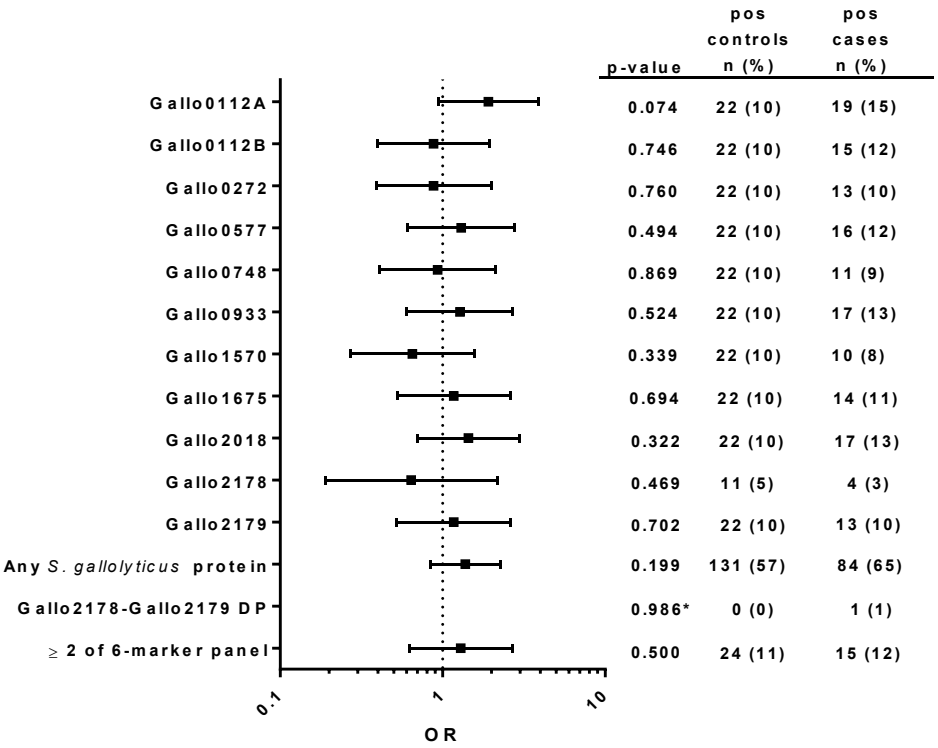


Figure 28: Antibody-positivity to individual *S. gallolyticus* proteins and protein combinations in relation to gastric cancer cases (GC) in DACHSplus. BliTz controls (n=228) served as reference. OR and 95% CI were estimated using logistic regression models with adjustment for age, sex, smoking, BMI and education. The dashed line indicates null association. Significant associations after Bonferroni-correction are underlined. 6-marker panel includes Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178 and Gallo2179 *Crude analyses using Chi-square test. pos = antibody-positive.

3.4.4. Antibody responses to *S. gallolyticus* and incident CRC in SCCS

The observed associations in the BliTz-DACHSplus CRC case-control studies were assessed in a prospective setting with the CRC case-control study nested within SCCS. However, none of the associations of antibody responses to *S. gallolyticus* with CRC in the BliTz-DACHSplus study could be reproduced in the SCCS. OR for positivity to individual proteins, for positivity to any of the eleven proteins, double-positivity to Gallo2178-Gallo2179 as well as positivity to two or more proteins of the 6-marker panel with CRC risk ranged around the null in the SCCS (Figure 29A). Also exclusion of cases diagnosed within two years from the analyses did not alter the estimates (Figure 29B).

Stratification of the study by baseline characteristics age at blood draw, sex and ethnicity did not reveal a certain group at specifically increased odds for CRC with positivity to two or more proteins of the 6-marker panel (Figure 30A). The very small group of Caucasian-Americans included in this study showed an elevated odds ratio (OR: 3.94, 95% CI: 0.37-41.54) compared to African-Americans (OR: 1.03, 95% CI: 0.56-1.89), however, due to the small sample size confidence intervals were widely overlapping. Subgroup analyses by age at diagnosis, stage or tumor site did not reveal any group at increased risk either (Figure 30B)

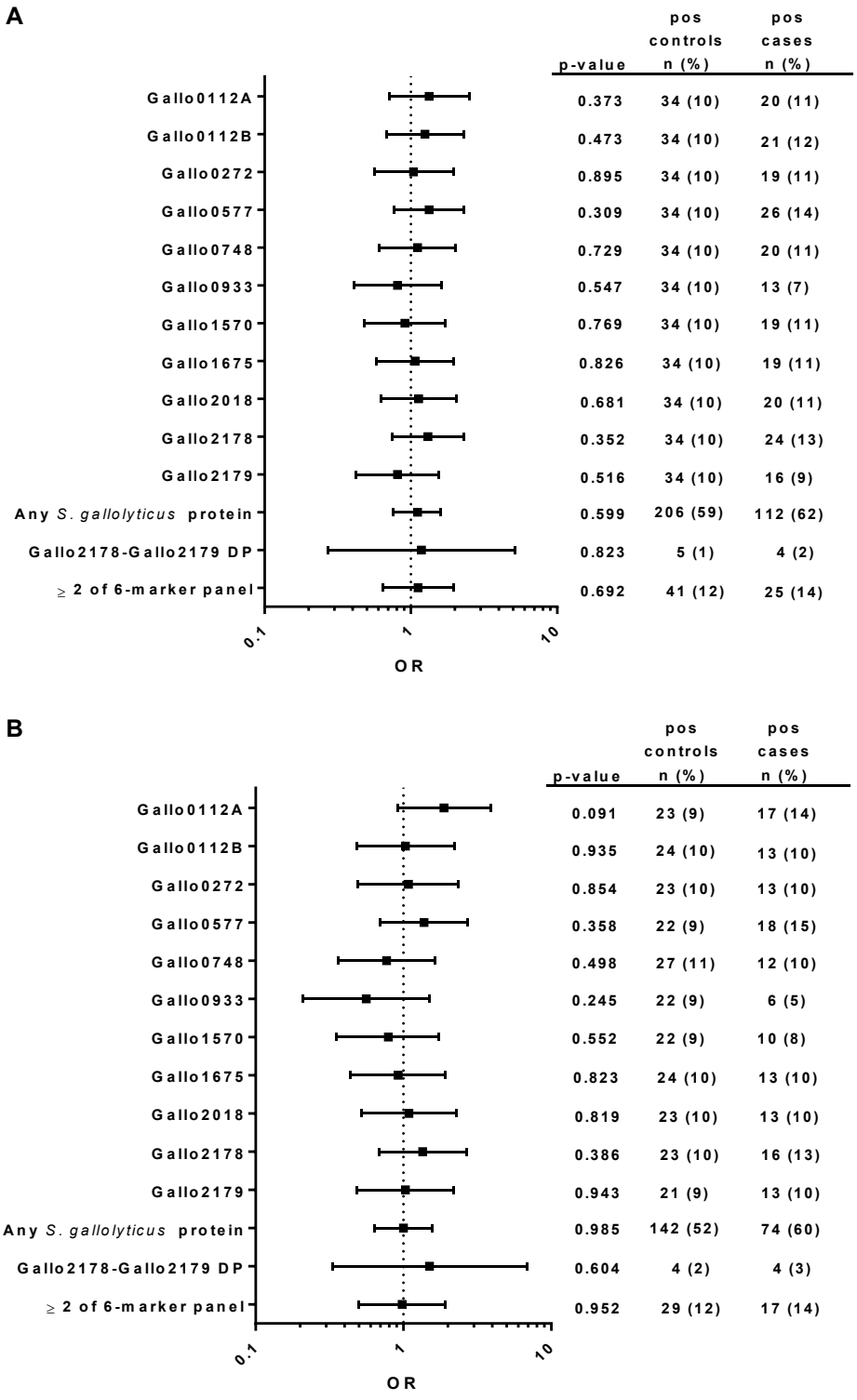


Figure 29: Antibody-positivity to individual *S. gallolyticus* proteins and combinations in relation to CRC risk in SCCS. A) Analyses with all individuals. B) Analyses of individuals diagnosed more than two years after blood draw. OR and 95% CI were estimated using conditional logistic regression models with adjustment for smoking, BMI and education. The dashed line indicates null association. 6-marker panel includes Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178 and Gallo2179. pos = antibody-positive.

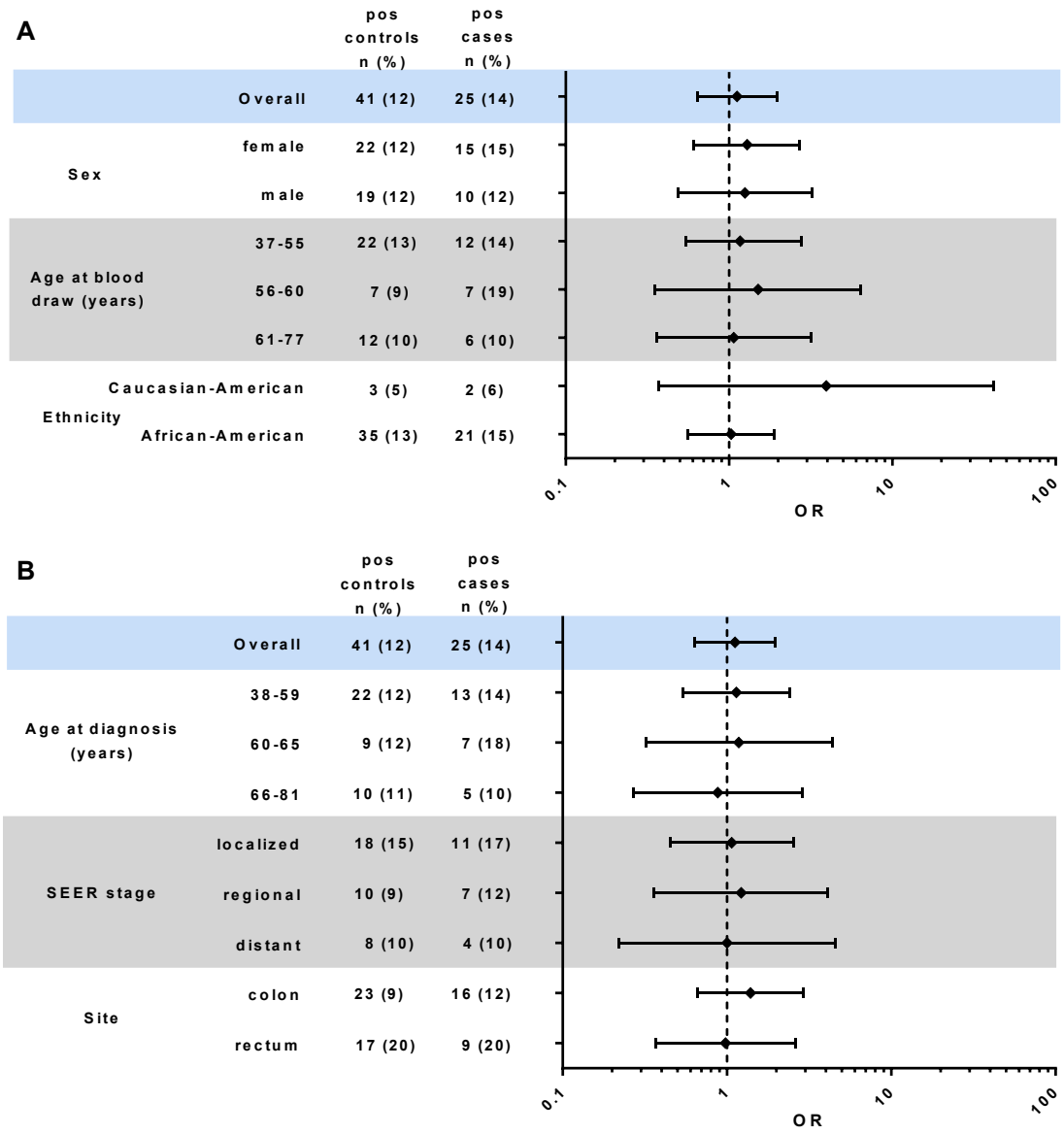


Figure 30: OR and 95% CI for antibody-positivity to two or more proteins of the *S. gallolyticus* 6-marker panel in relation to CRC risk in SCCS, overall and in separate conditional logistic regression models for (A) sex, age at blood draw and ethnicity as well as (B) for age at diagnosis, SEER stage and tumor site under adjustment for BMI, education, and smoking. The dashed line indicates null association. pos = antibody-positive.

3.4.5. Antibody responses to *S. gallolyticus* and incident CRC in the EPIC study

A second prospective nested CRC case-control study was analyzed for an association of antibody responses to *S. gallolyticus* with CRC risk. In this study I was able to reproduce findings from BliTz-DACHSplus in a prospective setting: positivity to Gallo2178 individually (OR: 2.74, 95% CI: 1.39-5.40), double-positivity to Gallo2178-Gallo2179 (OR: 7.02, 95% CI: 1.52-32.51), positivity to any of the eleven *S. gallolyticus* proteins (OR: 1.36, 95%CI: 1.05-1.78) as well as to two or more proteins of the 6-maker panel (OR: 2.10, 95% CI: 1.40-3.14) were significantly associated with CRC risk (Figure 31A). Positivity to two or more proteins of the 6-marker panel even remained significant (p-value 0.0004) after Bonferroni-correction for multiple testing (p-value < 0.0036). In addition to the replicated findings also antibody responses to individual proteins Gallo0272 (OR: 1.52, 95% CI: 1.01-2.29) and Gallo0748 (OR: 1.49, 95% CI: 1.02-2.17) were significantly associated with CRC, however, only without correction for multiple testing. Sensitivity analyses with exclusion of those participants diagnosed within 2 years after blood draw revealed similar associations as described above with the exception of Gallo0748, which was not significantly associated anymore (OR: 1.38, 95% CI: 0.89-2.15). In addition, the statistical power with the reduced sample size was not sufficient for positivity to two or more proteins of the 6-marker panel being significantly associated (p-value 0.006) after correction for multiple testing (Figure 31B). Thus, I was able to reproduce findings of the BliTz-DACHSplus study in the prospective EPIC study setting showing that antibody responses to *S. gallolyticus* proteins were significantly associated with CRC more than two years prior to diagnosis.

Interestingly, stratification by age at blood draw showed that CRC in the youngest age group (37-55 years) was not significantly associated with positivity to two or more proteins of the 6-marker panel (OR: 1.12, 95% CI: 0.49-2.56) compared to the two other age groups examined (56-60 years (OR: 3.94, 95% CI: 1.27-12.25) and 61-77 years (OR: 2.41, 95% CI: 1.32-4.39)) (Figure 32A). Stratification by country of residence showed a slight difference in the estimates ranging from Germany/Netherlands with a non-significant OR of 1.73 to a significant association in Italy/Greece (OR: 3.48, 95% CI: 1.21-10.03). There was no difference in the estimate by stage, however, by tumor site: rectal cancer cases showed a stronger association with positivity to two or more proteins of the 6-marker panel (OR: 22.16, 95% CI: 1.58-311.11) than colon cancer cases (OR: 1.92, 95% CI: 1.26-2.94). However, confidence intervals were strongly overlapping due to the small sample size of rectal cancers (Figure 32B).

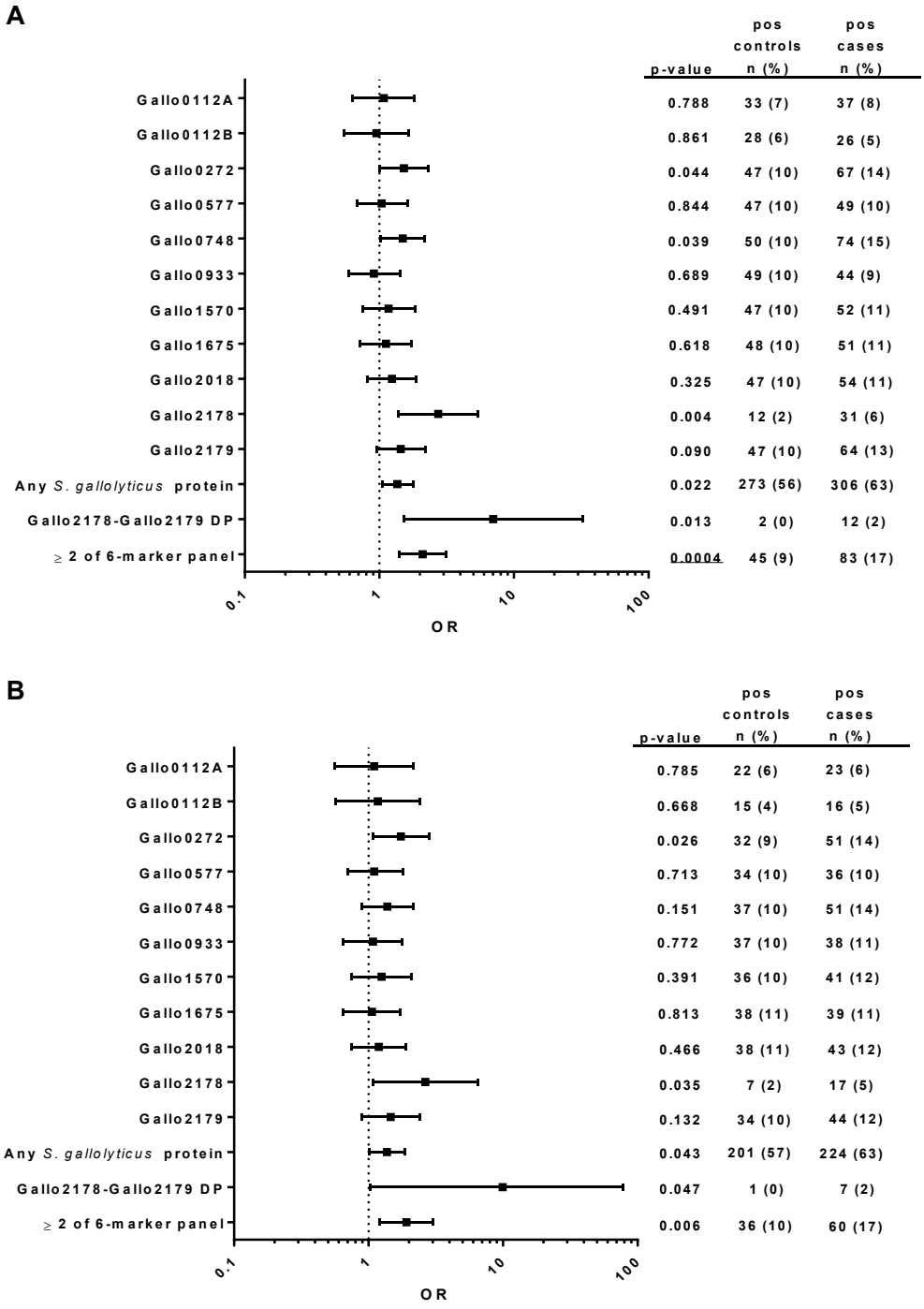


Figure 31: Antibody-positivity to *S. gallolyticus* proteins and combinations in relation to CRC risk in EPIC. A) Analyses with all individuals. B) Analyses with individuals diagnosed more than two years after blood draw. OR and 95% CI were estimated using conditional logistic regression models with adjustment for smoking, alcohol status, BMI and education. The dashed line indicates null association. Significant associations after Bonferroni-correction (p-value < 0.0036) are underlined. 6-marker panel includes Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178 and Gallo2179. pos = antibody-positive.

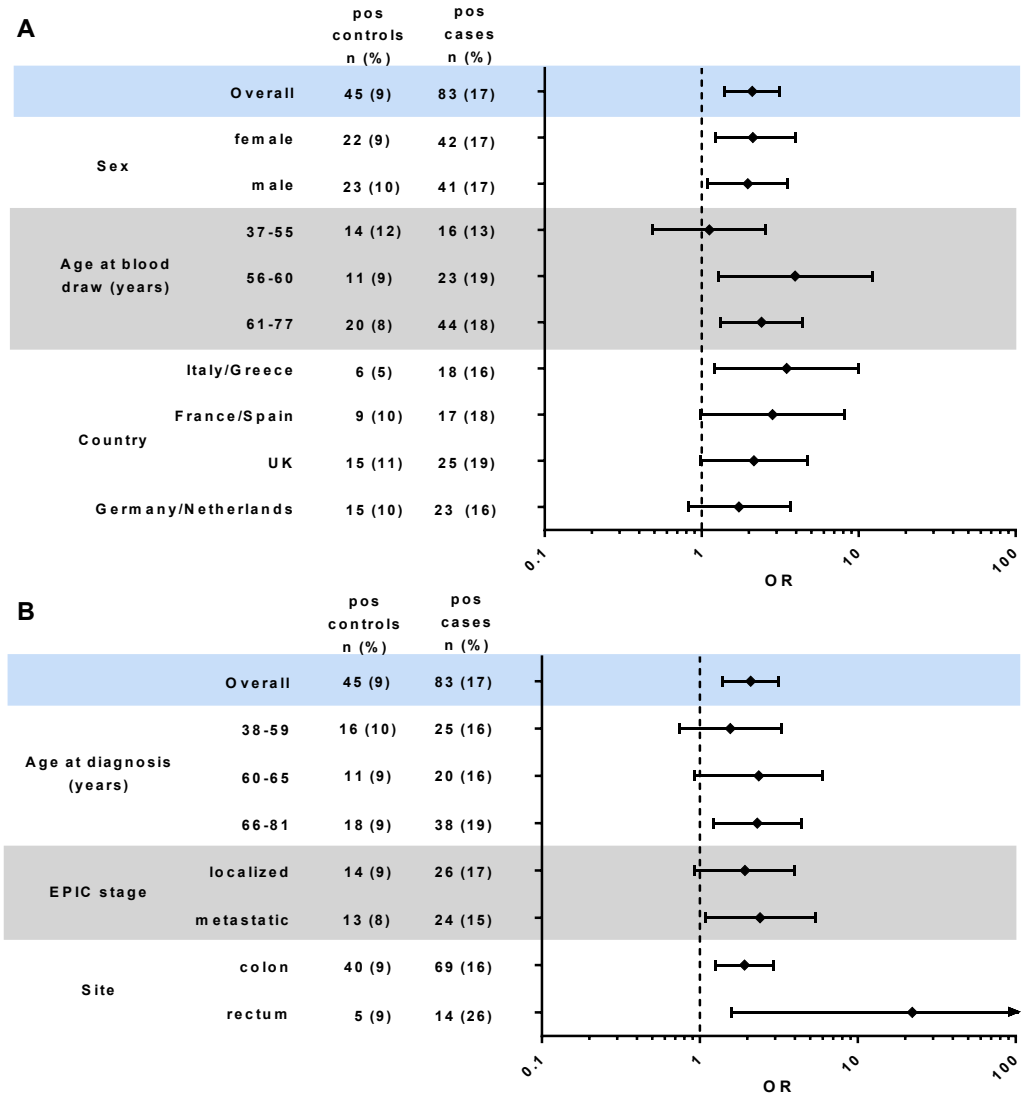


Figure 32: OR and 95% CI for antibody-positivity to two or more proteins of the *S. gallolyticus* 6-marker panel in relation to CRC risk in EPIC, overall and in separate conditional logistic regression models for (A) sex, age at blood draw and country of residence as well as (B) for age at diagnosis, stage (EPIC classification) and tumor site under adjustment for BMI, education, alcohol and smoking status. The dashed line indicates null association. pos = antibody-positive.

3.4.6. Exploration of *S. gallolyticus* multiplex serology in precancerous colorectal lesions

I found a significant association of antibody responses to *S. gallolyticus* proteins with prevalent CRC in the BliTz-DACHSplus study but also with the risk of developing CRC in the prospective EPIC study. I analyzed two studies, BliTz (1:100 serum dilution) and the Irish CRC case-control study (1:1000 serum dilution), for an association of antibody responses to *S. gallolyticus* proteins to assess the hypothesis that antibody responses can be detected prior to diagnosis and thus already in precancerous lesions. Both studies provided only small sample numbers of precancerous lesions ($n < 100$) and analyses should therefore be considered exploratory. Antibody responses to *S. gallolyticus* proteins in precancerous lesions and CRC cases in BliTz were compared to those in BliTz controls with the above described cut-offs (3.4.1). The Irish case-control study consisted of only 37 controls to generate antigen specific cut-offs (Table 34). Applied cut-offs were used to compare the fraction of positive individuals in Irish controls to those with minor diagnoses and colorectal neoplasm.

The fraction of individuals positive to two or more proteins of the 6-marker panel was 11% in BliTz controls and 8% in Irish controls. These numbers were exceeded by all diagnosed colorectal neoplasms (14%-27% in BliTz and 18%-26% in the Irish study) and those individuals with minor diagnoses in the Irish study (25%) (Table 35 and Table 36, respectively). These differences were significant for non-advanced adenoma cases in the BliTz study and for minor diagnoses, polyp and adenoma cases in the Irish study. Positivity to any *S. gallolyticus* protein was significantly different between Irish controls and individuals with minor diagnoses as well as with polyps. Double-positivity to Gallo2178-Gallo2179 was a rare event and did not differ among groups in the Irish study, however, was significantly different between controls and non-advanced adenoma cases in the BliTz study, even after Bonferroni-correction for multiple testing. Positivity to individual proteins differed significantly for Gallo0272 between Irish controls and individuals with polyps, adenomas and high-grade dysplasia; for Gallo0748 between Irish controls and individuals with minor diagnoses as well as high-grade dysplasia; for Gallo0933 between Irish controls and individuals with polyps; for Gallo2178 between BliTz controls and CRC cases; and for Gallo2179 with BliTz controls and non-advanced adenoma cases as well as with Irish controls and CRC cases.

Thus, despite small sample numbers in the BliTz and Irish CRC case-control studies I observed an overall trend for a higher fraction of individuals antibody-positive to *S. gallolyticus* proteins and combinations in precancerous lesions and CRC cases compared to controls. However, both studies were rather inconsistent in their individual results, probably due to the statistical imprecision resulting from the small sample numbers in both studies.

Table 34: *S. gallolyticus* antigen-specific cut-offs in the Irish CRC case-control study

Antigen	Cut-off (MFI)
Gallo0112A	30
Gallo0112B	30
Gallo0272	91
Gallo0577	193
Gallo0748	73
Gallo0933	108
Gallo1570	162
Gallo1675	109
Gallo2018	111
Gallo2178	14
Gallo2179	132

Table 35: Fraction of *S. gallolyticus* antibody-positive individuals among controls in comparison to individuals with colorectal neoplasm identified during colonoscopy screening in the BliTz study

	Controls (n=228)	Non-advanced adenoma[1] (n=30)	p-value[3]	Advanced adenoma[2] (n=100)	p-value[3]	CRC (n=50)	p-value[3]
Gallo0112A	22 (10)	2 (7)	0.597	5 (5)	0.159	3 (6)	0.414
Gallo0112B	22 (10)	3 (10)	0.951	12 (12)	0.520	1 (2)	0.075
Gallo0272	22 (10)	5 (17)	0.238	12 (12)	0.520	6 (12)	0.617
Gallo0577	22 (10)	4 (13)	0.529	5 (5)	0.159	4 (8)	0.717
Gallo0748	22 (10)	3 (10)	0.951	16 (16)	0.098	8 (16)	0.190
Gallo0933	22 (10)	4 (13)	0.529	17 (17)	0.058	7 (14)	0.362
Gallo1570	22 (10)	4 (13)	0.529	4 (4)	0.081	5 (10)	0.940
Gallo1675	22 (10)	6 (20)	0.087	8 (8)	0.633	6 (12)	0.617
Gallo2018	22 (10)	3 (10)	0.951	12 (12)	0.520	9 (18)	0.089
Gallo2178	11 (5)	3 (10)	0.240	3 (3)	0.452	7 (14)	**0.017**
Gallo2179	22 (10)	7 (23)	**0.026**	11 (11)	0.708	3 (6)	0.414
Any *S. gallolyticus* protein	131 (57)	20 (67)	0.336	65 (65)	0.200	34 (68)	0.169
Gallo2178-Gallo2179 DP	0 (0)	2 (7)	**<0.0001**	2 (2)	**0.032**	0 (0)	
≥ 2 of 6-marker panel[4]	24 (11)	8 (27)	**0.012**	14 (14)	0.366	8 (16)	0.272

[1]tubular adenoma, adenoma <10mm; [2]High grade dysplasia, villous adenoma with high-grade dysplasia, large adenoma (> 10 mm) with neither high-grade dysplasia nor villous architecture; [3]Pearson's Chi-square test in comparison to controls; [4]6-marker panel includes: Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178 and Gallo2179; Significant associations after Bonferroni-correction (p-value < 0.0036) are underlined. DP = double-positivity

Table 36: Fraction of *S. gallolyticus* antibody-positive individuals among controls in comparison to individuals with minor diagnoses or individuals with colorectal neoplasm diagnosed with colonoscopy in the Irish CRC and precursors sample collection

	Controls (n=37)	Minor diagnoses[1] (n=63)	p-value[4]	Polyp[2] (n=85)	p-value[4]	Adenoma[3] (n=60)	p-value[4]	High grade dysplasia (n=22)	p-value[4]	CRC (n=25)	p-value[4]
Gallo0112A	2 (5)	4 (6)	0.848	7 (8)	0.583	4 (7)	0.802	2 (9)	0.586	3 (12)	0.350
Gallo0112B	1 (3)	1 (2)	0.701	6 (7)	0.342	2 (3)	0.862	0 (0)	0.437	1 (4)	0.777
Gallo0272	3 (8)	14 (22)	0.070	23 (27)	**0.019**	17 (28)	**0.017**	6 (27)	**0.048**	1 (4)	0.518
Gallo0577	3 (8)	9 (14)	0.359	13 (15)	0.280	3 (5)	0.537	2 (9)	0.896	4 (16)	0.335
Gallo0748	3 (8)	16 (25)	**0.033**	19 (22)	0.060	11 (18)	0.164	8 (36)	**0.007**	6 (24)	0.081
Gallo0933	3 (8)	14 (22)	0.070	23 (27)	**0.019**	6 (10)	0.755	5 (23)	0.113	3 (12)	0.611
Gallo1570	3 (8)	10 (16)	0.265	11 (13)	0.441	4 (7)	0.790	5 (23)	0.113	2 (8)	0.988
Gallo1675	3 (8)	6 (10)	0.811	11 (13)	0.441	5 (8)	0.969	1 (5)	0.599	1 (4)	0.518
Gallo2018	3 (8)	8 (13)	0.479	15 (18)	0.172	8 (13)	0.431	3 (14)	0.497	5 (20)	0.171
Gallo2178	2 (5)	5 (8)	0.632	10 (12)	0.278	3 (5)	0.930	1 (5)	0.884	2 (8)	0.683
Gallo2179	3 (8)	8 (13)	0.479	9 (11)	0.672	10 (17)	0.229	2 (9)	0.896	7 (28)	**0.037**
Any *S. gallolyticus* protein	21 (57)	48 (76)	**0.043**	65 (76)	**0.028**	40 (67)	0.326	16 (73)	0.220	18 (72)	0.223
Gallo2178-Gallo2179 DP	0	1 (2)	0.441	1 (1)	0.508	1 (2)	0.430	0		0	
≥2 of 6-marker panel[5]	3 (8)	16 (25)	**0.033**	22 (26)	**0.025**	15 (25)	**0.038**	4 (18)	0.247	6 (24)	0.081

[1]includes hemorrhoids, diverticulosis, mucosal ulceration/inflammation, melanosis coli, menorrhagia/disordered proliferative endometrium, mild active colitis, diarrhea, erythema; [2]hyperplastic polyp or small tubular adenoma (< 10mm); [3]tubular adenoma, tubulovillous adenoma, villous adenoma; [4]Pearson's Chi-square test in comparison to controls; [5]6-marker panel includes: Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178 and Gallo2179; DP = double-positivity

4. Discussion

CRC is the third most common type of cancer worldwide [1]. A connection of CRC with alterations in the gastrointestinal microbiome has gained increasing interest in recent years [25]. In particular, CRC prevalence was found to be associated with two bacterial species: *F. nucleatum* and *S. gallolyticus* [33]. However, it still remains unclear whether associated bacterial infections are a consequence of tumor development or whether they play a causal role in tumorigenesis or progression of the disease. The latter could have importance in prevention by eradication of the bacterium. Knowledge on the former, especially at which stage or precancerous lesion this association is present, might be of diagnostic value in the early detection of CRC.

Serology is a time- and cost-effective methodology, especially compared to PCR- and sequencing-based methods, to analyze infection markers in epidemiological studies. Multiplex serology, developed by Tim Waterboer et al. allows the simultaneous analyses of up to 100 antigens in one reaction and thus is a high-throughput serology application [114].

In this thesis, I developed multiplex serology assays for the detection of antibodies directed against eleven proteins each of *F. nucleatum* and *S. gallolyticus*. These were applied in one retrospective case-control study as well as two independent case-control studies nested within prospective cohorts. Gold-standard assays for serological validation of *F. nucleatum* and *S. gallolyticus* multiplex serology assays were not available. Attempts to validate the assays by comparison to DNA status in colon or fecal samples indicated specific antibody responses to *F. nucleatum* and *S. gallolyticus* proteins, however, small sample sizes resulted in non-significant differences. Experimentally infected mice and samples from humans with *S. gallolyticus* bacteremia further strengthened the observations, however, also here sample sizes were not sufficient to define robust cut-offs for antibody-positivity. Though, with arbitrary definitions for antibody-positivity I was able to identify repeatedly an association of prevalent and incident CRC cases with antibody responses to *S. gallolyticus* proteins in a German case-control study (BliTz-DACHSplus) and a prospective European study (EPIC) but interestingly not in an US-American study (SCCS). In contrast, I could not identify a serological association of *F. nucleatum* with CRC, neither retro- nor prospectively.

4.1. *F. nucleatum* multiplex serology and CRC

4.1.1. Development of *F. nucleatum* multiplex serology

Multiplex serology provides the possibility of the simultaneous analyses of several antigens in one reaction [114]. This is of great advantage especially in the seroepidemiological analysis of bacterial infections since bacteria express a large variety of potential immunogenic proteins that are presented to the immune system. *H. pylori* multiplex serology for example includes 15 different immunogenic proteins giving a high sensitivity and specificity for detection of present and past *H. pylori* infection [143]. Previous studies on *F. nucleatum* serology used whole cell protein extracts [144-146]. This method is prone to lack specificity due to cross-reacting antibody responses induced by proteins from other bacterial species, which share high homology to *F. nucleatum* proteins.

Only few species-specific and immunogenic proteins of *F. nucleatum* have been identified so far. The adhesin FadA (Fn0264) was applied in ELISA to analyze antibody responses in saliva and plasma of periodontitis patients [117]. The outer membrane protein FomA (Fn1859) [118] was used to immunize mice, which developed plasma IgG and IgA antibody responses against FomA. A very recent study identified FomA also as immunogenic protein in CRC patients that were *F. nucleatum* DNA-positive in stool [147]. These two proteins have been included in the *F. nucleatum* multiplex serology assay. The selection of additional nine proteins was based on their localization and potential function in the bacterium since experience from *H. pylori* multiplex serology had revealed that proteins located at the outer membrane or secreted as well as those important in virulence are immunogenic and associated with disease [113, 143]. As described in 1.2.1 proteins Fap2 (Fn1449) and RadD (Fn1526) may have virulence functions, on the one hand by modulating the immune system and on the other in adhesion and biofilm formation [52, 53, 59, 60]. These two proteins are members of a type Va secretion pathway [58]. They represent autotransporters, polypeptides which possess three domains for translocating effectors (signal sequence) over the inner (passenger domain) and outer membrane (translocation unit) to be finally secreted. Such autotransporters were described in other bacteria to be of importance in virulence of the bacterium. A well-known autotransporter in *H. pylori* is the Vacuolating Cytotoxin A (VacA), which inserts into the host cell plasma membrane and induces vacuoles which lead to cell death [148]. VacA also induces antibody responses that are associated with gastric cancer [149]. Other *F. nucleatum*

proteins included in the multiplex serology and belonging to the Type Va secretion system were Fn0387, Fn1426 and Fn1893.

Fn0131 and Fn1817, in contrast, belong to the type Vb secretion pathway, which is a two-partner secretion pathway of a transporter (Fn0131) and a secreted protein (Fn1817) [58, 148]. Finally Fn0253 was identified as a protein located at the outer membrane and therefore prone to be detected by the host immune system [122].

Specificity of selected proteins was addressed *in silico*. This approach is limited to genomes of sequenced bacteria and thus potentially misses sequences of yet unidentified and non-sequenced bacterial species. Amino acid sequences were compared for percentage identity to homologous proteins in other bacteria. The reference strain ATCC 25586 belongs to the *F. nucleatum* subspecies *nucleatum*. Proteins of other *F. nucleatum* subspecies shared a more than 74% amino acid sequence identity with the selected proteins. Thus, detected antibody responses are probably not subspecies-specific. The selected proteins have a high amino acid identity (up to 97%) to proteins of a newly identified fusobacterial species, *F. hwasookii* [124]. A serological distinction between *F. hwasookii* and *F. nucleatum* cannot be guaranteed. *F. hwasookii* was first isolated from a human periodontitis lesion in 2015 [124], however, since then was not further reported in the literature. The epidemiology of this bacterium has not been assessed to date and the impact in human infections cannot be inferred from the available literature. The next known bacterial relative is *F. periodonticum*. Except for Fn0264, selected proteins shared a sequence identity of less than 80% with proteins of this bacterium. Other, none fusobacterial species did not exceed an identity of 56%. Thus, although cross-reactive antibody responses to linear epitopes were unlikely, cross-reactive responses to conformational epitopes cannot be excluded. Homologous proteins could be generated to measure potential cross-reactive antibody responses, however, this was not possible within the scope of this thesis.

The eleven selected *F. nucleatum* proteins were recombinantly expressed and underwent quality controls, both, on the DNA- and protein level to ascertain expression of the correct and functional full-length antigen in acceptable quantity. Quality controls on the DNA level included an analytical restriction digestion of plasmids isolated from an intermediate step during the expression procedure as well as PCR of the final lysate with subsequent sequencing. Both controls identified the presence of all expected expression constructs in the respective lysates. Quality control on the protein level included Western blot with antibodies directed against the N-terminal GST and, to detect full-length fusion protein, against the C-

terminal tag. Additionally, an anti-tag ELISA was performed for the quantification of glutathione-binding full-length protein relative to GST-tag as reference. Seven out of eleven proteins could be detected in the anti-tag blot of the final lysate at the expected protein size. These proteins also needed less than 30-fold of the total protein concentration relative to GST-tag to reach the half-maximum OD of GST-tag. Low amounts of full-length fusion proteins were found for Fn0131, Fn1426, Fn1526 and Fn1859 in the final lysates indicating partial insolubility. All four proteins were predicted to be located in the cell membrane. Potential transmembrane domains that were not identified with the applied prediction programs might have led to hydrophobicity and consequently insolubility. Additionally, Fn1859 migrated like an approximately 10 kD smaller protein. The protein was most probably N-terminally degraded since the C-terminal tag was detectable. The polyclonal anti GST-antibody, in contrast to monoclonal anti-tag antibody, recognizes several epitopes in GST explaining why bands in the anti-GST blot could be detected despite the presumed N-terminal degradation.

A final quality control on the protein level was performed by loading of antigens onto glutathione-casein coupled beads at a lysate protein concentration of 1 mg/ml. Subsequent loading control with anti-tag antibody indicated similar amounts of full-length protein on the beads for all *F. nucleatum* proteins as compared to GST-tag lysate. Thus, all recombinantly expressed *F. nucleatum* proteins were finally applied in multiplex serology.

In addition to the technical controls it was attempted to serologically validate *F. nucleatum* multiplex serology. A clear limitation thereby was the lack of a serological gold standard assay for comparison to the newly developed multiplex serology. In the attempt to nonetheless validate the assay I used serum samples of patients with adenoma or CRC in whom *F. nucleatum* DNA had been found in the tumor and/or surrounding normal tissue. Antibody levels (MFI values) to none of the *F. nucleatum* proteins were able to discriminate between *F. nucleatum* colon-tissue DNA-negative and –positive individuals. Experience from *H. pylori* multiplex serology had shown that positivity to several proteins simultaneously increased specificity for detection of the infection [143]. Application of a cut-off derived from the mean MFI plus three times standard deviation in controls revealed a trend of a higher frequency of *F. nucleatum* antibody-positives in DNA-positives compared to –negatives, with positivity to individual but also to more than one of the proteins. However, due to the small

sample size and the weakness of the differences there was no significant correlation observable.

As described in 1.3.2 amplification of DNA of infectious agents only detects acute infections at the specific site analyzed. Serology, however, detects acute and past infections independent of the site of infection. *F. nucleatum* is most predominant in the oral cavity, another infection site possibly leading to antibody responses. This and the detection of antibody responses resulting from possible past infections may have led to the difficulties in discrimination between colonic-tissue DNA-negative and –positive individuals by multiplex serology.

4.1.2. Antibody responses to *F. nucleatum* and CRC in a retrospective case-control study and two independent case-control studies nested within prospective cohorts

F. nucleatum multiplex serology was applied in one German CRC case-control study (BliTz-DACHSplus) and two independent case-control studies nested within prospective cohorts: one multi-center cohort from southern US (SCCS) and one multi-national cohort conducted across Europe (EPIC). A clear limitation in the analyses was the lack of an infection-based cut-off for *F. nucleatum* antibody-positivity resulting from the above described absence of appropriate reference assay and samples. Therefore, the cut-off had to be defined arbitrarily and was set for individual *F. nucleatum* proteins to the MFI that defined 10% of control individuals as being positive. The distribution of MFI in BliTz controls supported this cut-off definition: The overall antibody responses to individual proteins were low and the cut-off was similar to the approximate point of inflection in the antibody distribution curve. Visual inspection of these so-called percentile plots was previously used in our laboratory to define cut-offs since it is assumed that a sudden rise in the antibody response over percentile of serum indicates the cut-off for antibody-positivity ([128] and Michael Pawlita, personal communication). Cut-offs were defined per study. The SCCS study was analyzed in the same experimental run as the BliTz-DACHSplus study and cut-offs can therefore directly be compared. Cut-offs were similar for the individual proteins among studies indicating robustness. However, for one protein, Fn1817_1, the cut-off defining 10% of controls as positive was 15-times higher in the SCCS (2039 MFI) than in the BliTz-DACHSplus study (133 MFI). The only baseline characteristic in SCCS controls significantly associated with antibody-positivity to Fn1817_1 was being African-American indicating a difference in

antibody response by ethnicity that will be discussed later in the context of *S. gallolyticus* multiplex serology.

The arbitrarily defined cut-offs for individual proteins were applied to assess whether there is an association of antibody responses measured in *F. nucleatum* multiplex serology with prevalent and incident CRC. Overall *F. nucleatum* positivity was defined as being positive to any of the eleven proteins to allow inter-individual differences in the immune response but also infection with different bacterial strains. This algorithm resulted in similar fractions of *F. nucleatum* antibody-positive controls in all three studies (57% in BliTz-DACHSplus, 59% in SCCS and 53% in EPIC). These were compared to *F. nucleatum* antibody-negative controls in their baseline characteristics. The only significant difference between *F. nucleatum* negative and positive controls was identified in the BliTz-DACHSplus study with sex distribution. Here, more males were *F. nucleatum* antibody-positive than females. This trend was also observed in SCCS and EPIC, however, it was not significant. A review by Albandar identified studies showing that periodontitis, a disease associated with increased gingival levels of *F. nucleatum* was more frequent in adult males than in females independent of age [150]. However, a specific association of sex with *F. nucleatum* infection was not reported so far.

I did not observe a significant association of positivity to any *F. nucleatum* protein or individual *F. nucleatum* proteins with neither prevalent nor incident CRC in any of the studies. Positivity to individual *F. nucleatum* proteins correlated, which was expected from proteins of the same organism, however, not specifically in CRC patients. These results are not concordant with the PCR-based studies presented in 1.2.1 that all found an association of *F. nucleatum* with CRC [37-42]. Recently, a study by Wang et al. was published where the authors used *F. nucleatum* whole cell protein ELISA to analyze serum samples of 200 healthy subjects, 50 benign colon disease patients and 258 CRC patients for differences in antibody responses to *F. nucleatum* [147]. They found significant differences in antibody responses between healthy controls and CRC patients. Application of a cut-off based on 96% specificity resulted in sensitivities of detecting overall CRC of 31% with a secondary antibody against IgA and 19% with anti-IgG. In addition they applied Western blot analyses with subsequent mass spectrometry analyses to identify specific immunogenic *F. nucleatum* proteins detected by anitbodies in serum from CRC patients with *F. nucleatum* DNA-positive stool. Fn1859, but none of the other *F. nucleatum* multiplex serology proteins, was among the identified proteins. However, also Fn1859 was neither associated with CRC nor with *F. nucleatum*

DNA-positivity in colon tissue (3.2.1) in the studies presented here. Different possible reasons for the discordance of the results presented here with the current literature can be inferred from the study by Wang et al. [147]:

i) The selection of antigens. The only identified immunogenic protein by Wang et al. included in multiplex serology was Fn1859. However, in their association study with CRC they even used *F. nucleatum* whole cell protein instead of recombinantly expressed proteins [147]. Potentially, the proteins selected for multiplex serology are of low sensitivity in detecting antibody responses against *F. nucleatum*, which could be reflected by the overall low antibody responses in all three studies presented here.

ii) The selection of secondary antibody. Wang et al. analyzed IgA and IgG antibody responses separately and found differing sensitivities in detection of antibody responses against *F. nucleatum* [147]. *F. nucleatum* was identified in CRC by detection of the DNA in tumor tissue, rather than by a secondary bacteremia as seen with *S. gallolyticus*. It is probable that *F. nucleatum* resides in the colonic epithelial/mucosal tissue, where the main isotype of antibodies secreted is IgA [151]. The secondary antibody applied in multiplex serology is directed against IgG, IgA and IgM simultaneously, however, a more specific detection of the three isotypes could be beneficial in *F. nucleatum* serology.

iii) The selection of study subjects. The association of *F. nucleatum* DNA with CRC was predominantly found in proximal tumors, higher stage tumors, tumors of larger size, MSI-H tumors, and those of the serrated pathway [47-50, 152]. Separate analyses by stage in the studies presented here did not reveal differences in the OR. The only significant differences observed between subgroups were an inverse association in females of the BliTz-DACHSplus and EPIC studies in contrast to a null association among males and an inverse association with rectal in contrast to colon cases in SCCS. These differences could not be reproduced over all studies and might result from an underlying sample characteristic that was not analyzable with the given data. However, the differences might also have occurred simply by chance. Wang et al. did not further specify the types of cases they included [147]. It would be interesting to further analyze studies with cases more specifically characterized by morphology and histology of the tumor to assess whether positivity with *F. nucleatum* multiplex serology is associated with a

certain subtype of CRC that may have been underrepresented in the studies analyzed in this thesis.

4.2. *S. gallolyticus* multiplex serology and CRC

4.2.1. Development of *S. gallolyticus* multiplex serology

In a previous study I used recombinantly expressed and purified his-tagged *S. gallolyticus* pilus proteins (Gallo1569, Gallo2039, Gallo2178 and Gallo2179) in multiplex serology to analyze a Spanish multi-center CRC case-control study [81]. Two of these pilus proteins, Gallo2178 and Gallo2179 were individually associated with CRC (OR of 1.5) and double-positivity to both significantly increased odds for CRC 3.6-fold. However, only 1% of controls were double-positive compared to 4%, and thus only a minor fraction, of cases.

In the attempt to possibly increase sensitivity by the inclusion of several more *S. gallolyticus* proteins I further selected nine additional proteins for *S. gallolyticus* multiplex serology. Similar to *F. nucleatum*, selection of proteins was based on predicted function and localization of the proteins in the bacterium since information on immunogenic proteins in literature was rare. *S. gallolyticus* is a gram-positive bacterium possessing a cell wall as an outer shell. Sillanpää et al. and Hinse et al. listed several proteins potentially localized at the cell wall due to the presence of a LpxTG signal motif [121, 123]. These proteins included Gallo2178 and Gallo2179 as well as Gallo0112, a putative fructan hydrolase, Gallo0272, a putative agglutinin receptor, Gallo0577, a protein containing a cnaB domain, which is frequently found in pilus structures, Gallo0748, a putative proteinase, Gallo1570, also a subunit of a pilus structure, Gallo1675 with unknown function and Gallo2018, a protein putatively involved in bacteriocin synthesis. I additionally included an enzyme that is putatively secreted from the bacterium, Gallo0933, a potential tannase. Tannases degrade toxic tannins found in plants to gallic acid, which can be further used as carbon supply for the bacterium and may represent a survival advantage compared to other bacteria not expressing tannases [119].

Comparison of amino acid sequences to proteins of other bacteria identified several *S. gallolyticus* proteins sharing a high sequence identity with proteins of *S. equinus* (up to 97%). However, this bacterium is rarely isolated from humans [66]. Other closely related species of *S. gallolyticus*, *S. pasteurianus* and *S. infantarius* have homologous proteins to Gallo1675 (99% sequence identity) and Gallo0748 (67%), respectively. The sequence identity was below

75% for all other *S. gallolyticus* proteins to those of species of the *S. equinus*/*S. bovis* complex, the family of Streptococcacae or none streptococcal species indicating a high specificity of the assay. However, as described for *F. nucleatum*, also here antibodies cross-reacting to conformational epitopes cannot be excluded.

Most of the recombinantly expressed *S. gallolyticus* proteins showed lower relative full-length protein concentration in anti-tag ELISA, while all proteins were identified at the expected sizes in the anti-tag blot of cleared lysate. The anti-tag epitope might have been hidden in anti-tag ELISA, where proteins should keep their conformation. In the application in multiplex serology, however, proteins loaded onto beads were detected in comparable quantities to GST-tag indicating also for *S. gallolyticus* proteins that there is a saturating excess of full-length protein in the amount of lysate applied in bead-loading.

There is no gold standard assay available for the serological validation of *S. gallolyticus* multiplex serology. However, I had three different possibilities to potentially address serological validation: i) A time-series of sera from mice after inoculation with *S. gallolyticus* UCN34 or PBS; ii) serum samples from dairy cows with additional information on *S. gallolyticus* DNA found in feces; iii) serum samples from three individuals with diagnosed *S. gallolyticus* bacteremia and three healthy controls. The three different attempts resulted in similar pattern of antibody responses to *S. gallolyticus* proteins. Antibody responses to Gallo0272 and Gallo2179 distinguished defined positives from negatives in all three hosts. Analysis of the human serum samples from individuals diagnosed with *S. gallolyticus* bacteremia probably resembles best the later application in serum collections of human origin. Gallo2178 was recognized specifically by sera of patients diagnosed with *S. gallolyticus* bacteremia together with Gallo0272 and Gallo2179 but also Gallo0577, Gallo0748, Gallo1570 and Gallo2018. Also these proteins mostly overlap with the proteins that best distinguished *S. gallolyticus* DNA-negative from -positive cows as well as experimentally infected from non-infected mice. In addition, the finding that patients with *S. gallolyticus* bacteremia raised antibody responses to several proteins simultaneously was concordant to the observation in bovine sera, where positivity to several proteins strengthened the specificity in comparison to DNA positivity.

Comparison of serology with PCR-based data has two major pitfalls: Serology measures present and past systemic infections, whereas PCR measures present infection at a specific site. I cannot exclude that DNA-negative cows never had a past *S. gallolyticus* infection

neither can I assure that *S. gallolyticus* is excreted with feces. Unfortunately, the number of human samples with known systemic *S. gallolyticus* infection status was small with three individuals in each, healthy control and *S. gallolyticus*-bacteremia patient groups. Larger serum collections would have been needed to define robust cut-offs for individual proteins and an algorithm for *S. gallolyticus* overall positivity.

4.2.2. Antibody responses to *S. gallolyticus* and CRC in a German case-control study and an independent case-control study nested within a European prospective cohort

A limitation, also for *S. gallolyticus*, was the arbitrary cut-off definition for the analyses of seroepidemiological studies in BliTz, SCCS and EPIC. The fecal carriage rate of *S. gallolyticus* in healthy individuals was found to range between 1 and 11% [78, 79]. In a previous multiplex serology study, where four *S. gallolyticus* pilus proteins were applied as antigens in a Spanish multi-center CRC case-control study, the cut-offs were arbitrarily set to define 10% of controls being positive to each of the four *S. gallolyticus* proteins. The overall antibody responses to the newly developed eleven *S. gallolyticus* GST-X-tag fusion proteins in colonoscopy-negative BliTz controls was low, reflecting the expected low fecal carriage rate: A MFI of 500 was achieved by more than 30% of the serum samples only with antigens Gallo0272, Gallo0933, Gallo1570 and Gallo2018. Cut-offs defining 10% of the controls as being positive for the individual proteins ranged from 30 MFI as a technical minimum cut-off for Gallo2178 to above 1000 MFI for Gallo0272, Gallo0933, Gallo1570 and Gallo2018. Comparison with inflection points of percentile plots showed that the here arbitrarily defined cut-offs were more stringent. However, to make the results comparable to the previous already published CRC case-control study it was decided to further apply the 10% cut-off. In addition, a more stringent cut-off provides a higher specificity in detecting *S. gallolyticus* and, if associated, CRC-specific antibody responses.

In a first attempt overall *S. gallolyticus* antibody-positivity was defined as being positive to any of the eleven proteins. This resulted in 57% positive BliTz controls, 59% positive SCCS controls and 56% positive EPIC controls. These fractions were much larger than the reported low fecal carriage rates (1-11%). However, to allow for inter-individual differences in the immune response but also infection with different bacterial strains, I applied this algorithm to assess potential risk factors for antibody responses to *S. gallolyticus* in controls of the three studies. None of the here assessed baseline characteristics was significantly associated with

S. gallolyticus among controls and thus considered as a potential risk factor in healthy individuals. Case-control studies by Abdulamir, Boleij and Garza-Gonzalez addressed age and sex as potential risk factors and did not identify a significant association with antibody-positivity measured in ELISA or Western blot with whole cell protein [82-84]. In Butt et al., 2016, I addressed the same risk factors as presented here, age, sex, BMI, smoking, education and family history of CRC and did not identify significant risk factors for antibody responses to the four *S. gallolyticus* pilus proteins either [81]. Whether other characteristics like diet, lifestyle or co-morbidities are associated, remains to be elucidated.

In a first step in the analysis of an association of *S. gallolyticus* with CRC I attempted to reproduce findings from Butt et al. [81] with respect to an association of antibody responses to proteins Gallo2178 and Gallo2179 with CRC in the BliTz-DACHSplus case-control study. Indeed, positivity to Gallo2178, but not Gallo2179, was significantly associated with prevalent CRC with an OR of 4.5. This association even remained significant after Bonferroni-correction for multiple testing. Double-positivity to Gallo2178-Gallo2179 was a rare event, as seen in Butt et al. [81], with 0% double-positive controls compared to 4% double-positive cases. Since none of the controls was double-positive it was not possible to estimate the strength of the association, however, a crude p-value obtained by Chi-square test was even significant after correction for multiple testing.

None of the other individual *S. gallolyticus* proteins was associated with prevalent CRC in the BliTz-DACHSplus study. Positivity to any of the proteins was significantly associated with 1.6-fold increased odds for CRC, however, significance was not retained after correction for multiple testing. As discussed above analyses of serum samples from cattle and from *S. gallolyticus* bacteremia patients indicated a higher specificity for detecting *S. gallolyticus* specific antibodies when being positive to several *S. gallolyticus* proteins simultaneously. I assessed whether positivity to *S. gallolyticus* proteins significantly correlated similar to Gallo2178-Gallo2179. These analyses were done separately for BliTz controls and DACHSplus cases, since I hypothesized that correlations would occur preferentially in cases if *S. gallolyticus* infection was CRC specific. Positivity to three proteins, Gallo0577, Gallo1570 and Gallo2179, correlated strongly but independently of being a control or case. Amino acid sequence comparison identified long homologous stretches at the C-terminal part of the proteins that most probably led to cross-reacting antibodies mimicking a correlation in positivity between these proteins. In contrast, there were three antigen pairs, in addition to

Gallo2178-Gallo2179 that specifically correlated among CRC cases: Gallo0272-Gallo0748, Gallo0272-Gallo1675 and Gallo2018-Gallo2179. Except for Gallo1675 all of these proteins were among those best distinguishing serologically the bacteremia patients from healthy controls in the validation attempts. So far, functions of the newly identified proteins were only predicted by amino acid similarities to proteins of other bacterial species: i) Gallo0272 is a putative agglutinin receptor. The expressed domain resembles a glucan binding protein C domain. Agglutinin receptors in the oral bacterium *Streptococcus gordinii* mediate binding to host cell and bacterial receptors and may represent important virulence factors [153]. ii) Gallo0748 is supposed to have an endopeptidase function, which is important in protein turnover. iii) Gallo2018 is putatively involved in bacteriocin synthesis, inhibiting the growth of other bacteria in the surrounding and thereby mediating a growth advantage. iv) Gallo1675 is a cell wall protein with unknown function [121]. The in total 6 proteins, Gallo0272, Gallo0748, Gallo1675, Galo2018, Gallo2178 and Gallo2179 were combined to a 6-marker panel. To reflect the correlation between these markers, positivity was defined as being positive to at least two proteins of the 6-marker panel. Eleven percent of controls were positive with this newly defined algorithm, reflecting again the fecal carriage rate in healthy individuals. Compared to 19% of positive prevalent CRC cases this resulted in significantly 2-fold increased odds for CRC in the BliTz-DACHSplus study. Probably due to the relatively small sample size this association was not significant after correction for multiple testing. To further show that the observed association is CRC specific I compared BliTz controls to GC cases of the DACHSplus study for positivity to *S. gallolyticus*. Indeed, I did not observe an association of antibody responses to *S. gallolyticus* with GC undermining the specificity of the association seen with CRC.

The findings of the CRC case-control study BliTz-DACHSplus were further elucidated in a prospective setting in two independent nested case-control studies to assess a temporality of the association. One of the studies, EPIC, was a multi-national study, conducted in several countries of Europe, including Spain, Italy, France, Greece, the Netherlands and also Germany. I was able to reproduce findings from prevalent CRC also with incident CRC cases in EPIC: Positivity to Gallo2178 (OR: 2.7), double-positivity to Gallo2178-Gallo2179 (OR: 7), positivity to any *S. gallolyticus* protein (OR: 1.4) and positivity to two or more proteins of the 6-marker panel (OR: 2) were significantly associated. Additionally, positivity to Gallo0272 (OR: 1.5) and Gallo0748 (OR: 1.5), two proteins included in the 6-marker panel, were also individually associated with CRC risk. The positivity to two or more proteins of the

6-marker panel even remained significant after correction for multiple testing. A sensitivity analyses was carried out including only those cases diagnosed after two years from blood draw to exclude that observed associations are attributed to those cases most probably already having advanced colorectal tumors at time of blood draw. All associations seen with the complete dataset remained significant except for Gallo0748, which was not associated anymore with CRC diagnosed more than two years from blood draw. Also with correction for multiple testing the association of positivity to two or more proteins of the 6-marker panel with these cases becomes insignificant, however, also the sample size was reduced leading to reduced statistical power.

Thus, I showed that the association of antibody responses to *S. gallolyticus* with CRC in EPIC was present already several years prior to diagnosis. This is in concordance with the bacterial-driver passenger model developed by Harold Tjalsma (1.2) [33], defining *S. gallolyticus* as a passenger bacterium invading adenomatous tissue in the colon. The lag-time in EPIC ranged from 0.4 to 8.5 years with a median time of 3.3 years. Since CRC development is a long-term process taking several years from an initial polyp to malignant disease, most probably incident CRC cases already had a precancerous lesion at time of blood draw. Studies with a lag-time of more than ten years would be needed to assess whether antibody responses to *S. gallolyticus* can be detected even prior to development of early precancerous lesions. If so, *S. gallolyticus* could be causally linked to initiation of CRC development being a so-called driver bacterium in the driver-passenger model [33] or even an alpha-bac in the alpha-bac hypothesis [32], similar to *H. pylori* in the initiation of GC. However, this cannot be inferred from the available data and needs to be further elucidated.

An additional question, independently of the time-point of infection, arising from the results is whether *S. gallolyticus* may be involved in progression of the tumor. As discussed above incident CRC cases most probably already had a precancerous lesion at time of blood draw. However, also among controls presence of a certain number of precancerous lesions has to be expected. Data from the German screening colonoscopy registry showed that the detection rate of non-advanced adenomas was 22.3% among males and 14.9% among females aged above 55 years in 2012 [154]. The detection rate for advanced adenomas in the same age groups was 9% for males and 5.2% for females. Rough estimates from these numbers would suggest that also about 25% of control samples might have had an adenoma at time of blood draw. Thus, the question arises which factors make an adenoma further progress to cancer and

whether *S. gallolyticus* might be among these contributing factors. The given data does also not answer this question, however, may stimulate further more mechanistic investigations into the connection of *S. gallolyticus* infection and CRC development.

I performed separate analyses by sex, age at blood draw, country of residence but also case characteristics like age at diagnosis, stage and tumor site. I observed an effect of age on the estimate in EPIC but not in BliTz-DACHSplus. The youngest age-group at blood draw (37 to 55 years) showed a null association with CRC risk as opposed to age groups older than 55 years, where positivity to two or more proteins of the 6-marker panel was significantly associated with CRC. This is reflected in the analyses separated by age at diagnosis, since with a median lag-time of 3.3 years most individuals from the age group 37 to 55 years at blood draw were below 60 years at diagnosis. CRC diagnosed early in life is mostly of different etiology than CRC diagnosed later in life and represents a group of patients at specific risk for CRC [155]. They might represent a specific group of tumors that is not as prone to *S. gallolyticus* invasion as others. Interestingly, in the published Spanish CRC case-control study I observed an increased association with prevalent CRC cases younger than 65 years [81]. Both studies are difficult to compare due to the retrospective and prospective study designs but also the different sets of antigens applied. Nevertheless, the observed discordant age-effect should be regarded with caution and might also originate from cohort effects.

Apart from age, I also observed a difference in the estimate between the distinct participating countries with CRC cases from Italy/Greece having the strongest association as opposed to Germany/Netherlands with a null association. Regional differences in the prevalence of *S. gallolyticus* bacteremia were previously observed and ascribed to cattle farming and fishing areas [77]. The underlying reasons here for an increased number of *S. gallolyticus* positive CRC cases in Italy/Greece remain unclear. I observed neither in BliTz-DACHSplus nor in EPIC an effect of sex or tumor stage on the estimate. However, the estimate differed with tumor site with a stronger association in rectal compared to colon cases. The number of rectal cancer cases was small (n=53) resulting in wide confidence intervals. Whether this difference between colon and rectal cancer cases is true should be assessed in studies with larger sample sizes of rectal cancers.

In a brief exploratory analysis, I wanted to assess whether results from the prospective analyses can be inferred to precancerous lesions. Two independent colonoscopy screening trials (BliTz and the Irish CRC cohort) included colonoscopy-verified adenoma-negative

controls, different types of precancerous lesions as well as CRC cases. Interestingly, associations with antibody responses to *S. gallolyticus* proteins were found already with precancerous lesions, especially non-advanced adenomas and polyps indicating that *S. gallolyticus* infection of colorectal tissue may happen very early. The frequency of positives to two or more proteins of the 6-marker panel among non-advanced adenoma cases (27%) in BliTz and polyp cases (25%) in the Irish cohort thereby resembled the frequency of positives found by Garza-Gonzalez et al. (29%) with a Western blot using *S. gallolyticus* whole cell protein [83]. The Irish cohort additionally included serum samples of individuals with minor diagnoses in colonoscopy, mainly hemorrhoids and diverticulosis. Also in this group the frequency of antibody responses to *S. gallolyticus* was higher than in colonoscopy-negative controls. A study by Reynolds et al. from 1983 found that *S. bovis* (not further specified to subspecies)-induced endocarditis was found in patients with diverticulosis [156]. A potential association of diverticulosis with CRC is controversially discussed in the literature [157], however, it does not seem unlikely that pouches in the intestinal wall increase the invading potential for *S. gallolyticus* .

4.2.3. Antibody responses to *S. gallolyticus* and CRC in a case-control study nested within a southern US prospective cohort

Observed associations with the risk of developing CRC in EPIC could not be reproduced in SCCS. Neither antibody responses to individual *S. gallolyticus* proteins nor to established protein combinations were significantly associated with CRC risk in this study. Major differences between EPIC and SCCS that can be inferred from the given data include the geographic area in which the study was conducted, Europe versus US, the ethnicity of the majority of study participants, mainly Caucasian versus mainly African-American, as well as the mean age at diagnosis, 63 versus 59 years, respectively. Interestingly, the overall CRC incidence differs substantially between the Caucasian-American (CA) and African-American (AA) population in the US, between 1992 and 2006 with an age-adjusted incidence rate of 71.0 in males and 54.8 in females per 100,000 among AA and 61.8 and 45.3 per 100,000 among CA [105]. Age at diagnosis is generally lower in AA compared to CA which is reflected by new recommendations for screening of AA already from age 45 years onwards in contrast to the usually recommended age of 50 years [158, 159]. With respect to an association with *S. gallolyticus* infection the differences between the ethnicities in CRC development might have distinct impacts:

i) Differences in biological mechanisms between CRC in CA and AA are unknown, however, might impact susceptibility for *S. gallolyticus* infection.

ii) Concordant with increased incidences for CRC also the risk for adenoma is increased in AA compared to CA, especially below age 50 [160]. The cut-off that defines 10% of controls as positive for individual *S. gallolyticus* proteins was higher for many of the antigens (Gallo0272, Gallo0933, Gallo2018, Gallo2178, Gallo2179) in the SCCS controls compared to colonoscopy-negative controls in BliTz indicating a higher baseline antibody response in the SCCS controls. A higher rate of *S. gallolyticus* positive adenoma or even CRC already in the SCCS control group, not yet diagnosed, might result in the lowered difference in *S. gallolyticus* positivity to SCCS CRC cases.

iii) Factors potentially underlying the increased risk for CRC in the AA population might affect *S. gallolyticus* colonization of the intestine. Participants of the SCCS included in the CRC nested case-control study for example had a high rate of self-reported obesity-related morbidities like diabetes (27.7% among controls, 35.1% among cases) or hypercholesterol (35.8% among controls, 38.3% among cases) [134]. In contrast, the prevalence of diabetes in Europe had been much lower with 7.9% in 2015 [161]. In addition, rates of *H. pylori* infection, which was itself shown to be associated with CRC in this population differed substantially between the ethnicities in SCCS (AA: 92%; CA: 68%) [134]. *H. pylori* was shown to reshape the gastric microbiota upon infection [162] and also seems to have an influence on the intestinal microbiome [163, 164] Whether obesity-related morbidities or underlying factors in diet and lifestyle as well as co-infections with other bacteria, like *H. pylori*, may affect the association with *S. gallolyticus* needs to be further elucidated.

Analyses of the association of *S. gallolyticus* with CRC separately by ethnicity in the SCCS revealed a higher effect estimate in the CA population than in the AA population, however, due to small sample numbers, confidence intervals were strongly overlapping and this difference in the effect estimate should be regarded with caution and just as indication for above discusses hypotheses for the difference between SCCS and EPIC.

4.3. Natural history of *F. nucleatum* and *S. gallolyticus* infection

High antibody responses to proteins of both, *F. nucleatum* and *S. gallolyticus*, have been rare events in the studies presented here. Both bacteria have in common to be considered as opportunists, i.e. commensals with the potential to turn pathogenic under certain circumstances. However, *F. nucleatum* and *S. gallolyticus* have been identified in different sites and diseases in the human body.

S. gallolyticus is a rare commensal in the intestine: the reported fecal carriage rates for *S. gallolyticus* are low and range from 1 to 11% in healthy individuals [78, 79]. However, species of the *S.bovis*/*S. equinus* complex, including *S. gallolyticus*, were also found to be inducers of infective endocarditis and accounted for 5 to 15% of infective endocarditis patients in the US [165]. It is hypothesized that a colorectal neoplasm in the intestine is the entry port for *S. gallolyticus* to the bloodstream building the connection to a potential systemic infection and pathogenicity of the commensal bacterium in the heart valves [90]. This hypothesis is supported by the findings in this thesis with high antibody responses being a rare event in colonoscopy-negative controls and a significantly larger fraction of individuals with high antibody responses in CRC cases. Risk factors for and the time points of intestinal colonization but also infection and seroconversion have not been reported so far but would be of great importance to further understand the etiology of *S. gallolyticus* in the development of CRC.

In contrast to the apparently rare colonizer *S. gallolyticus*, *F. nucleatum* has been reported to be present in the saliva of 91% of infants already at one year of age [166]. A study of the subgingival periodontal flora in adults found *F. nucleatum* in 58% to 71% of all isolates in individuals with no or mild gingivitis increasing to 82% to 91% in patients with severe gingivitis [35]. Gingivitis is a very frequent disease with a prevalence of more than 50% among adults in the US [167]. Thus, *F. nucleatum* represents a dominant species in the oral microflora and was found to be associated with inflammatory disease at this specific site. I therefore expected to detect strong *F. nucleatum* antibody responses in a substantial portion of study individuals independent of being a CRC case or control, however, as discussed above high antibody responses to the selected *F. nucleatum* proteins were only rarely detected. Other serological studies on *F. nucleatum* conducted in periodontitis patients found lower antibody titers against *F. nucleatum* whole cell protein compared to those against other oral bacteria, although the bacterial load was similar [117, 168].

Commensals are usually tolerated by the mucosal immune system, whereas pathogens like *H. pylori* induce a strong immune response also elicited by high antibody responses [143]. A pathogen-turned commensal would therefore be expected to induce antibody responses, as seen with *S. gallolyticus* in CRC patients. Indeed, as shown by Tew et al., antibody responses to *F. nucleatum* are almost absent in healthy subjects but higher in patients with severe periodontitis, however, still remain lower compared to other bacteria, e.g. *Bacteroides gingivales*, examined in the same study [169]. Unfortunately, attempts to get access to serum samples of periodontitis patients failed but would have been interesting for the serological validation of *F. nucleatum* multiplex serology. Why the overall antibody responses to *F. nucleatum* were comparably low, as shown in this thesis but also by others [117, 168, 169], remains to be elucidated. Some reviews even discuss whether *F. nucleatum* is able to actively inhibit an adaptive immune response [170, 171], however, this is not proven yet.

It is important to further elucidate the natural history, epidemiology and seroconversion pattern of *S. gallolyticus* and *F. nucleatum* to understand differences in the antibody response compared to other bacterial infections, like for example *H. pylori*. The time-point of seroconversion might be of special interest to see when antibody responses are first detectable and whether this is dependent or independent of diseases like periodontitis or intestinal adenoma. One out of two mice inoculated with *S. gallolyticus* bacterial culture showed successful colonization of the intestine and antibody responses of up to 2000 MFI against Gallo2179 already two weeks after inoculation (3.2.2). Factors determining infection success and whether these can be transferred to human infection with *S. gallolyticus* remain to be elucidated. Furthermore, identification of risk factors for carrier state and infection is important to identify individuals at specific risk for the disease. Large cross-sectional studies conducted in children and adults of different age with comprehensive questionnaires would be of high impact in the general understanding of antibody responses to opportunists *F. nucleatum* and *S. gallolyticus.*

4.4. Conclusions and Outlook

In conclusion, antibody responses detected by *F. nucleatum* multiplex serology were not associated with prevalent and incident CRC in the studies presented here. It could be attempted to further improve the assay. An advantage of multiplex serology is the possibility to analyze antibody responses to several antigens in one reaction. However, the number of antigens that can be analyzed is limited to 100 and recombinant expression of potential

antigens with adequate quality controls would not be feasible for very many or the event of all of the approximate 2000 open reading frames of *F. nucleatum*. Screening of protein microarrays representing the complete set of proteins potentially expressed by *F. nucleatum* for sero-reactive proteins could circumvent this problem. A technique developed in our laboratory by Katrin Hufnagel, Smith Lueong and Tim Waterboer uses in vitro transcription and translation of selected genes directly spotted on a Nickel-coated chip. Subsequent incubation of expressed proteins with serum of patients of interest may reveal the identification of immunogenic proteins and disease-specific antibody pattern. This system was already established for *Chlamydia trachomatis* (unpublished data) and is theoretically applicable for all infectious agents with a sequenced genome. However, also for this approach it would be necessary to obtain a set of appropriate reference sera. Reference sera should either be tested with a gold standard serological assay or retrieved from patients with a diagnosed *F. nucleatum* infection or from negative controls without infection, respectively. So far, a gold standard serological assay is not available, however, as described above, other studies used whole cell protein of *F. nucleatum* in ELISA, which could serve as a rather unspecific but potentially sensitive comparison assay. A collection of reference sera, e.g. from patients with severe periodontitis and diagnosed *F. nucleatum* involvement could be an alternative for an infection-based validation. A collaboration with a dentist clinic would have to be established, attempts during my thesis time have failed.

A different technical improvement could arise from applying secondary antibodies for IgA and IgG separately. Application of biotinylated IgA and IgG antibodies in multiplex serology would first have to be established and presented studies would have to be reanalyzed. Apart from technical improvements of the assay, analysis of studies with more information on histology and morphology of the tumor could help assessing whether antibody responses to *F. nucleatum* might serve as specific marker for a subgroup of CRC cases, e.g. tumors of the serrated pathway [49]. If all these attempts fail we would have contradictory results to Wang et al. [147] and could not verify with the developed *F. nucleatum* multiplex serology that serology is applicable to identify specific *F. nucleatum* infection markers for CRC. Antibody responses resulting from *F. nucleatum* infection at its natural site, the gingiva, might lead to a lack of specificity in the detection of CRC specific antibody responses. More site-specific analyses in stool or even tumor tissue as found in the literature is probably advantageous in this respect, however, also less cost- and time-effective than multiplex serology in the analyses of larger epidemiological studies.

In contrast to the absence of a serological association of *F. nucleatum* with CRC, antibody responses to multiple *S. gallolyticus* proteins were significantly associated with CRC. I reproduced and extended earlier findings from Butt et al. [81] by expanding the CRC-specific antigen panel from Gallo2178 and Gallo2179 to a 6-marker panel that was associated with prevalent and incident CRC in two studies conducted in Europe, however, not in a prospective study from southern US with a majority of African-American subjects. The results discussed here raised many new open questions in the association of *S. gallolyticus* with CRC. Positivity to two or more proteins of the 6-marker panel was found in 17% of CRC cases in the prospective EPIC study, two times more than in controls. However, in the southern US study SCCS this association was absent and it remains to be elucidated what the underlying reasons for this difference in the association is. Both studies were conducted on different continents and in distinct ethnicities bearing differences in lifestyle, environment, diet, microbiota and genetics that may have contributed to the differing results. A large cross-sectional study specifically addressing potential risk factors, for example red meat consumption as proposed by Harald zur Hausen [172], could identify risk factors for colonization of the intestine with *S. gallolyticus*. Additionally, it should be analyzed whether *S. gallolyticus* infection is restricted to a certain morphological or histological subtype of CRC.

So far, the data obtained with the prospective study EPIC and with the small sample sets of polyp and adenoma cases from the BliTz and Irish CRC case-control study follow the bacterial-driver passenger model of Harold Tjalsma [33]: Antibody responses to *S. gallolyticus* could be detected in precancerous lesions and in CRC cases prior to diagnosis, showing that infection happens already in precursors. However, whether *S. gallolyticus* is involved in progression of tumor development and thus plays a causal role needs to be further elucidated. The Bradford Hill criteria for causation include: 1) Strength of the association; 2) Consistency; 3) Specificity; 4) Temporality; 5) Biological gradient; 6) Plausibility; 7) Coherence; 8) Experiment; 9) Analogy. Points 1) to 4) are addressed by the data and literature (1.2.2) presented in this thesis. Plausibility and coherence are given by the analogy to *H. pylori*, which is causally related to the development of gastric cancer by the induction of a chronic inflammation. However, a definite causal relation can only be inferred from experimental evidence. Abdulamir et al. found *S. gallolyticus* DNA directly in tumor tissue of CRC patients as a first indicator for a direct involvement of the bacterium in tumorigenesis [80, 82]. In addition they found pro-inflammatory markers elevated in tumor tissue of *S. gallolyticus* positive adenoma and CRC cases [80, 82] indicating a pro-carcinogenic

potential, similar to *H. pylori* in the gastric mucosa [173]. These findings, however, are necessary to be confirmed by independent studies. Additionally, more mechanistic studies should be undertaken. It could be studied in vitro whether infection of cell lines with *S. gallolyticus* leads to altered activation of cellular pathways and which proteins of *S. gallolyticus* might interfere with host metabolics. *H. pylori,* for example, was shown to activate MAPK (Mitogen-activated protein kinase), which has pro-inflammatory and pro-proliferative consequences for the cells [174]. $APC^{min/+}$ mice are a frequently applied in vivo model to study the influence of environmental factors on CRC development [175]. These mice develop, due to a point mutation in the tumor suppressor *APC* gene, multiple intestinal metaplasia. This type of mice was used for studies of the relationship of *F. nucleatum* and CRC and Yang et al. could show that $APC^{min/+}$ mice gavaged with *F. nucleatum* develop more colorectal tumors than mice fed with PBS. Additionally, these mice had shorter survival times. This mouse model could also be applied to study whether colonization with *S. gallolyticus* promotes tumor progression. The finding of a causal relationship could be valuable in prevention of the disease by eradication of the bacterium. However, CRC development is considered a multifactorial process [176] and the impact of eradication of one specific bacterium remains to be elucidated.

Independently of a causal relationship, antibody responses to proteins in *S. gallolyticus* multiplex serology, in contrast to *F. nucleatum* multiplex serology, might serve as an early marker for the development of CRC in European populations and could therefore be of diagnostic value for a subgroup of CRC cases. The identified 6-marker panel should be verified in additional independent prospective studies and analyzed for the diagnostic potential in larger studies of colorectal precancerous lesions.

5. References

1 http://globocan.iarc.fr/, 2016,

2 http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-028312.pdf, 2016,

3 Liang PS, Chen TY, Giovannucci E: Cigarette smoking and colorectal cancer incidence and mortality: Systematic review and meta-analysis. International journal of cancer Journal international du cancer 2009;124:2406-2415.

4 Fedirko V, Tramacere I, Bagnardi V, Rota M, Scotti L, Islami F, Negri E, Straif K, Romieu I, La Vecchia C, Boffetta P, Jenab M: Alcohol drinking and colorectal cancer risk: An overall and dose-response meta-analysis of published studies. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 2011;22:1958-1972.

5 Chan DS, Lau R, Aune D, Vieira R, Greenwood DC, Kampman E, Norat T: Red and processed meat and colorectal cancer incidence: Meta-analysis of prospective studies. PloS one 2011;6:e20456.

6 Samad AK, Taylor RS, Marshall T, Chapman MA: A meta-analysis of the association of physical activity with reduced risk of colorectal cancer. Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland 2005;7:204-213.

7 Jiang Y, Ben Q, Shen H, Lu W, Zhang Y, Zhu J: Diabetes mellitus and incidence and mortality of colorectal cancer: A systematic review and meta-analysis of cohort studies. Eur J Epidemiol 2011;26:863-876.

8 Ma Y, Yang Y, Wang F, Zhang P, Shi C, Zou Y, Qin H: Obesity and risk of colorectal cancer: A systematic review of prospective studies. PloS one 2013;8:e53916.

9 Jess T, Rungoe C, Peyrin-Biroulet L: Risk of colorectal cancer in patients with ulcerative colitis: A meta-analysis of population-based cohort studies. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2012;10:639-645.

10 Algra AM, Rothwell PM: Effects of regular aspirin on long-term cancer incidence and metastasis: A systematic comparison of evidence from observational studies versus randomised trials. The lancet oncology 2012;13:518-527.

11 Taylor DP, Burt RW, Williams MS, Haug PJ, Cannon-Albright LA: Population-based family history-specific risks for colorectal cancer: A constellation approach. Gastroenterology 2010;138:877-885.

12 Lynch HT, de la Chapelle A: Hereditary colorectal cancer. The New England journal of medicine 2003;348:919-932.

13 Galiatsatos P, Foulkes WD: Familial adenomatous polyposis. Am J Gastroenterol 2006;101:385-398.

14 Conteduca V, Sansonno D, Russi S, Dammacco F: Precancerous colorectal lesions (review). International journal of oncology 2013;43:973-984.

15 Simon K: Colorectal cancer development and advances in screening. Clinical interventions in aging 2016;11:967-976.

16 Pino MS, Chung DC: The chromosomal instability pathway in colon cancer. Gastroenterology 2010;138:2059-2072.

17 Sweetser S, Smyrk TC, Sinicrope FA: Serrated colon polyps as precursors to colorectal cancer. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association 2013;11:760-767; quiz e754-765.

18 Brenner H, Stock C, Hoffmeister M: Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: Systematic review and meta-analysis of randomised controlled trials and observational studies. Bmj 2014;348:g2467.

19 Zavoral M, Suchanek S, Zavada F, Dusek L, Muzik J, Seifert B, Fric P: Colorectal cancer screening in europe. World journal of gastroenterology : WJG 2009;15:5907-5915.

20 Brenner H, Tao S: Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. European journal of cancer 2013;49:3049-3054.

21 Niedermaier T, Weigl K, Hoffmeister M, Brenner H: Fecal immunochemical tests combined with other stool tests for colorectal cancer and advanced adenoma detection: A systematic review. Clinical and translational gastroenterology 2016;7:e175.

22 Brenner H, Hoffmeister M, Birkner B, Stock C: Which adenomas are detected by fecal occult blood testing? A state-wide analysis from bavaria, germany. International journal of cancer Journal international du cancer 2015;136:1672-1679.

23 Nguyen MT, Weinberg DS: Biomarkers in colorectal cancer screening. Journal of the National Comprehensive Cancer Network : JNCCN 2016;14:1033-1040.

24 Baxter NT, Ruffin MTt, Rogers MA, Schloss PD: Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome medicine 2016;8:37.

25 Marchesi JR, Adams DH, Fava F, Hermes GD, Hirschfield GM, Hold G, Quraishi MN, Kinross J, Smidt H, Tuohy KM, Thomas LV, Zoetendal EG, Hart A: The gut microbiota and host health: A new clinical frontier. Gut 2015

26 Rodriguez JM, Murphy K, Stanton C, Ross RP, Kober OI, Juge N, Avershina E, Rudi K, Narbad A, Jenmalm MC, Marchesi JR, Collado MC: The composition of the gut microbiota throughout life, with an emphasis on early life. Microbial ecology in health and disease 2015;26:26050.

27 Flint HJ, Scott KP, Louis P, Duncan SH: The role of the gut microbiota in nutrition and health. Nature reviews Gastroenterology & hepatology 2012;9:577-589.

28 Kelly D, Mulder IE: Microbiome and immunological interactions. Nutr Rev 2012;70 Suppl 1:S18-30.

29 Gagniere J, Raisch J, Veziant J, Barnich N, Bonnet R, Buc E, Bringer MA, Pezet D, Bonnet M: Gut microbiota imbalance and colorectal cancer. World journal of gastroenterology : WJG 2016;22:501-518.

30 Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Bohm J, Brunetti F, Habermann N, Hercog R, Koch M, Luciani A, Mende DR, Schneider MA, Schrotz-King P, Tournigand C, Tran Van Nhieu J, Yamada T, Zimmermann J, Benes V, Kloor M, Ulrich CM, von Knebel Doeberitz M, Sobhani I, Bork P: Potential of fecal

microbiota for early-stage detection of colorectal cancer. Molecular systems biology 2014;10:766.

31 Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, O'Riordain M, Shanahan F, O'Toole PW: Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut 2016

32 Sears CL, Pardoll DM: Perspective: Alpha-bugs, their microbial partners, and the link to colon cancer. J Infect Dis 2011;203:306-311.

33 Tjalsma H, Boleij A, Marchesi JR, Dutilh BE: A bacterial driver-passenger model for colorectal cancer: Beyond the usual suspects. Nature reviews Microbiology 2012;10:575-582.

34 Bolstad AI, Jensen HB, Bakken V: Taxonomy, biology, and periodontal aspects of fusobacterium nucleatum. Clinical microbiology reviews 1996;9:55-71.

35 Moore WE, Moore LV: The bacteria of periodontal diseases. Periodontology 2000 1994;5:66-77.

36 Han YW, Wang X: Mobile microbiome: Oral bacteria in extra-oral infections and inflammation. Journal of dental research 2013;92:485-491.

37 Fukugaiti MH, Ignacio A, Fernandes MR, Ribeiro Junior U, Nakano V, Avila-Campos MJ: High occurrence of fusobacterium nucleatum and clostridium difficile in the intestinal microbiota of colorectal carcinoma patients. Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology] 2015;46:1135-1140.

38 Liang JQ, Chiu J, Chen Y, Huang Y, Higashimori A, Fang JY, Brim H, Ashktorab H, Ng SC, Ng SS, Zheng S, Chan FK, Sung JJ, Yu J: Fecal bacteria act as novel biomarkers for non-invasive diagnosis of colorectal cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 2016

39 McCoy AN, Araujo-Perez F, Azcarate-Peril A, Yeh JJ, Sandler RS, Keku TO: Fusobacterium is associated with colorectal adenomas. PloS one 2013;8:e53653.

40 Mira-Pascual L, Cabrera-Rubio R, Ocon S, Costales P, Parra A, Suarez A, Moris F, Rodrigo L, Mira A, Collado MC: Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. J Gastroenterol 2015;50:167-179.

41 Wong SH, Kwong TN, Chow TC, Luk AK, Dai RZ, Nakatsu G, Lam TY, Zhang L, Wu JC, Chan FK, Ng SS, Wong MC, Ng SC, Wu WK, Yu J, Sung JJ: Quantitation of faecal fusobacterium improves faecal immunochemical test in detecting advanced colorectal neoplasia. Gut 2016

42 Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, Wang X, Xu X, Chen N, Wu WK, Al-Aama J, Nielsen HJ, Kiilerich P, Jensen BA, Yau TO, Lan Z, Jia H, Li J, Xiao L, Lam TY, Ng SC, Cheng AS, Wong VW, Chan FK, Xu X, Yang H, Madsen L, Datz C, Tilg H, Wang J, Brunner N, Kristiansen K, Arumugam M, Sung JJ, Wang J: Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut 2015

43 Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, Allen-Vercoe E, Moore RA, Holt RA: Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome research 2012;22:299-306.

44 Flanagan L, Schmid J, Ebert M, Soucek P, Kunicka T, Liska V, Bruha J, Neary P, Dezeeuw N, Tommasino M, Jenab M, Prehn JH, Hughes DJ: Fusobacterium nucleatum

associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology 2014;33:1381-1390.

45 Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J, Baselga J, Liu C, Shivdasani RA, Ogino S, Birren BW, Huttenhower C, Garrett WS, Meyerson M: Genomic analysis identifies association of fusobacterium with colorectal carcinoma. Genome research 2012;22:292-298.

46 Li YY, Ge QX, Cao J, Zhou YJ, Du YL, Shen B, Wan YJ, Nie YQ: Association of fusobacterium nucleatum infection with colorectal cancer in chinese patients. World journal of gastroenterology : WJG 2016;22:3227-3233.

47 Mima K, Nishihara R, Qian ZR, Cao Y, Sukawa Y, Nowak JA, Yang J, Dou R, Masugi Y, Song M, Kostic AD, Giannakis M, Bullman S, Milner DA, Baba H, Giovannucci EL, Garraway LA, Freeman GJ, Dranoff G, Garrett WS, Huttenhower C, Meyerson M, Meyerhardt JA, Chan AT, Fuchs CS, Ogino S: Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. Gut 2015

48 Wei Z, Cao S, Liu S, Yao Z, Sun T, Li Y, Li J, Zhang D, Zhou Y: Could gut microbiota serve as prognostic biomarker associated with colorectal cancer patients' survival? A pilot study on relevant mechanism. Oncotarget 2016

49 Ito M, Kanno S, Nosho K, Sukawa Y, Mitsuhashi K, Kurihara H, Igarashi H, Takahashi T, Tachibana M, Takahashi H, Yoshii S, Takenouchi T, Hasegawa T, Okita K, Hirata K, Maruyama R, Suzuki H, Imai K, Yamamoto H, Shinomura Y: Association of fusobacterium nucleatum with clinical and molecular features in colorectal serrated pathway. International journal of cancer Journal international du cancer 2015;137:1258-1268.

50 Nosho K, Sukawa Y, Adachi Y, Ito M, Mitsuhashi K, Kurihara H, Kanno S, Yamamoto I, Ishigami K, Igarashi H, Maruyama R, Imai K, Yamamoto H, Shinomura Y: Association of fusobacterium nucleatum with immunity and molecular alterations in colorectal cancer. World journal of gastroenterology : WJG 2016;22:557-566.

51 Mima K, Sukawa Y, Nishihara R, Qian ZR, Yamauchi M, Inamura K, Kim SA, Masuda A, Nowak JA, Nosho K, Kostic AD, Giannakis M, Watanabe H, Bullman S, Milner DA, Harris CC, Giovannucci E, Garraway LA, Freeman GJ, Dranoff G, Chan AT, Garrett WS, Huttenhower C, Fuchs CS, Ogino S: Fusobacterium nucleatum and t cells in colorectal carcinoma. JAMA oncology 2015;1:653-661.

52 Kaplan CW, Ma X, Paranjpe A, Jewett A, Lux R, Kinder-Haake S, Shi W: Fusobacterium nucleatum outer membrane proteins fap2 and radd induce cell death in human lymphocytes. Infection and immunity 2010;78:4773-4778.

53 Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, Enk J, Bar-On Y, Stanietsky-Kaynan N, Coppenhagen-Glazer S, Shussman N, Almogy G, Cuapio A, Hofer E, Mevorach D, Tabib A, Ortenberg R, Markel G, Miklic K, Jonjic S, Brennan CA, Garrett WS, Bachrach G, Mandelboim O: Binding of the fap2 protein of fusobacterium nucleatum to human inhibitory receptor tigit protects tumors from immune cell attack. Immunity 2015;42:344-355.

54 Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL, El-Omar EM, Brenner D, Fuchs CS, Meyerson M, Garrett WS: Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell host & microbe 2013;14:207-215.

55 Yang Y, Weng W, Peng J, Hong L, Yang L, Toiyama Y, Gao R, Liu M, Yin M, Pan C, Li H, Guo B, Zhu Q, Wei Q, Moyer MP, Wang P, Cai S, Goel A, Qin H, Ma Y: Fusobacterium nucleatum increases proliferation of colorectal cancer cells and tumor development in mice by activating tlr4 signaling to nfkappab, upregulating expression of microrna-21. Gastroenterology 2016

56 Han YW, Ikegami A, Rajanna C, Kawsar HI, Zhou Y, Li M, Sojar HT, Genco RJ, Kuramitsu HK, Deng CX: Identification and characterization of a novel adhesin unique to oral fusobacteria. Journal of bacteriology 2005;187:5330-5340.

57 Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW: Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating e-cadherin/beta-catenin signaling via its fada adhesin. Cell host & microbe 2013;14:195-206.

58 Desvaux M, Khan A, Beatson SA, Scott-Tucker A, Henderson IR: Protein secretion systems in fusobacterium nucleatum: Genomic identification of type 4 piliation and complete type v pathways brings new insight into mechanisms of pathogenesis. Biochimica et biophysica acta 2005;1713:92-112.

59 Abed J, Emgard JE, Zamir G, Faroja M, Almogy G, Grenov A, Sol A, Naor R, Pikarsky E, Atlan KA, Mellul A, Chaushu S, Manson AL, Earl AM, Ou N, Brennan CA, Garrett WS, Bachrach G: Fap2 mediates fusobacterium nucleatum colorectal adenocarcinoma enrichment by binding to tumor-expressed gal-galnac. Cell host & microbe 2016;20:215-225.

60 Kaplan CW, Lux R, Haake SK, Shi W: The fusobacterium nucleatum outer membrane protein radd is an arginine-inhibitable adhesin required for inter-species adherence and the structured architecture of multispecies biofilm. Mol Microbiol 2009;71:35-47.

61 Devriese LA, Uyttebroek E, Gevaert D, Vandekerckhove P, Ceyssens K: Streptococcus bovis infections in pigeons. Avian pathology : journal of the WVPA 1990;19:429-434.

62 Devriese LA, Vandamme P, Pot B, Vanrobaeys M, Kersters K, Haesebrouck F: Differentiation between streptococcus gallolyticus strains of human clinical and veterinary origins and streptococcus bovis strains from the intestinal tracts of ruminants. J Clin Microbiol 1998;36:3520-3523.

63 Jans C, de Wouters T, Bonfoh B, Lacroix C, Kaindi DW, Anderegg J, Bock D, Vitali S, Schmid T, Isenring J, Kurt F, Kogi-Makau W, Meile L: Phylogenetic, epidemiological and functional analyses of the streptococcus bovis/streptococcus equinus complex through an overarching mlst scheme. BMC microbiology 2016;16:117.

64 Sekizaki T, Nishiya H, Nakajima S, Nishizono M, Kawano M, Okura M, Takamatsu D, Nishino H, Ishiji T, Osawa R: Endocarditis in chickens caused by subclinical infection of streptococcus gallolyticus subsp. Gallolyticus. Avian diseases 2008;52:183-186.

65 Mevius D, Devriese L, Butaye P, Vandamme P, Verschure M, Veldman K: Isolation of glycopeptide resistant streptococcus gallolyticus strains with vana, vanb, and both vana and vanb genotypes from faecal samples of veal calves in the netherlands. The Journal of antimicrobial chemotherapy 1998;42:275-276.

66 Schlegel L, Grimont F, Ageron E, Grimont PA, Bouvet A: Reappraisal of the taxonomy of the streptococcus bovis/streptococcus equinus complex and related species: Description of streptococcus gallolyticus subsp. Gallolyticus subsp. Nov., s. Gallolyticus subsp. Macedonicus subsp. Nov. And s. Gallolyticus subsp. Pasteurianus subsp. Nov. Int J Syst Evol Microbiol 2003;53:631-645.

67 Osawa R, Fujisawa T, Sly LI: Streptococcus gallolyticus sp-nov - gallate degrading organisms formerly assigned to streptococcus-bovis. Systematic and applied microbiology 1995;18:74-78.

68 Klein RS, Recco RA, Catalano MT, Edberg SC, Casey JI, Steigbigel NH: Association of streptococcus bovis with carcinoma of the colon. The New England journal of medicine 1977;297:800-802.

69 Murray HW, Roberts RB: Streptococcus bovis bacteremia and underlying gastrointestinal disease. Archives of internal medicine 1978;138:1097-1099.

70 Noble CJ, Uttley AH, Falk RH, Richardson PJ: Streptococcus bovis endocarditis and colonic cancer. Lancet 1978;1:766.

71 Steinberg D, Naggar CZ: Streptococcus bovis endocarditis with carcinoma of the colon. The New England journal of medicine 1977;297:1354-1355.

72 Ruoff KL, Miller SI, Garner CV, Ferraro MJ, Calderwood SB: Bacteremia with streptococcus bovis and streptococcus salivarius: Clinical correlates of more accurate identification of isolates. J Clin Microbiol 1989;27:305-308.

73 Corredoira JC, Alonso MP, Garcia JF, Casariego E, Coira A, Rodriguez A, Pita J, Louzao C, Pombo B, Lopez MJ, Varela J: Clinical characteristics and significance of streptococcus salivarius bacteremia and streptococcus bovis bacteremia: A prospective 16-year study. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology 2005;24:250-255.

74 Corredoira-Sanchez J, Garcia-Garrote F, Rabunal R, Lopez-Roses L, Garcia-Pais MJ, Castro E, Gonzalez-Soler R, Coira A, Pita J, Lopez-Alvarez MJ, Alonso MP, Varela J: Association between bacteremia due to streptococcus gallolyticus subsp. Gallolyticus (streptococcus bovis i) and colorectal neoplasia: A case-control study. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 2012;55:491-496.

75 Corredoira JC, Alonso MP, Garcia-Pais MJ, Rabunal R, Garcia-Garrote F, Lopez-Roses L, Lancho A, Coira A, Pita J, Velasco D, Lopez-Alvarez MJ, Tjalsma H, Varela J: Is colonoscopy necessary in cases of infection by streptococcus bovis biotype ii? European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology 2014;33:171-177.

76 Sheng WH, Chuang YC, Teng LJ, Hsueh PR: Bacteraemia due to streptococcus gallolyticus subspecies pasteurianus is associated with digestive tract malignancies and resistance to macrolides and clindamycin. The Journal of infection 2014;69:145-153.

77 Corredoira J, Grau I, Garcia-Rodriguez JF, Alonso-Garcia P, Garcia-Pais MJ, Rabunal R, Garcia-Garrote F, Ardanuy C, Coira A, Lopez-Alvarez MJ, Pallares R: The clinical epidemiology and malignancies associated with streptococcus bovis biotypes in 506 cases of bloodstream infections. The Journal of infection 2015;71:317-325.

78 Chirouze C, Patry I, Duval X, Baty V, Tattevin P, Aparicio T, Pagenault M, Carbonnel F, Couetdic G, Hoen B: Streptococcus bovis/streptococcus equinus complex fecal carriage, colorectal carcinoma, and infective endocarditis: A new appraisal of a complex connection. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology 2013;32:1171-1176.

79 Lopes PG, Cantarelli VV, Agnes G, Costabeber AM, d'Azevedo PA: Novel real-time pcr assays using taqman minor groove binder probes for identification of fecal carriage of

streptococcus bovis/streptococcus equinus complex from rectal swab specimens. J Clin Microbiol 2014;52:974-976.

80 Abdulamir AS, Hafidh RR, Bakar FA: Molecular detection, quantification, and isolation of streptococcus gallolyticus bacteria colonizing colorectal tumors: Inflammation-driven potential of carcinogenesis via il-1, cox-2, and il-8. Molecular cancer 2010;9:249.

81 Butt J, Romero-Hernandez B, Perez-Gomez B, Willhauck-Fleckenstein M, Holzinger D, Martin V, Moreno V, Linares C, Dierssen-Sotos T, Barricarte A, Tardon A, Altzibar JM, Moreno-Osset E, Franco F, Requena RO, Huerta JM, Michel A, Waterboer T, Castano-Vinyals G, Kogevinas M, Pollan M, Boleij A, de Sanjose S, Del Campo R, Tjalsma H, Aragones N, Pawlita M: Association of streptococcus gallolyticus subspecies gallolyticus with colorectal cancer: Serological evidence. International journal of cancer Journal international du cancer 2016;138:1670-1679.

82 Abdulamir AS, Hafidh RR, Mahdi LK, Al-jeboori T, Abubaker F: Investigation into the controversial association of streptococcus gallolyticus with colorectal cancer and adenoma. BMC cancer 2009;9:403.

83 Garza-Gonzalez E, Rios M, Bosques-Padilla FJ, Francois F, Cho I, Gonzalez GM, Perez-Perez GI: Immune response against streptococcus gallolyticus in patients with adenomatous polyps in colon. International journal of cancer Journal international du cancer 2012;131:2294-2299.

84 Boleij A, Roelofs R, Schaeps RM, Schulin T, Glaser P, Swinkels DW, Kato I, Tjalsma H: Increased exposure to bacterial antigen rpl7/l12 in early stage colorectal cancer patients. Cancer 2010;116:4014-4022.

85 Sillanpaa J, Nallapareddy SR, Singh KV, Ferraro MJ, Murray BE: Adherence characteristics of endocarditis-derived streptococcus gallolyticus ssp. Gallolyticus (streptococcus bovis biotype i) isolates to host extracellular matrix proteins. FEMS microbiology letters 2008;289:104-109.

86 Vollmer T, Hinse D, Kleesiek K, Dreier J: Interactions between endocarditis-derived streptococcus gallolyticus subsp. Gallolyticus isolates and human endothelial cells. BMC microbiology 2010;10:78.

87 Cole WG, Chan D, Hickey AJ, Wilcken DE: Collagen composition of normal and myxomatous human mitral heart valves. The Biochemical journal 1984;219:451-460.

88 Zou X, Feng B, Dong T, Yan G, Tan B, Shen H, Huang A, Zhang X, Zhang M, Yang P, Zheng M, Zhang Y: Up-regulation of type i collagen during tumorigenesis of colorectal cancer revealed by quantitative proteomic analysis. Journal of proteomics 2013;94:473-485.

89 Danne C, Entenza JM, Mallet A, Briandet R, Debarbouille M, Nato F, Glaser P, Jouvion G, Moreillon P, Trieu-Cuot P, Dramsi S: Molecular characterization of a streptococcus gallolyticus genomic island encoding a pilus involved in endocarditis. J Infect Dis 2011;204:1960-1970.

90 Boleij A, Muytjens CM, Bukhari SI, Cayet N, Glaser P, Hermans PW, Swinkels DW, Bolhuis A, Tjalsma H: Novel clues on the specific association of streptococcus gallolyticus subsp gallolyticus with colorectal cancer. J Infect Dis 2011;203:1101-1109.

91 Martins M, Aymeric L, du Merle L, Danne C, Robbe-Masselot C, Trieu-Cuot P, Sansonetti P, Dramsi S: Streptococcus gallolyticus pil3 pilus is required for adhesion to colonic mucus and for colonization of mouse distal colon. J Infect Dis 2015;212:1646-1655.

92 Martins M, Porrini C, du Merle L, Danne C, Robbe-Masselot C, Trieu-Cuot P, Dramsi S: The pil3 pilus of streptococcus gallolyticus binds to intestinal mucins and to fibrinogen. Gut microbes 2016;7:526-532.

93 Sanchez-Diaz AM, Romero-Hernandez B, Conde-Moreno E, Kwak YK, Zamora J, Colque-Navarro P, Mollby R, Ruiz-Garbajosa P, Canton R, Garcia-Bermejo L, Del Campo R: New insights into the enterococcus faecium and streptococcus gallolyticus subsp. Gallolyticus host interaction mechanisms. PloS one 2016;11:e0159159.

94 Boleij A, Dutilh BE, Kortman GA, Roelofs R, Laarakkers CM, Engelke UF, Tjalsma H: Bacterial responses to a simulated colon tumor microenvironment. Molecular & cellular proteomics : MCP 2012;11:851-862.

95 Porta M: A dictionary of epidemiology, Oxford University Press, 2014, pp 95.

96 Rothman KJ: Epidemiology: An introduction, Oxford University Press, 2012, pp 87-94.

97 Rothman KJ: Epidemiology: An introduction, Oxford University Press, 2012, pp 69-85.

98 Hill AB: The environment and disease: Association or causation? Proceedings of the Royal Society of Medicine 1965;58:295-300.

99 Rothman KJ: Epidemiology: An introduction, Oxford University Press, 2012, pp 88-90.

100 Rothman KJ: Epidemiology: An introduction, Oxford University Press, 2012, pp 211-216.

101 Rothman KJ: Modern epidemiology, Lippincott Williams&Wilkins, 2013, pp 157.

102 Rothman KJ: Modern epidemiology, Lippincott Williams&Wilkins, 2013, pp 260.

103 Rothman KJ: Modern epidemiology, Lippincott Williams&Wilkins, 2013, pp 171-178.

104 Rothman KJ: Modern epidemiology, Lippincott Williams&Wilkins, 2013, pp 259.

105 Murphy G, Devesa SS, Cross AJ, Inskip PD, McGlynn KA, Cook MB: Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. International journal of cancer Journal international du cancer 2011;128:1668-1675.

106 Washington JA: Principles of diagnosis; Medical microbiology, University of Texas Medical Branch at Galveston, 1996

107 Saito I, Miyamura T, Ohbayashi A, Harada H, Katayama T, Kikuchi S, Watanabe Y, Koi S, Onji M, Ohta Y, et al.: Hepatitis c virus infection is associated with the development of hepatocellular carcinoma. Proceedings of the National Academy of Sciences of the United States of America 1990;87:6547-6549.

108 Forman D, Newell DG, Fullerton F, Yarnell JW, Stacey AR, Wald N, Sitas F: Association between infection with helicobacter pylori and risk of gastric cancer: Evidence from a prospective investigation. Bmj 1991;302:1302-1305.

109 Tsang RK, Vlantis AC, Ho RW, Tam JS, To KF, van Hasselt CA: Sensitivity and specificity of epstein-barr virus iga titer in the diagnosis of nasopharyngeal carcinoma: A three-year institutional review. Head & neck 2004;26:598-602.

110 Meschede W, Zumbach K, Braspenning J, Scheffner M, Benitez-Bribiesca L, Luande J, Gissmann L, Pawlita M: Antibodies against early proteins of human papillomaviruses as diagnostic markers for invasive cervical cancer. J Clin Microbiol 1998;36:475-480.

111 Kreimer AR, Brennan P, Lang Kuhs KA, Waterboer T, Clifford G, Franceschi S, Michel A, Willhauck-Fleckenstein M, Riboli E, Castellsague X, Hildesheim A, Fortner RT, Kaaks R, Palli D, Ljuslinder I, Panico S, Clavel-Chapelon F, Boutron-Ruault MC, Mesrine S, Trichopoulou A, Lagiou P, Trichopoulos D, Peeters PH, Cross AJ, Bueno-de-Mesquita HB, Vineis P, Larranaga N, Pala V, Sanchez MJ, Navarro C, Barricarte A, Tumino R, Khaw KT, Wareham N, Boeing H, Steffen A, Travis RC, Quiros JR, Weiderpass E, Pawlita M, Johansson M: Human papillomavirus antibodies and future risk of anogenital cancer: A nested case-control study in the european prospective investigation into cancer and nutrition study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2015;33:877-884.

112 Paulson KG, Carter JJ, Johnson LG, Cahill KW, Iyer JG, Schrama D, Becker JC, Madeleine MM, Nghiem P, Galloway DA: Antibodies to merkel cell polyomavirus t antigen oncoproteins reflect tumor burden in merkel cell carcinoma patients. Cancer research 2010;70:8388-8397.

113 Gao L, Michel A, Weck MN, Arndt V, Pawlita M, Brenner H: Helicobacter pylori infection and gastric cancer risk: Evaluation of 15 h. Pylori proteins determined by novel multiplex serology. Cancer research 2009;69:6164-6170.

114 Waterboer T, Sehr P, Michael KM, Franceschi S, Nieland JD, Joos TO, Templin MF, Pawlita M: Multiplex human papillomavirus serology based on in situ-purified glutathione s-transferase fusion proteins. Clinical chemistry 2005;51:1845-1853.

115 MacArthur H, Walter G: Monoclonal antibodies specific for the carboxy terminus of simian virus 40 large t antigen. Journal of virology 1984;52:483-491.

116 Boleij A, Roelofs R, Danne C, Bellais S, Dramsi S, Kato I, Tjalsma H: Selective antibody response to streptococcus gallolyticus pilus proteins in colorectal cancer patients. Cancer Prev Res (Phila) 2012;5:260-265.

117 Shin J, Kho SA, Choi YS, Kim YC, Rhyu IC, Choi Y: Antibody and t cell responses to fusobacterium nucleatum and treponema denticola in health and chronic periodontitis. PloS one 2013;8:e53703.

118 Nakagaki H, Sekine S, Terao Y, Toe M, Tanaka M, Ito HO, Kawabata S, Shizukuishi S, Fujihashi K, Kataoka K: Fusobacterium nucleatum envelope protein foma is immunogenic and binds to the salivary statherin-derived peptide. Infection and immunity 2010;78:1185-1192.

119 Rusniok C, Couve E, Da Cunha V, El Gana R, Zidane N, Bouchier C, Poyart C, Leclercq R, Trieu-Cuot P, Glaser P: Genome sequence of streptococcus gallolyticus: Insights into its adaptation to the bovine rumen and its ability to cause endocarditis. Journal of bacteriology 2010;192:2266-2276.

120 Kapatral V, Anderson I, Ivanova N, Reznik G, Los T, Lykidis A, Bhattacharyya A, Bartman A, Gardner W, Grechkin G, Zhu L, Vasieva O, Chu L, Kogan Y, Chaga O, Goltsman E, Bernal A, Larsen N, D'Souza M, Walunas T, Pusch G, Haselkorn R, Fonstein M, Kyrpides N, Overbeek R: Genome sequence and analysis of the oral bacterium fusobacterium nucleatum strain atcc 25586. Journal of bacteriology 2002;184:2005-2018.

121 Hinse D, Vollmer T, Ruckert C, Blom J, Kalinowski J, Knabbe C, Dreier J: Complete genome and comparative analysis of streptococcus gallolyticus subsp. Gallolyticus, an emerging pathogen of infective endocarditis. BMC genomics 2011;12:400.

122 Kapatral V, Ivanova N, Anderson I, Reznik G, Bhattacharyya A, Gardner WL, Mikhailova N, Lapidus A, Larsen N, D'Souza M, Walunas T, Haselkorn R, Overbeek R, Kyrpides N: Genome analysis of f. Nucleatum sub spp vincentii and its comparison with the genome of f. Nucleatum atcc 25586. Genome research 2003;13:1180-1189.

123 Sillanpaa J, Nallapareddy SR, Qin X, Singh KV, Muzny DM, Kovar CL, Nazareth LV, Gibbs RA, Ferraro MJ, Steckelberg JM, Weinstock GM, Murray BE: A collagen-binding adhesin, acb, and ten other putative mscramm and pilus family proteins of streptococcus gallolyticus subsp. Gallolyticus (streptococcus bovis group, biotype i). Journal of bacteriology 2009;191:6643-6653.

124 Cho E, Park SN, Lim YK, Shin Y, Paek J, Hwang CH, Chang YH, Kook JK: Fusobacterium hwasookii sp. Nov., isolated from a human periodontitis lesion. Current microbiology 2015;70:169-175.

125 Https://www.Ncbi.Nlm.Nih.Gov/nuccore/cdey01000010,

126 Sehr P, Zumbach K, Pawlita M: A generic capture elisa for recombinant proteins fused to glutathione s-transferase: Validation for hpv serology. Journal of immunological methods 2001;253:153-162.

127 Waterboer T, Sehr P, Pawlita M: Suppression of non-specific binding in serological luminex assays. Journal of immunological methods 2006;309:200-204.

128 Antonsson A, Green AC, Mallitt KA, O'Rourke PK, Pandeya N, Pawlita M, Waterboer T, Neale RE: Prevalence and stability of antibodies to 37 human papillomavirus types--a population-based longitudinal study. Virology 2010;407:26-32.

129 Chen H, Werner S, Butt J, Zornig I, Knebel P, Michel A, Eichmuller SB, Jager D, Waterboer T, Pawlita M, Brenner H: Prospective evaluation of 64 serum autoantibodies as biomarkers for early detection of colorectal cancer in a true screening setting. Oncotarget 2016

130 Hundt S, Haug U, Brenner H: Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. Ann Intern Med 2009;150:162-169.

131 Brenner H, Tao S, Haug U: Low-dose aspirin use and performance of immunochemical fecal occult blood tests. JAMA : the journal of the American Medical Association 2010;304:2513-2520.

132 Brenner H, Chang-Claude J, Seiler CM, Rickert A, Hoffmeister M: Protection from colorectal cancer after colonoscopy: A population-based, case-control study. Ann Intern Med 2011;154:22-30.

133 Werner S, Chen H, Butt J, Michel A, Knebel P, Holleczek B, Zornig I, Eichmuller SB, Jager D, Pawlita M, Waterboer T, Brenner H: Evaluation of the diagnostic value of 64 simultaneously measured autoantibodies for early detection of gastric cancer. Scientific reports 2016;6:25467.

134 Epplein M, Pawlita M, Michel A, Peek RM, Jr., Cai Q, Blot WJ: Helicobacter pylori protein-specific antibodies and risk of colorectal cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2013;22:1964-1974.

135 Signorello LB, Hargreaves MK, Steinwandel MD, Zheng W, Cai Q, Schlundt DG, Buchowski MS, Arnold CW, McLaughlin JK, Blot WJ: Southern community cohort study: Establishing a cohort to investigate health disparities. Journal of the National Medical Association 2005;97:972-979.

136 Young JL Jr RS, Ries LAG, Fritz AG, Hurlbut AA: Seer summary staging manual - 2000: Codes and coding instructions. Bethesda, MD, 2001, National Cancer Institute NIH

137 Http://epic.Iarc.Fr/, 2016,

138 Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondiere UR, Hemon B, Casagrande C, Vignat J, Overvad K, Tjonneland A, Clavel-Chapelon F, Thiebaut A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D, Bueno-De-Mesquita HB, Peeters PH, Lund E, Engeset D, Gonzalez CA, Barricarte A, Berglund G, Hallmans G, Day NE, Key TJ, Kaaks R, Saracci R: European prospective investigation into cancer and nutrition (epic): Study populations and data collection. Public health nutrition 2002;5:1113-1124.

139 Jenab M, Bueno-de-Mesquita HB, Ferrari P, van Duijnhoven FJ, Norat T, Pischon T, Jansen EH, Slimani N, Byrnes G, Rinaldi S, Tjonneland A, Olsen A, Overvad K, Boutron-Ruault MC, Clavel-Chapelon F, Morois S, Kaaks R, Linseisen J, Boeing H, Bergmann MM, Trichopoulou A, Misirli G, Trichopoulos D, Berrino F, Vineis P, Panico S, Palli D, Tumino R, Ros MM, van Gils CH, Peeters PH, Brustad M, Lund E, Tormo MJ, Ardanaz E, Rodriguez L, Sanchez MJ, Dorronsoro M, Gonzalez CA, Hallmans G, Palmqvist R, Roddam A, Key TJ, Khaw KT, Autier P, Hainaut P, Riboli E: Association between pre-diagnostic circulating vitamin d concentration and risk of colorectal cancer in european populations:A nested case-control study. Bmj 2010;340:b5500.

140 Fedirko V, Riboli E, Tjonneland A, Ferrari P, Olsen A, Bueno-de-Mesquita HB, van Duijnhoven FJ, Norat T, Jansen EH, Dahm CC, Overvad K, Boutron-Ruault MC, Clavel-Chapelon F, Racine A, Lukanova A, Teucher B, Boeing H, Aleksandrova K, Trichopoulou A, Benetou V, Trichopoulos D, Grioni S, Vineis P, Panico S, Palli D, Tumino R, Siersema PD, Peeters PH, Skeie G, Brustad M, Chirlaque MD, Barricarte A, Ramon Quiros J, Sanchez MJ, Dorronsoro M, Bonet C, Palmqvist R, Hallmans G, Key TJ, Crowe F, Khaw KT, Wareham N, Romieu I, McKay J, Wark PA, Romaguera D, Jenab M: Prediagnostic 25-hydroxyvitamin d, vdr and casr polymorphisms, and survival in patients with colorectal cancer in western european ppulations. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2012;21:582-593.

141 McNamara D, Qasim A, Lee N, Condon C, O'Morain C: Round one of the adelaide and meath hospital/trinity college colorectal cancer screening programme: Programme report and analysis based on established international key performance indices. Irish journal of medical science 2011;180:549-552.

142 Brenner H, Kloor M, Pox CP: Colorectal cancer. Lancet 2014;383:1490-1502.

143 Michel A, Waterboer T, Kist M, Pawlita M: Helicobacter pylori multiplex serology. Helicobacter 2009;14:525-535.

144 Shet U, Oh HK, Chung HJ, Kim YJ, Kim OS, Lim HJ, Shin MH, Lee SW: Humoral immune responses to periodontal pathogens in the elderly. Journal of periodontal & implant science 2015;45:178-183.

145 Naito Y, Okuda K, Takazoe I: Immunoglobulin g response to subgingival gram-negative bacteria in human subjects. Infection and immunity 1984;45:47-51.

146 Vincent JW, Falkler WA, Jr., Cornett WC, Suzuki JB: Effect of periodontal therapy on specific antibody responses to suspected periodontopathogens. Journal of clinical periodontology 1987;14:412-417.

147 Wang HF, Li LF, Guo SH, Zeng QY, Ning F, Liu WL, Zhang G: Evaluation of antibody level against fusobacterium nucleatum in the serological diagnosis of colorectal cancer. Scientific reports 2016;6:33440.

148 Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC, Ala'Aldeen D: Type v protein secretion pathway: The autotransporter story. Microbiology and molecular biology reviews : MMBR 2004;68:692-744.

149 Cai H, Ye F, Michel A, Murphy G, Sasazuki S, Taylor PR, Qiao YL, Park SK, Yoo KY, Jee SH, Cho ER, Kim J, Chen SC, Abnet CC, Tsugane S, Cai Q, Shu XO, Zheng W, Pawlita M, Epplein M: Helicobacter pylori blood biomarker for gastric cancer risk in east asia. International journal of epidemiology 2016;45:774-781.

150 Albandar JM: Global risk factors and risk indicators for periodontal diseases. Periodontology 2000 2002;29:177-206.

151 Woof JM, Kerr MA: The function of immunoglobulin a in immunity. The Journal of pathology 2006;208:270-282.

152 Mima K, Cao Y, Chan AT, Qian ZR, Nowak JA, Masugi Y, Shi Y, Song M, da Silva A, Gu M, Li W, Hamada T, Kosumi K, Hanyuda A, Liu L, Kostic AD, Giannakis M, Bullman S, Brennan CA, Milner DA, Baba H, Garraway LA, Meyerhardt JA, Garrett WS, Huttenhower C, Meyerson M, Giovannucci EL, Fuchs CS, Nishihara R, Ogino S: Fusobacterium nucleatum in colorectal carcinoma tissue according to tumor location. Clinical and translational gastroenterology 2016;7:e200.

153 Demuth DR, Duan Y, Brooks W, Holmes AR, McNab R, Jenkinson HF: Tandem genes encode cell-surface polypeptides sspa and sspb which mediate adhesion of the oral bacterium streptococcus gordonii to human and bacterial receptors. Mol Microbiol 1996;20:403-413.

154 Brenner H, Altenhofen L, Kretschmann J, Rosch T, Pox C, Stock C, Hoffmeister M: Trends in adenoma detection rates during the first 10 years of the german screening colonoscopy program. Gastroenterology 2015;149:356-366 e351.

155 Silla IO, Rueda D, Rodriguez Y, Garcia JL, de la Cruz Vigo F, Perea J: Early-onset colorectal cancer: A separate subset of colorectal cancer. World journal of gastroenterology : WJG 2014;20:17288-17296.

156 Reynolds JG, Silva E, McCormack WM: Association of streptococcus bovis bacteremia with bowel disease. J Clin Microbiol 1983;17:696-697.

157 Ekbom A: Is diverticular disease associated with colonic malignancy? Digestive diseases 2012;30:46-50.

158 Agrawal S, Bhupinderjit A, Bhutani MS, Boardman L, Nguyen C, Romero Y, Srinivasan R, Figueroa-Moseley C, Committee of Minority A, Cultural Diversity ACoG: Colorectal cancer in african americans. Am J Gastroenterol 2005;100:515-523; discussion 514.

159 Rex DK, Johnson DA, Anderson JC, Schoenfeld PS, Burke CA, Inadomi JM, American College of G: American college of gastroenterology guidelines for colorectal cancer screening 2009 [corrected]. Am J Gastroenterol 2009;104:739-750.

160 Wallace K, Burke CA, Ahnen DJ, Barry EL, Bresalier RS, Saibil F, Baron JA: The association of age and race and the risk of large bowel polyps. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2015;24:448-453.

161 Http://www.Euro.Who.Int/en/health-topics/noncommunicable-diseases/diabetes/data-and-statistics, 2016,

162 Schulz C, Schutte K, Koch N, Vilchez-Vargas R, Wos-Oxley ML, Oxley AP, Vital M, Malfertheiner P, Pieper DH: The active bacterial assemblages of the upper gi tract in individuals with and without helicobacter infection. Gut 2016

163 Kienesberger S, Cox LM, Livanos A, Zhang XS, Chung J, Perez-Perez GI, Gorkiewicz G, Zechner EL, Blaser MJ: Gastric helicobacter pylori infection affects local and distant microbial populations and host responses. Cell reports 2016;14:1395-1407.

164 Wang L, Zhou J, Xin Y, Geng C, Tian Z, Yu X, Dong Q: Bacterial overgrowth and diversification of microbiota in gastric cancer. Eur J Gastroenterol Hepatol 2016;28:261-266.

165 McDonald JR: Acute infective endocarditis. Infectious disease clinics of North America 2009;23:643-664.

166 Kononen E, Kanervo A, Takala A, Asikainen S, Jousimies-Somer H: Establishment of oral anaerobes during the first year of life. Journal of dental research 1999;78:1634-1639.

167 Oliver RC, Brown LJ, Loe H: Periodontal diseases in the united states population. Journal of periodontology 1998;69:269-278.

168 Hwang AM, Stoupel J, Celenti R, Demmer RT, Papapanou PN: Serum antibody responses to periodontal microbiota in chronic and aggressive periodontitis: A postulate revisited. Journal of periodontology 2014;85:592-600.

169 Tew JG, Marshall DR, Moore WE, Best AM, Palcanis KG, Ranney RR: Serum antibody reactive with predominant organisms in the subgingival flora of young adults with generalized severe periodontitis. Infection and immunity 1985;48:303-311.

170 Bashir A, Miskeen AY, Hazari YM, Asrafuzzaman S, Fazili KM: Fusobacterium nucleatum, inflammation, and immunity: The fire within human gut. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 2016;37:2805-2810.

171 Gur C, Mandelboim O, Bachrach G: "Messieurs, c'est les microbes qui auront le dernier mot": Gentlemen, it is the microbes who have the last word (louis pasteur)-fusobacterium nucleatum protect tumors from killing by immune cells. Oncoimmunology 2015;4:e1038690.

172 zur Hausen H: Streptococcus bovis: Causal or incidental involvement in cancer of the colon? International journal of cancer Journal international du cancer 2006;119:xi-xii.

173 Shimada T, Terano A: Chemokine expression in helicobacter pylori-infected gastric mucosa. J Gastroenterol 1998;33:613-617.

174 Allison CC, Kufer TA, Kremmer E, Kaparakis M, Ferrero RL: Helicobacter pylori induces mapk phosphorylation and ap-1 activation via a nod1-dependent mechanism. J Immunol 2009;183:8099-8109.

175 Yamada Y, Mori H: Multistep carcinogenesis of the colon in apc(min/+) mouse. Cancer science 2007;98:6-10.

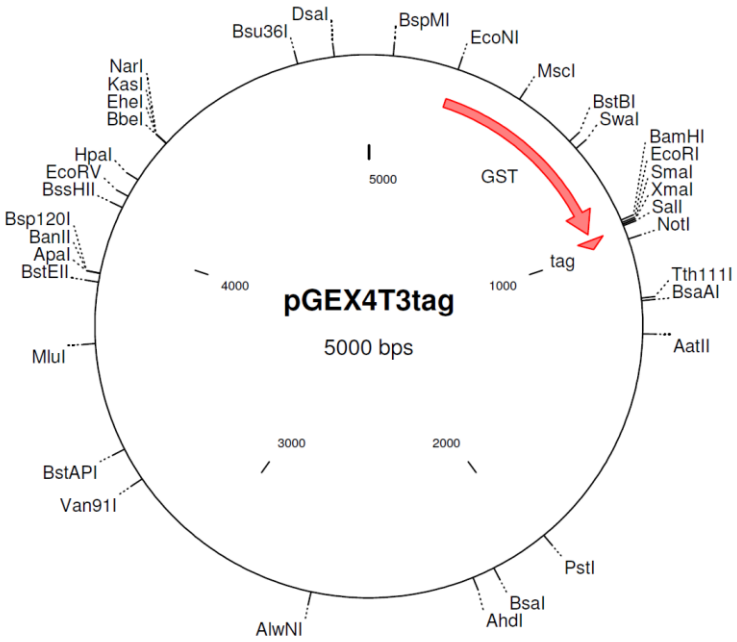
References

176 Aran V, Victorino AP, Thuler LC, Ferreira CG: Colorectal cancer: Epidemiology, disease mechanisms and interventions to reduce onset and mortality. Clinical colorectal cancer 2016;15:195-203.

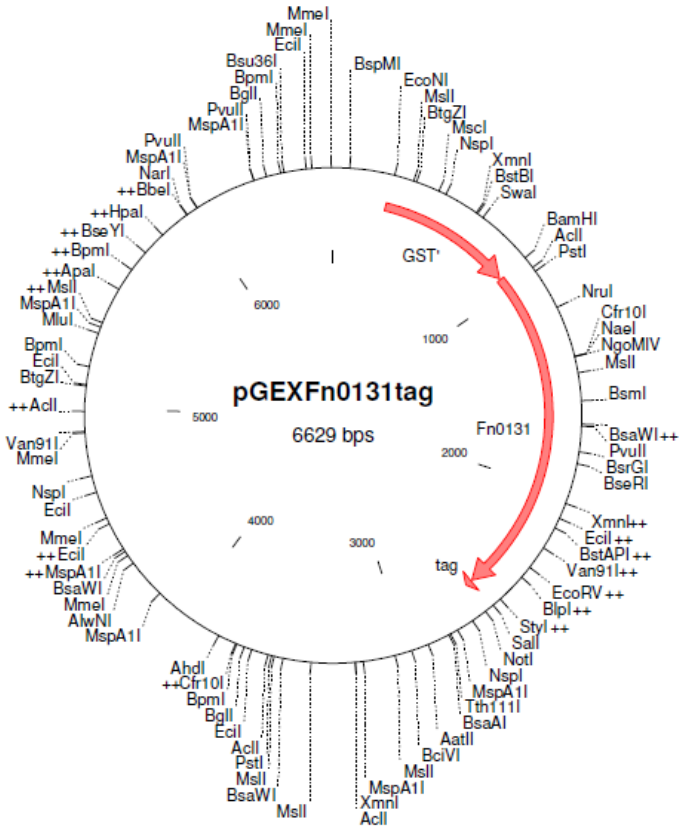
Appendix

I. Vector maps

pGEX4T3tag

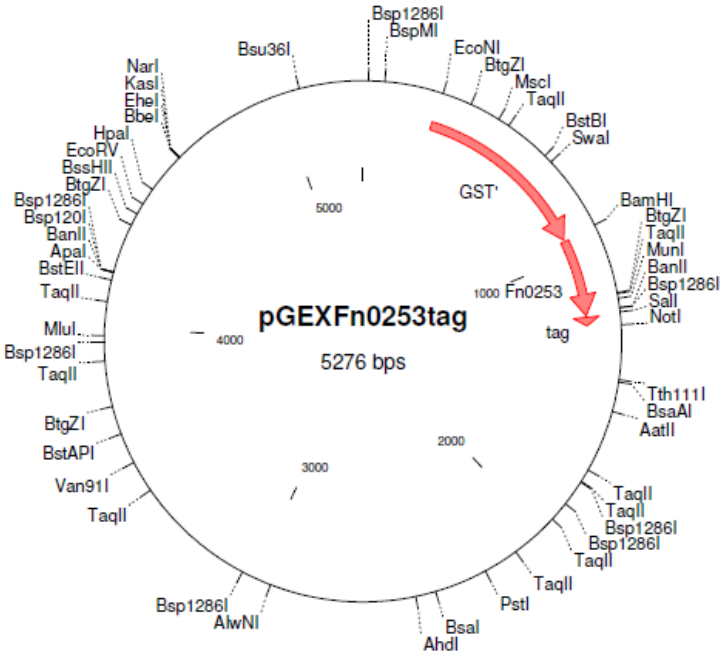


pGEXFn0131tag

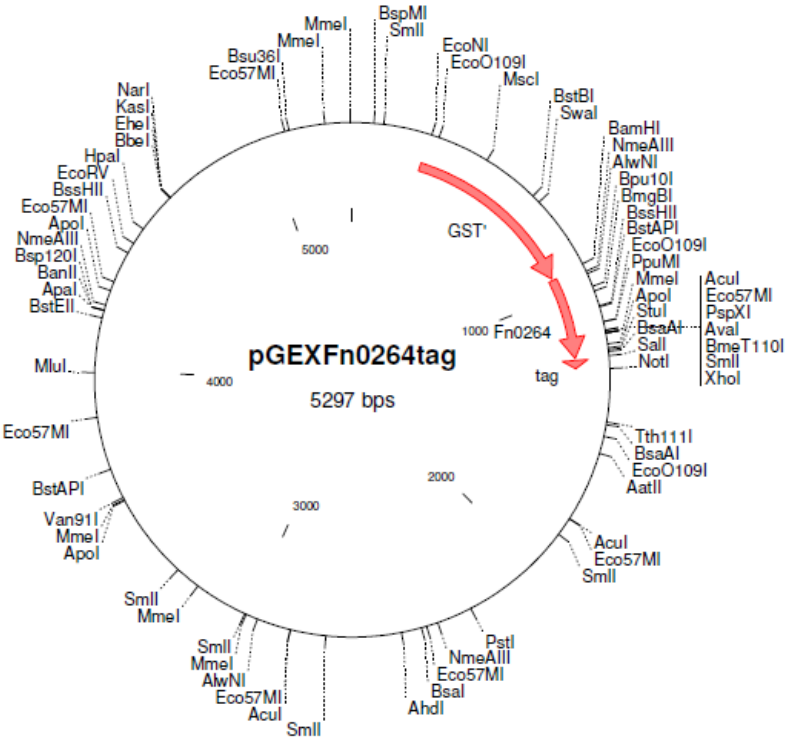


Appendix

pGEXFn0253tag

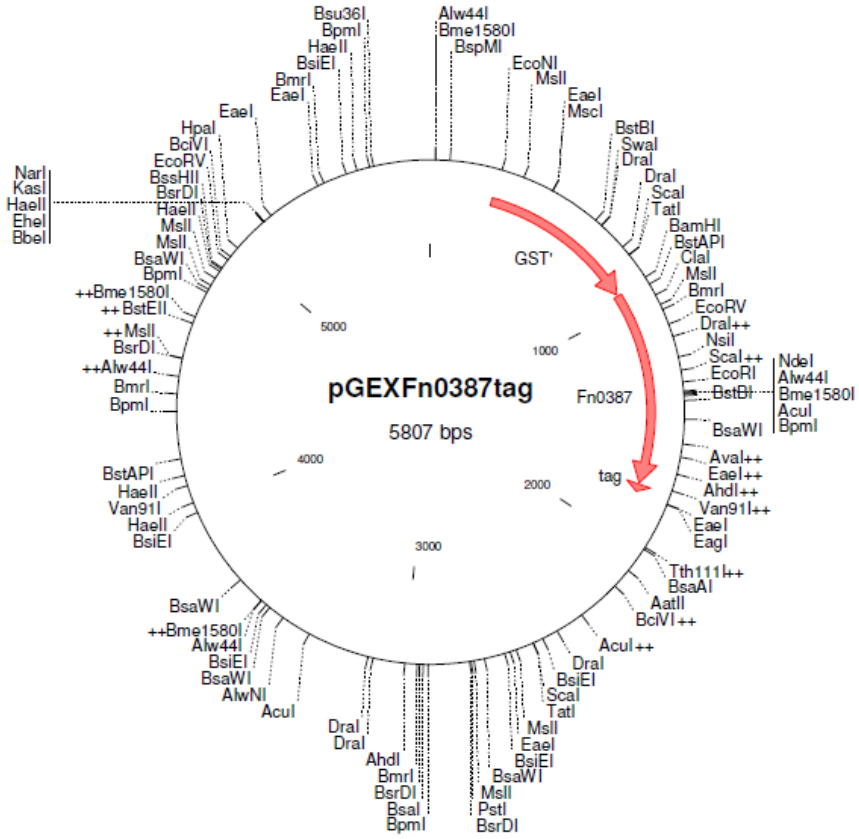


pGEXFn0264tag

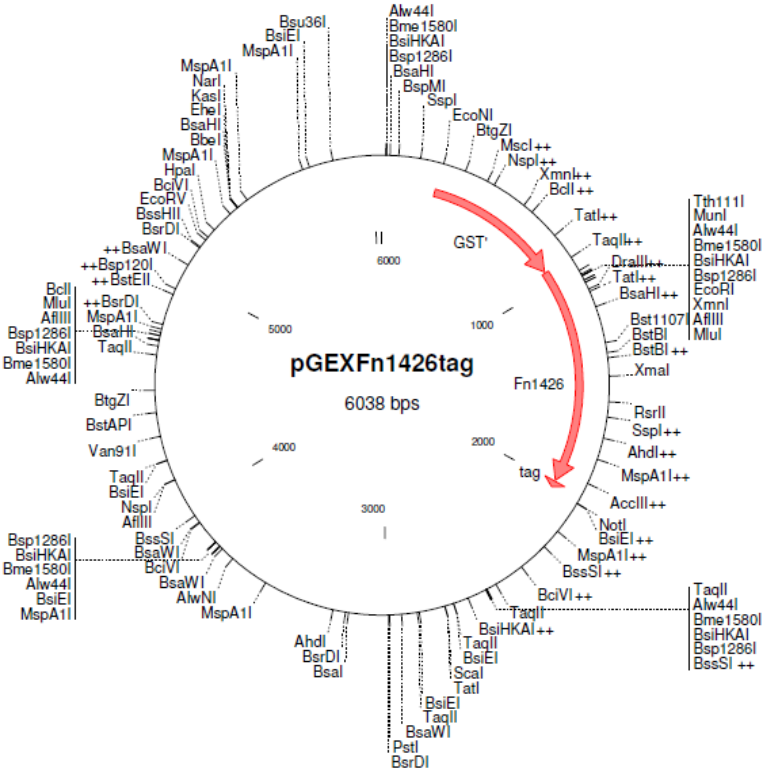


Appendix

pGEXFn0387tag

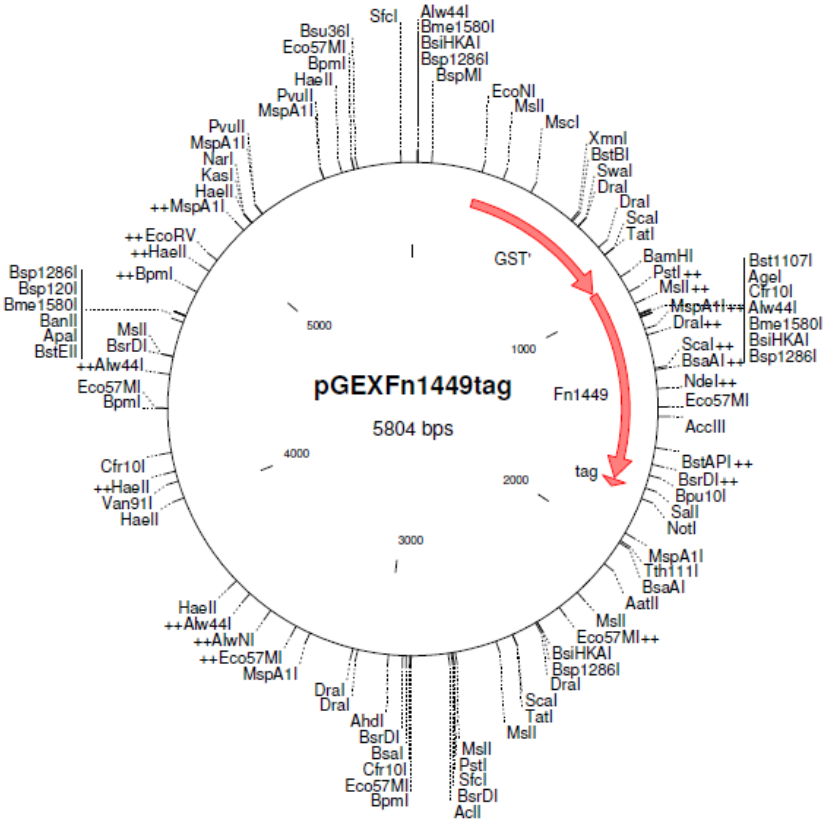


pGEXFn1426tag

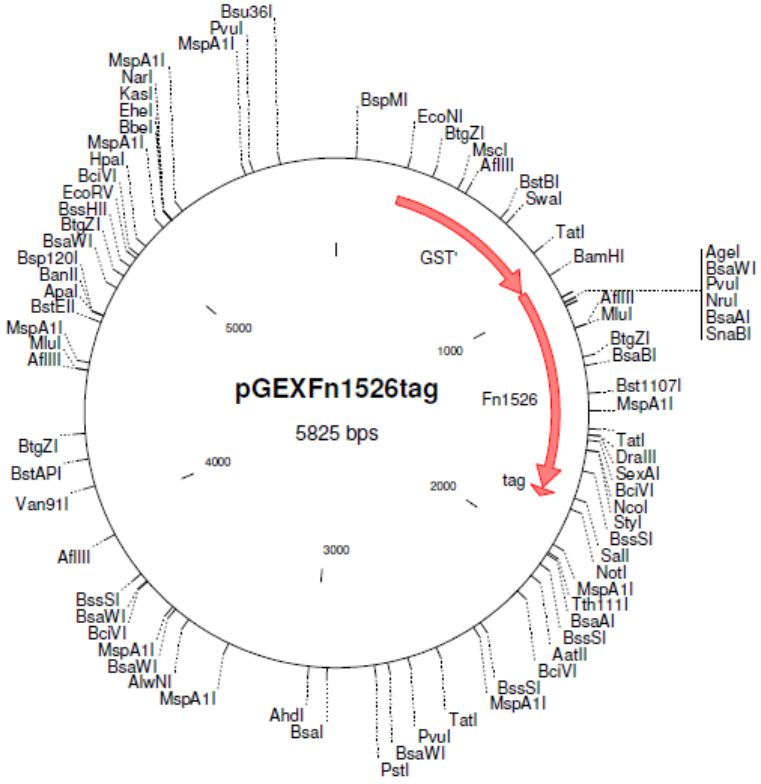


Appendix

pGEXFn1449tag

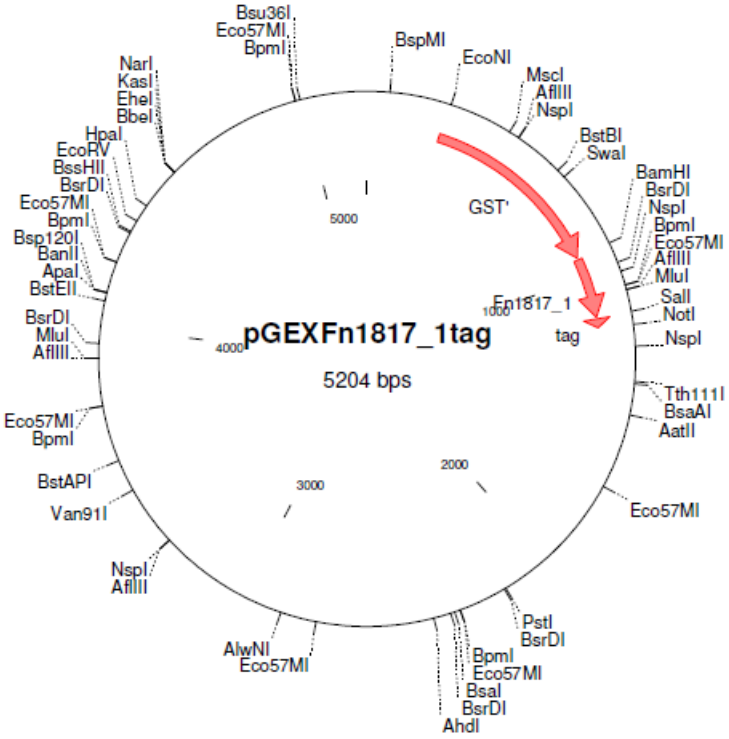


pGEXFn1526tag

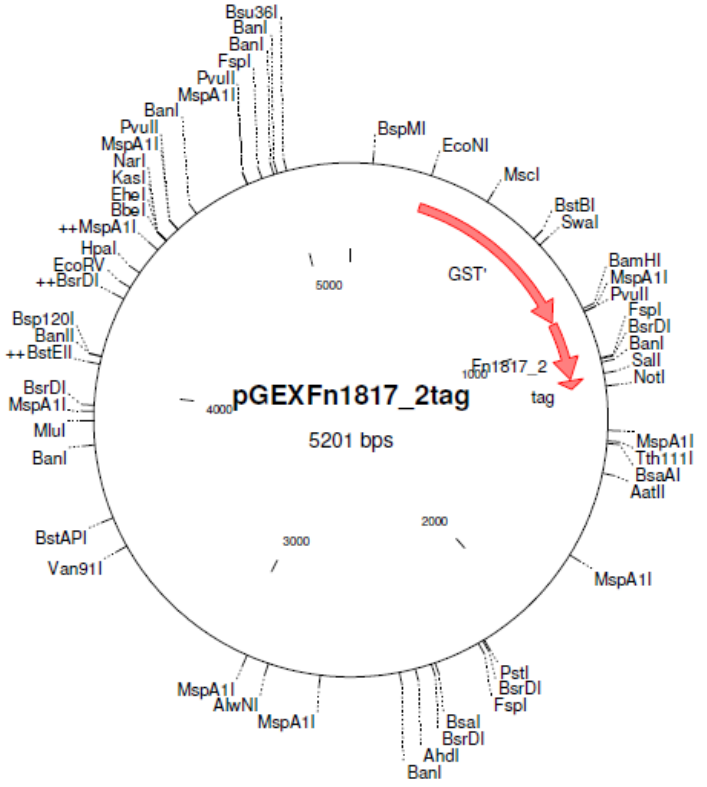


Appendix

pGEXFn1817_1tag

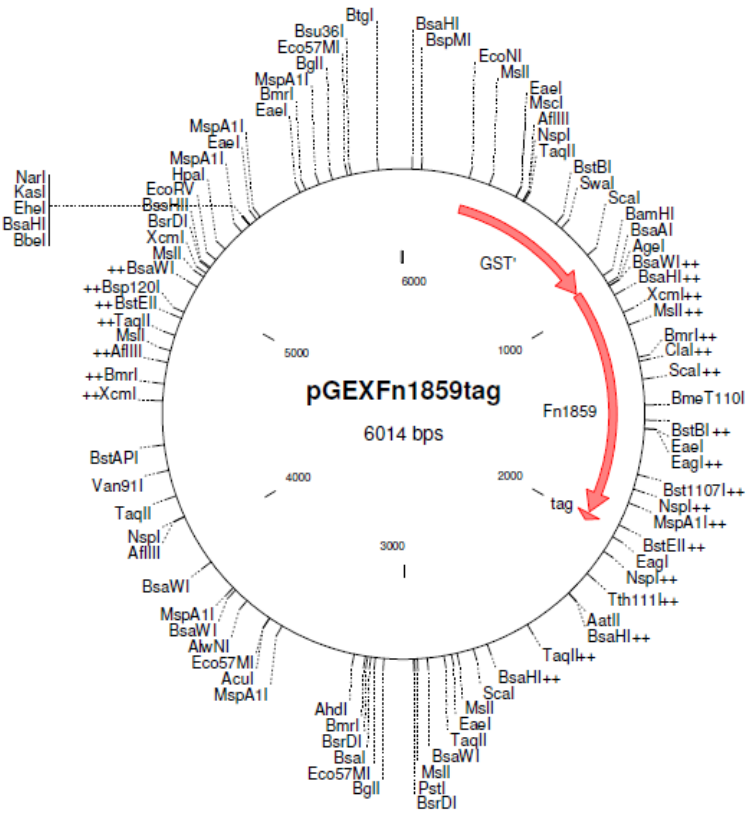


pGEXFn1817_2tag

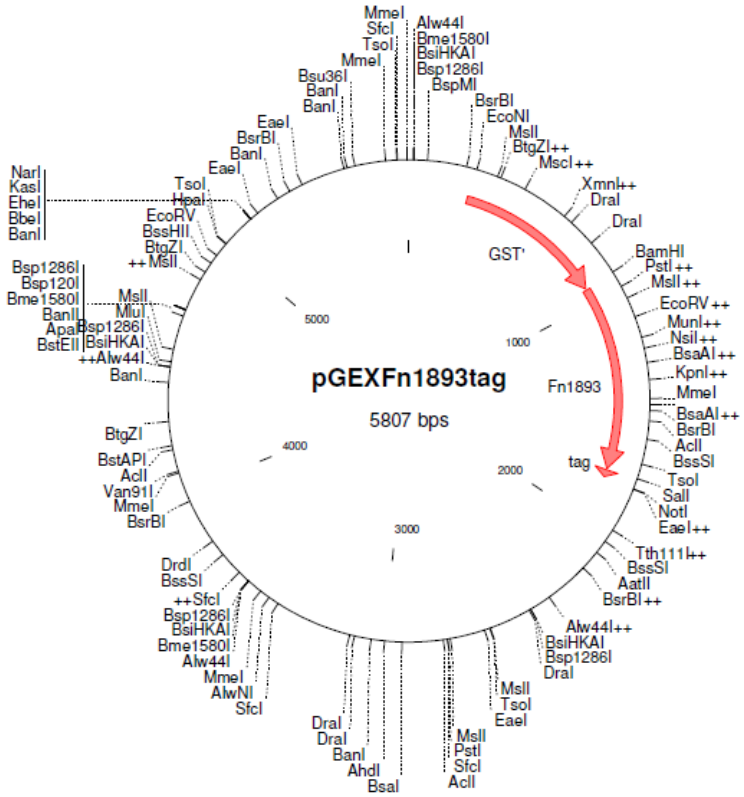


Appendix

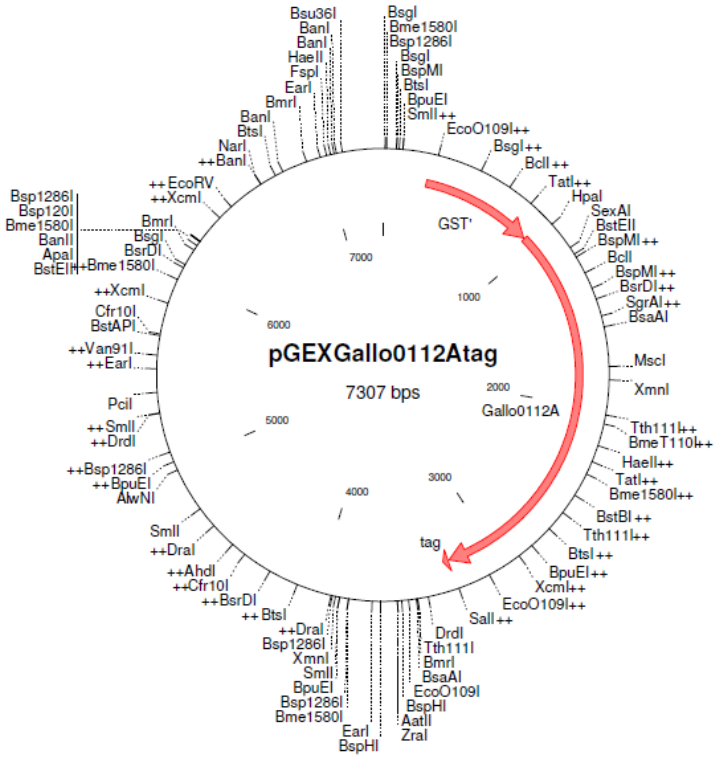
pGEXFn1859tag



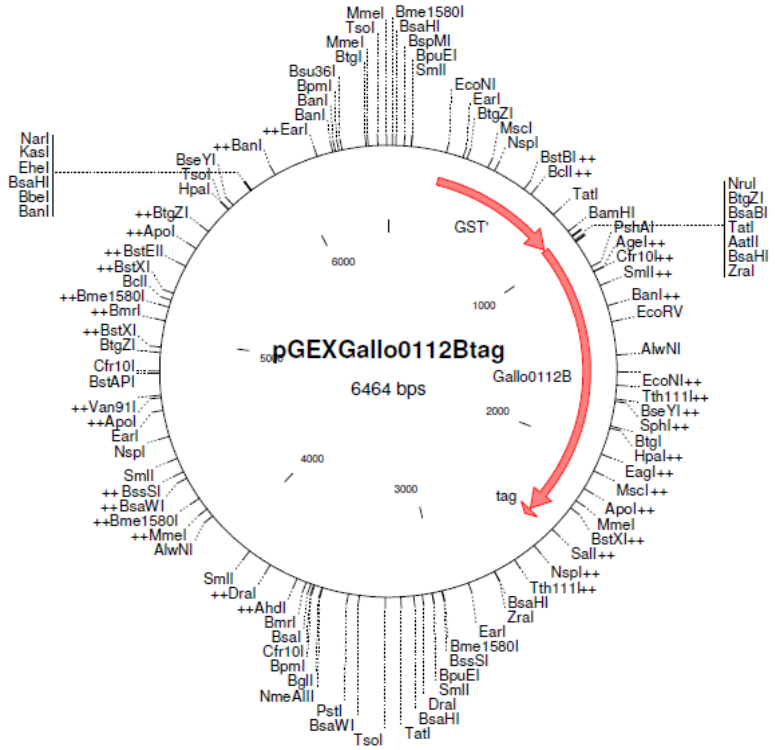
pGEXFn1893tag



pGEXGallo0112Atag

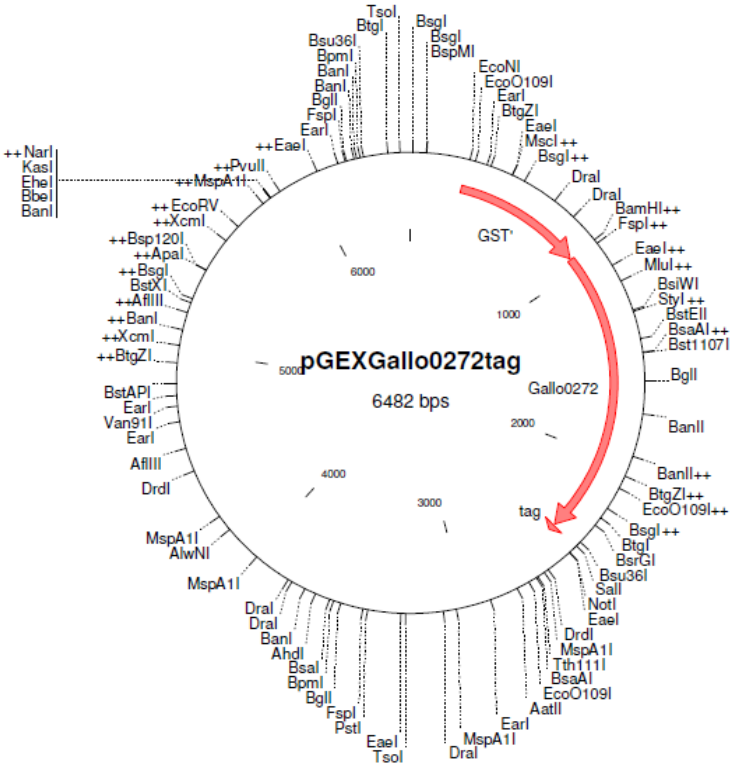


pGEXGallo0112Btag

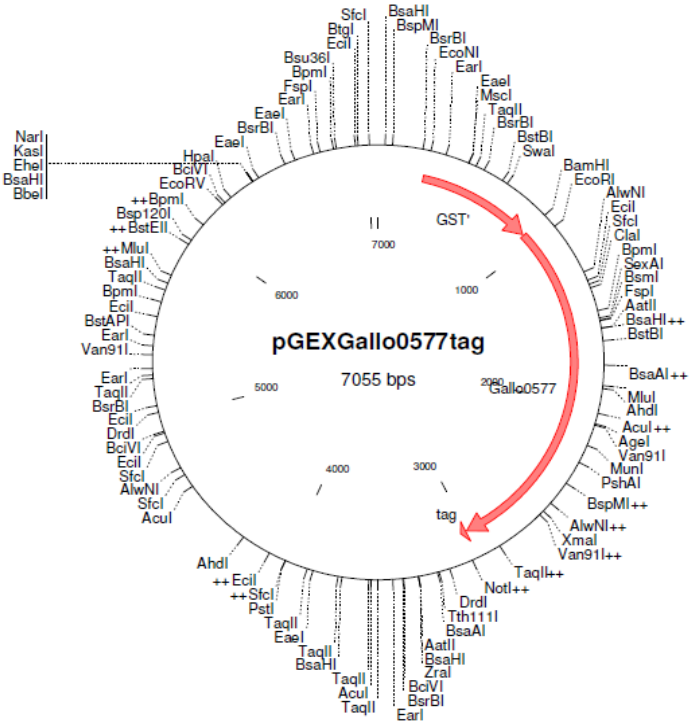


Appendix

pGEXGallo0272tag

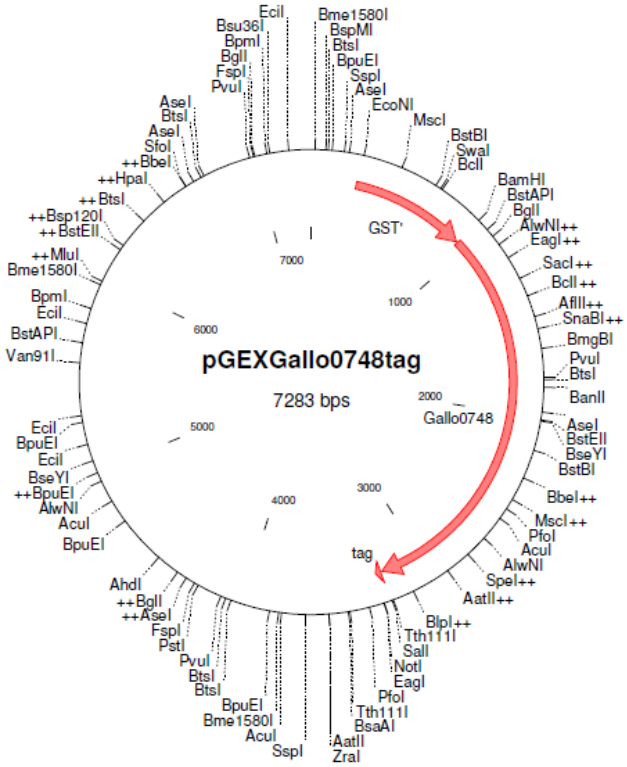


pGEXGallo0577tag

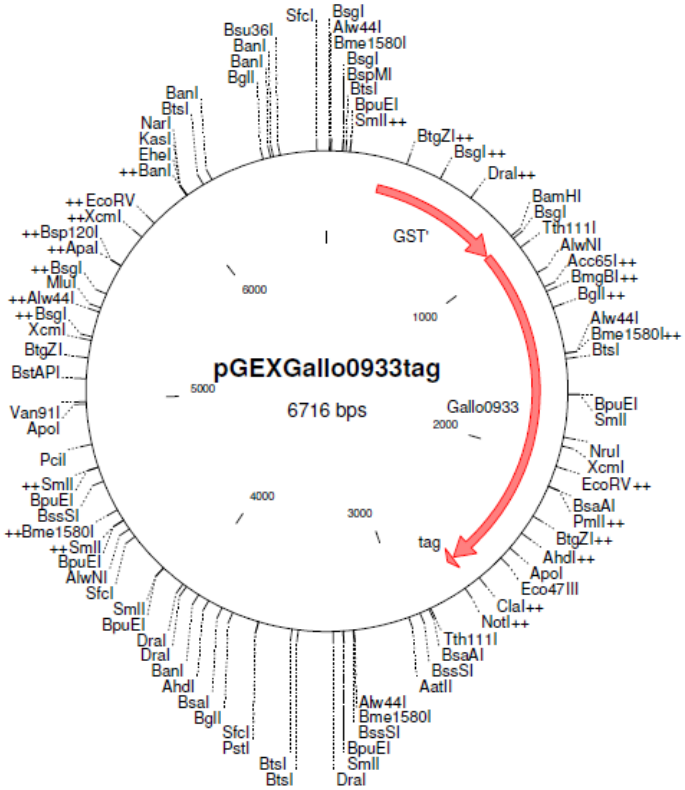


Appendix

pGEXGallo0748tag

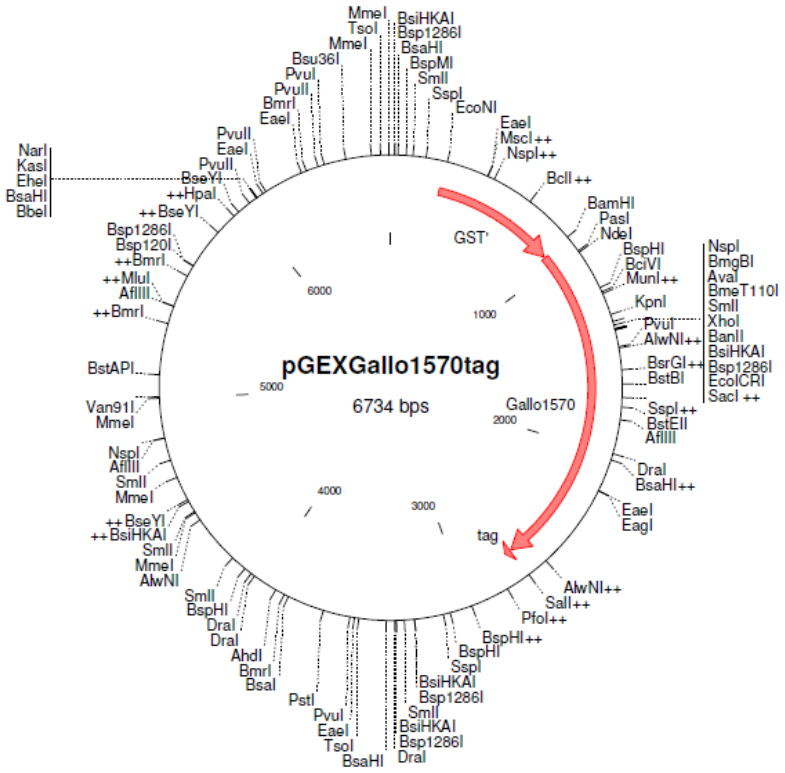


pGEXGallo0933tag

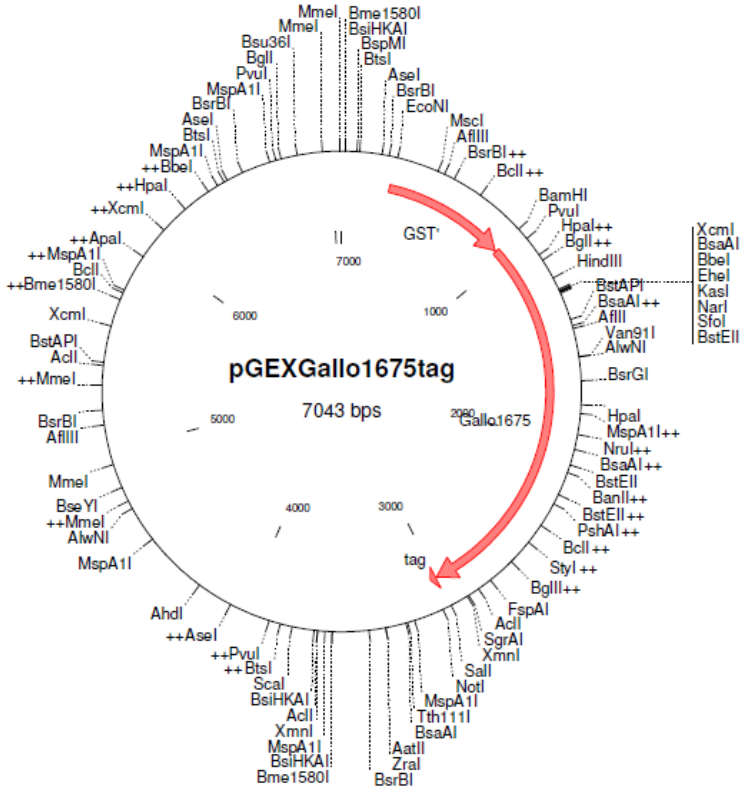


Appendix

pGEXGallo1570tag

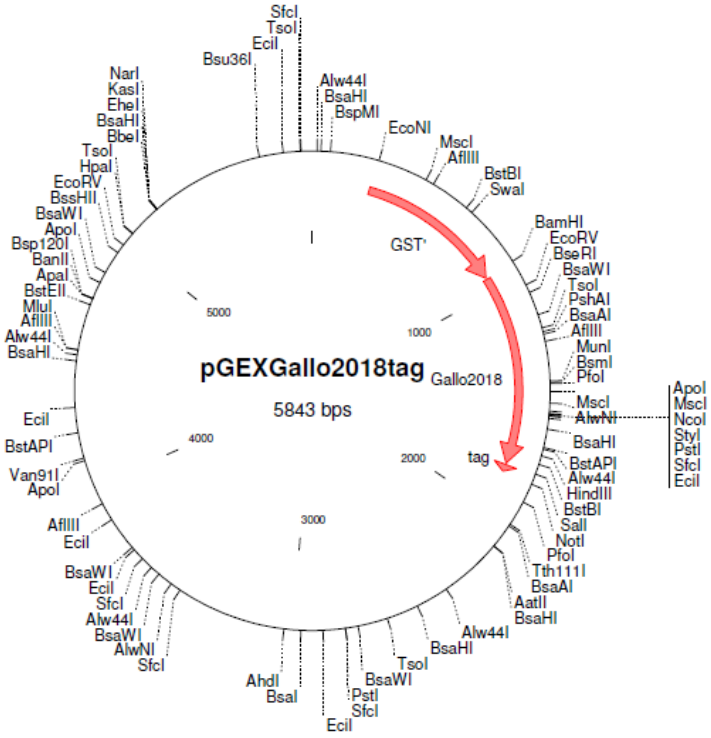


pGEXGallo1675tag

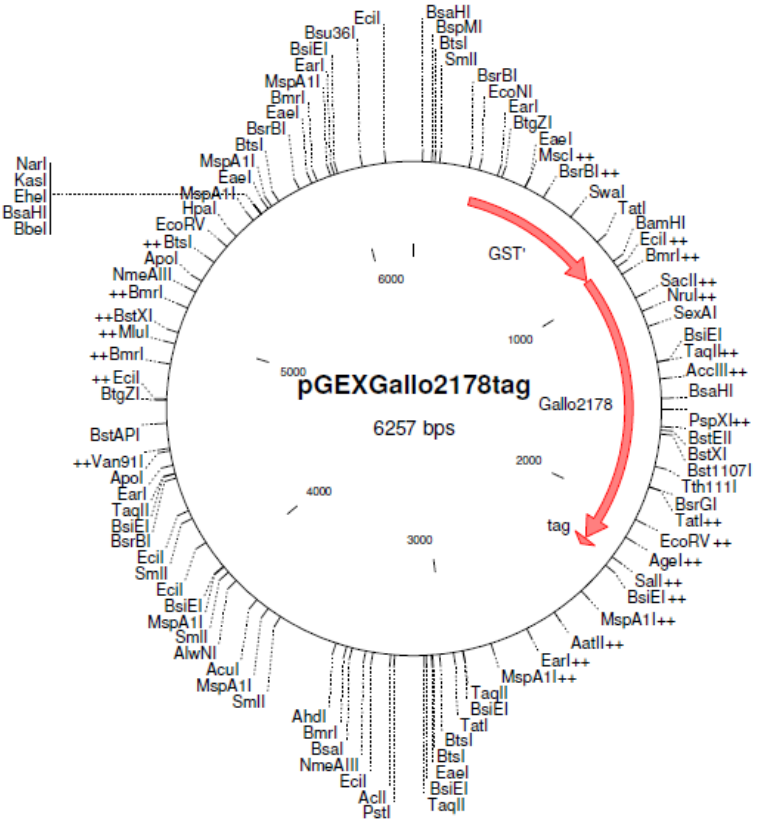


Appendix

pGEXGallo2018tag

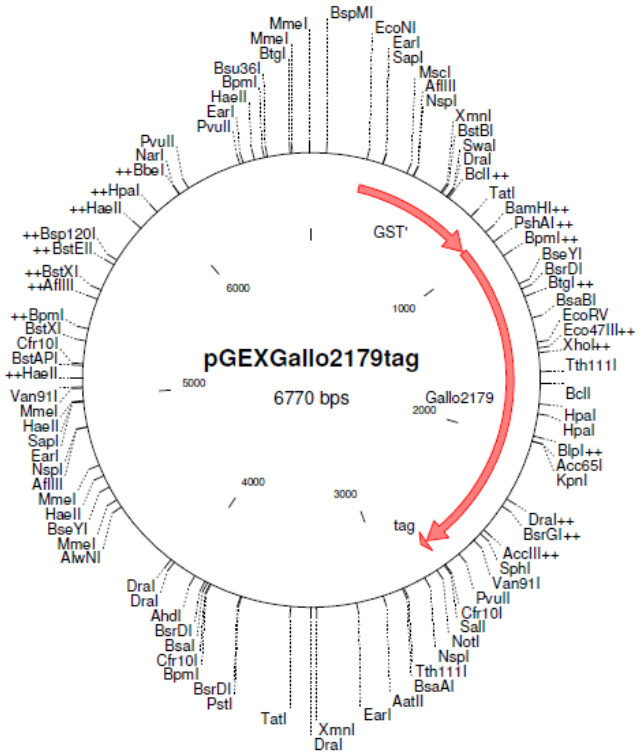


pGEXGallo2178tag



Appendix

pGEXGallo2179tag



II. Alignments of sequenced inserts to reference insert sequences

(light blue: BamHI restriction site, light green: SalI restriction site, orange: pGEXfor, brown: pGEXrev, purple: pGEXs, red: pGEXas)

Fn0131 (forward sequencing of miniPrep DNA via primer pGEXfor, reverse sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn0131_exp      ATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA
Fn0131_Seq      ----------------------------------------------------ATCCTCCAAA
                                                                     *  * ***

Fn0131_exp      TCGGATCTGGTTCCGCGTGGATCCTCTTTTAACGAGAACGAGGATGAACGTACCATTCTG
Fn0131_Seq      TCGGATCTGGTTCCGCGTGGATCCTCTTTTAACGAGAACGAGGATGAACGTACCATTCTG
                ************************************************************

Fn0131_exp      AAACAGGAACAACGTTCGGAACAAGAACGTCTGCAGAAAGAATTTCAGAAACGCGAAGAA
Fn0131_Seq      AAACAGGAACAACGTTCGGAACAAGAACGTCTGCAGAAAGAATTTCAGAAACGCGAAGAA
                ************************************************************

Fn0131_exp      ATTTTCAACCAACTGAAAAGCGAGAAAACGGACAAGCAAGAAGTGAGCACCAACGAAATC
Fn0131_Seq      ATTTTCAACCAACTGAAAAGCGAGAAAACGGACAAGCAAGAAGTGAGCACCAACGAAATC
                ************************************************************

Fn0131_exp      AAATTCCACATTTCACAGATCAATCTGGAAGATAATGAACGGCTCCTGAATGAAATCGAA
Fn0131_Seq      AAATTCCACATTTCACAGATCAATCTGGAAGATAATGAACGGCTCCTGAATGAAATCGAA
                ************************************************************

Fn0131_exp      AAGGAGAATATTCTGGGCAAATACATCAATCGCGATTTAGGGTCTACAGACATCACGAAT
Fn0131_Seq      AAGGAGAATATTCTGGGCAAATACATCAATCGCGATTTAGGGTCTACAGACATCACGAAT
                ************************************************************

Fn0131_exp      CTGATTACGGACCTTACCAATCGTCTGATCGCTAAAGGCTATATTACGAGCGTAGCGACC
Fn0131_Seq      CTGATTACGGACCTTACCAATCGTCTGATCGCTAAAGGCTATATTACGAGCGTAGCGACC
                ************************************************************

Fn0131_exp      ATCAGTGAGGATAACGACTTATCTACTAAAACGCTTAATCTCAAAATCATTCCGGGGAAG
Fn0131_Seq      ATCAGTGAGGATAACGACTTATCTACTAAAACGCTTAATCTCAAAATCATTCCGGGGAAG
                ************************************************************

Fn0131_exp      ATCGAGAAAATTATCCTGAACGAAGATAAAACCCTCGACAACCTGAAGAAATATTTTCTG
Fn0131_Seq      ATCGAGAAAATTATCCTGAACGAAGATAAAACCCTCGACAACCTGAAGAAATATTTTCTG
                ************************************************************

Fn0131_exp      GTGGACACTAAAGCCGGCAAGGTGTTAAACATTCGTGATTTGGACACTACGACAGAAAAT
Fn0131_Seq      GTGGACACTAAAGCCGGCAAGGTGTTAAACATTCGTGATTTGGACACTACGACAGAAAAT
                ************************************************************

Fn0131_exp      TTCAATTATCTGGAAGCCAACAACATGACTATGGAGATCATTCCGAGTGAAATCCAAAAC
Fn0131_Seq      TTCAATTATCTGGAAGCCAACAACATGACTATGGAGATCATTCCGAGTGAAATCCAAAAC
                ************************************************************

Fn0131_exp      CATTCCATTGTCAAACTGAAGAACGAAATGAAAGAGAAGTTCACGGTGAGTGTGCTGACC
Fn0131_Seq      CATTCCATTGTCAAACTGAAGAACGAAATGAAAGGAGAAGGTTCACGGTGGAGTGTGCTG
                ********************************   *     *   * *     **

Fn0131_exp      AACAACTACGGCGAAGATCGTCAGAATGCTATTTGGCGCGGCGGTGTCTCAATTAACATT
Fn0131_Seq      ACCAA-------------------------------------------------------
                * ***

Fn0131_exp      GATTCGCCATTAGGGATCGGTGATCGCGTGTACTTTTCCTATATGACAGTGCACAAGAAA
Fn0131_Seq      ----------------------------------------------CAGTGCACAAGAAA
                                                              **************
```

```
Fn0131_exp      AAACCGGATCGCAGCTGGAAACGGACAACAGAATCCCTCAAACCTGGCGAAATTGCACCT
Fn0131_Seq      AAACCGGATCGCAGCTGGAAACGGACAACAGAATCCCTCAAACCTGGCGAAATTGCACCT
                ************************************************************

Fn0131_exp      ATTGGTCCGAAAGGCTATGATCCACGCAAAGATACCTTGCCGTATAAACGCGACTTGGAT
Fn0131_Seq      ATTGGTCCGAAAGGCTATGATCCACGCAAAGATACCTTGCCGTATAAACGCGACTTGGAT
                ************************************************************

Fn0131_exp      TTGTACAATTTCCGCTATACACTGAAATTCAATTCCTATACCCTGAGCTTAGGCTCCTCC
Fn0131_Seq      TTGTACAATTTCCGCTATACACTGAAATTCAATTCCTATACCCTGAGCTTAGGCTCCTCC
                ************************************************************

Fn0131_exp      CGCATTGAAAATACGAGTTCGTTTTATACCCCGAATACCGTGTACGACATGGAAACCGTT
Fn0131_Seq      CGCATTGAAAATACGAGTTCGTTTTATACCCCGAATACCGTGTACGACATGGAAACCGTT
                ************************************************************

Fn0131_exp      AGCAATACGTTTTCAGTAAACTTGGATAAAGTCCTGTTACGCAACCAGAAGAATAAACTG
Fn0131_Seq      AGCAATACGTTTTCAGTAAACTTGGATAAAGTCCTGTTACGCAACCAGAAGAATAAACTG
                ************************************************************

Fn0131_exp      ACGTTTGGTATTGGGCTGAAACGGAAACATAATCAGTCGTACATCGAGGAAGCGATTCTG
Fn0131_Seq      ACGTTTGGTATTGGGCTGAAACGGAAACATAATCAGTCGTACATCGAGGAAGCGATTCTG
                ************************************************************

Fn0131_exp      AGTGATCGTGTCTTAACGATTGGAGACATTAGCCTGAACGGCACTACCACCTTTTATGGC
Fn0131_Seq      AGTGATCGTGTCTTAACGATTGGAGACATTAGCCTGAACGGCACTACCACCTTTTATGGC
                ************************************************************

Fn0131_exp      GGACTGCTGGGTGCAAGCCTGGGATACGAACGTGGCATGCGTGCACTGGGTGCGGAACGT
Fn0131_Seq      GGACTGCTGGGTGCAAGCCTGGGATACGAACGTGGCATGCGTGCACTGGGTGCGGAACGT
                ************************************************************

Fn0131_exp      GATAAGAATAAAGGCGTTCGCTCTCCCAAAGCGGAGTTTATGAAGTATACCCTGAACACT
Fn0131_Seq      GATAAGAATAAAGGCGTTCGCTCTCCCAAAGCGGAGTTTATGAAGTATACCCTGAACACT
                ************************************************************

Fn0131_exp      AACTACTACAAACCCCTTACCCAGAAACTGGTATACCGCTTTAACACCAATATCACCTAT
Fn0131_Seq      AACTACTACAAACCCCTTACCCAGAAACTGGTATACCGCTTTAACACCAATATCACCTAT
                ************************************************************

Fn0131_exp      TCGAACGATGTTCTCTATGGGTCGGAGAAACACTCTATTGGTGGTGTCGGCTCAGTTGGT
Fn0131_Seq      TCGAACGATGTTCTCTATGGGTCGGAGAAACACTCTATTGGTGGTGTCGGCTCAGTTGGT
                ************************************************************

Fn0131_exp      GGATATCATCGCACTGGGAATATTCAGGGTGATAAAGCCATCGAGATTGAAAACGAGCTG
Fn0131_Seq      GGATATCATCGCACTGGGAATATTCAGGGTGATAAAGCCATCGAGATTGAAAACGAGCTG
                ************************************************************

Fn0131_exp      AGCTACCGTGTTCTGGACTCTGAAAAGTTCGGCAAAATCACCCCGTATCTTAGCTACTCA
Fn0131_Seq      AGCTACCGTGTTCTGGACTCTGAAAAGTTCGGCAAAATCACCCCGTATCTTAGCTACTCA
                ************************************************************

Fn0131_exp      TATGGTAAAGTTCGCAACAACAAAAACAATAGCAAATACCGCAAAGGTTACATGAGTGGC
Fn0131_Seq      TATGGTAAAGTTCGCAACAACAAAAACAATAGCAAATACCGCAAAGGTTACATGAGTGGC
                ************************************************************

Fn0131_exp      GCGATCCTTGGTTTGCGCTATAACATGAAATATCTCCAGTTGACCGTCGACAAACCTCCC
Fn0131_Seq      GCGATCCTTGGTTTGCGCTATAACATGAAATATCTCCAGTTGACCGTCGACAAACCTCCC
                ************************************************************

Fn0131_exp      ACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATCGTGACTGACTGACGATCTGCCT
Fn0131_Seq      ACACCTCCCCCTAAC---------------------------------------------
                ************ *

Fn0131_exp      CGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTC
Fn0131_Seq      -----------------------------------------------
```

Fn0253 (sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn0253_exp      --------------------TCCAAAATCGGATCTGGTTCCGCGTGGATCCAACTTCGA
Fn0253_Seq      TTAATACGACTCACTATAGGGTCCAAAATCGGATCTGGTTCCGCGTGGATCCAACTTCGA
                                    ************************************

Fn0253_exp      CTTTGACAAGTCGAATGTGAAACCGCAGTATTACGATCTGCTGAACAACATCAAGGAGTT
Fn0253_Seq      CTTTGACAAGTCGAATGTGAAACCGCAGTATTACGATCTGCTGAACAACATCAAGGAGTT
                ************************************************************

Fn0253_exp      TGTGGAACAGAACAACTATGAGATCACCATTGTAGGACATACGGATTCAATCGGCTCTAA
Fn0253_Seq      TGTGGAACAGAACAACTATGAGATCACCATTGTAGGACATACGGATTCAATCGGCTCTAA
                ************************************************************

Fn0253_exp      TGCCTACAACTTCAAACTGAGTCGTCGTCGCGCAGAAAGCGTCAAAGCGAAACTCTTGGA
Fn0253_Seq      TGCCTACAACTTCAAACTGAGTCGTCGTCGCGCAGAAAGCGTCAAAGCGAAACTCTTGGA
                ************************************************************

Fn0253_exp      GTTTGGGTTATCCGAAGATCGCATTGTTGGCATTGAAGCGATGGGTGAAGAACAGCCAAT
Fn0253_Seq      GTTTGGGTTATCCGAAGATCGCATTGTTGGCATTGAAGCGATGGGTGAAGAACAGCCAAT
                ************************************************************

Fn0253_exp      TGCCACTAATGCGACCAAAGAAGGTCGGGCTCAAAATCGCGTCGACAAACCTCCCACACC
Fn0253_Seq      TGCCACTAATGCGACCAAAGAAGGTCGGGCTCAAAATCGCGTCGACAAACCTCCCACACC
                ************************************************************

Fn0253_exp      TCCCCCTGAACCTGAAACATAAGCGGCCGCATC
Fn0253_Seq      TCCCCCT-------------------------
                *******
```

Fn0264 (sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn0264_exp      ------------------------------TCCAAAATCGGATCTGGTTCCGCGTGGAT
Fn0264_Seq      ACTGAAGCTTTTAATACGACTCACTATAGGGTCCAAAATCGGATCTGGTTCCGCGTGGAT
                                              ****************************

Fn0264_exp      CCGGCGAACTGCAAGCTCTGGATGCCGAGTATCAGAACCTGGCGAATCAGGAAGAAGCCC
Fn0264_Seq      CCGGCGAACTGCAAGCTCTGGATGCCGAGTATCAGAACCTGGCGAATCAGGAAGAAGCCC
                ************************************************************

Fn0264_exp      GGTTTAACGAAGAACGCGCTCAGGCTGATGCAGCACGTCAGGCATTGGCACAGAACGAAC
Fn0264_Seq      GGTTTAACGAAGAACGCGCTCAGGCTGATGCAGCACGTCAGGCATTGGCACAGAACGAAC
                ************************************************************

Fn0264_exp      AGGTGTATAACGAACTGAGTCAACGCGCGCAACGTCTGCAAGCGGAAGCGAATACCCGCT
Fn0264_Seq      AGGTGTATAACGAACTGAGTCAACGCGCGCAACGTCTGCAAGCGGAAGCGAATACCCGCT
                ************************************************************

Fn0264_exp      TCTACAAATCGCAGTATCAGGACCTTGCGAGCAAATACGAAGATGCCCTGAAGAAACTCG
Fn0264_Seq      TCTACAAATCGCAGTATCAGGACCTTGCGAGCAAATACGAAGATGCCCTGAAGAAACTCG
                ************************************************************

Fn0264_exp      AGTCTGAGATGGAACAACAGAAAGCCATTATCTCCGACTTTGAGAAAATTCAGGCCTTAC
Fn0264_Seq      AGTCTGAGATGGAACAACAGAAAGCCATTATCTCCGACTTTGAGAAAATTCAGGCCTTAC
                ************************************************************

Fn0264_exp      GTGCGGGTAATGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAGCGGCCG
Fn0264_Seq      GTGCGGGTAATGTCGACAAACCTCCCACACCTCCCCCTAA--------------------
                ************************************.*

Fn0264_exp      CATC
Fn0264_Seq      ----
```

Fn0387 (sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn0387_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCTCTAAAGACTCGAACAAAATCAAAGCATT
Fn0387_Seq      ----AAATTGGATCTGGTTCCGCGTGGATCCTCTAAAGACTCGAACAAAATCAAAGCATT
                    **** **************************************************

Fn0387_exp      CGGTGCTCGTGGTGAATACAAGACCAATACTGCTGGTGTGATCGATTACAAGAACTATGC
Fn0387_Seq      CGGTGCTCGTGGTGAATACAAGACCAATACTGCTGGTGTGATCGATTACAAGAACTATGC
                ************************************************************

Fn0387_exp      GTATGGTGTTGCCTATATTCACGAGAATGAAAGTGTAAAACTGGGGAAAGACATCGGCTG
Fn0387_Seq      GTATGGTGTTGCCTATATTCACGAGAATGAAAGTGTAAAACTGGGGAAAGACATCGGCTG
                ************************************************************

Fn0387_exp      GTATACCGGGTTCGTACACAACACGTTTCGCTTTGAGGATATCGGTAAATCCAAAGAGGA
Fn0387_Seq      GTATACCGGGTTCGTACACAACACGTTTCGCTTTGAGGATATCGGTAAATCCAAAGAGGA
                ************************************************************

Fn0387_exp      AATGCTGTTAGGCAAAATCGGAATGTTTAAAAGCATTCCGTTTGATGACGATAATTCACT
Fn0387_Seq      AATGCTGTTAGGCAAAATCGGAATGTTTAAAAGCATTCCGTTTGATGACGATAATTCACT
                ************************************************************

Fn0387_exp      GAACTGGACAGTCAGTGGCAATGTGTTTGTTGGTCGCAACAAAATGCATCGGAAATTCCT
Fn0387_Seq      GAACTGGACAGTCAGTGGCAATGTGTTTGTTGGTCGCAACAAAATGCATCGGAAATTCCT
                ************************************************************

Fn0387_exp      CATTGTGGACGAAATCTTCAACGCGAAAAGCAAGTACTACGCGTATGGCATTGGGGTCAA
Fn0387_Seq      CATTGTGGACGAAATCTTCAACGCGAAAAGCAAGTACTACGCGTATGGCATTGGGGTCAA
                ************************************************************

Fn0387_exp      GAACGAAATTGGCAAAGAATTCCGGCTTTCTGAGGACTTTAGCATTCGTCCATATGGTGC
Fn0387_Seq      GAACGAAATTGGCAAAGAATTCCGGCTTTCTGAGGACTTTAGCATTCGTCCATATGGTGC
                ************************************************************

Fn0387_exp      ACTGAAGCTGGAGTACGGACGCATTTCGAAAATCAAAGAGAAAACGGGCGAAATTCGCCT
Fn0387_Seq      ACTGAAGCTGGAGTACGGACGCATTTCGAAAATCAAAGAGAAAACGGGCGAAATTCGCCT
                ************************************************************

Fn0387_exp      GGAAGTGAAAAGCAACGATTATGTCTCCATTAAACCGGAAATTGGCACGGAACTTAAATA
Fn0387_Seq      GGAAGTGAAAAGCAACGATTATGTCTCCATTAAACCGGAAATTGGCACGGAACTTAAATA
                ************************************************************

Fn0387_exp      CAAATATCTGTTCACCAACCGCAAAACCTTGACCGTTGGTTTGGGCGTTGCGTATGAGAA
Fn0387_Seq      CAAATATCTGTTCACCAACCGCAAAACCTTGACCGTTGGTTTGGGCGTTGCGTATGAGAA
                ************************************************************

Fn0387_exp      TGAACTCGGGAAAGTCGCCAATCCTAAGAACAAAGCCCGCGTAGCGTATACTGCGGCCGA
Fn0387_Seq      TGAACTCGGGAAAGTCGCCAATCCTAAGAACAAAGCCCGCGTAGCGTATACTGCGGCCGA
                ************************************************************

Fn0387_exp      TTGGTACAACTTACGCGGTGAAAAGGAAGATCGTCGTGGCAACATCAAAACGGATCTGAC
Fn0387_Seq      TTGGTACAACTTACGCGGTGAAAAGGAAGATCGTCGTGGCAACATCAAAACGGATCTGAC
                ************************************************************

Fn0387_exp      AATTGGTCTGGAAAATACCCGTTTTGGCGCTACTGCAAATGTGGGATACGATACCAAAGG
Fn0387_Seq      AATTGGTCTGGAAAATACCCGTTTTGGCGCTACTGCAAATGTGGGATACGATACCAAAGG
                ************************************************************

Fn0387_exp      CCATAATGTGGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAGCGGCCGC
Fn0387_Seq      CCATAATGTGGTCGACAAACCTCCCACACCTCCCCCT----------------------
                ***********************************

Fn0387_exp      ATC
Fn0387_Seq      ---
```

Fn1426 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn1426_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCACCATTCCGACCGAAACCATTTCGTACAA
Fn1426_Seq      ------------------TCGCGTGGATCCACCATTCCGACCGAAACCATTTCGTACAA
                                  ************************************

Fn1426_exp      TGGTAGCACCGTTAAAATCGGTATCCTTGACAGCGACTTTACGGACCCAGTCCGCAAAGC
Fn1426_Seq      TGGTAGCACCGTTAAAATCGGTATCCTTGACAGCGACTTTACGGACCCAGTCCGCAAAGC
                ************************************************************

Fn1426_exp      GCAATTGAGTGCACGTTACCCTGGCATTGAATTCATTCCACGCGTCAATTCGGACACAAG
Fn1426_Seq      GCAATTGAGTGCACGTTACCCTGGCATTGAATTCATTCCACGCGTCAATTCGGACACAAG
                ************************************************************

Fn1426_exp      CACCAGCTCTCACGGGGTGCAAGTACTGGAGGTAATGATGGACACCCTCGAAGATCGCAC
Fn1426_Seq      CACCAGCTCTCACGGGGTGCAAGTACTGGAGGTAATGATGGACACCCTCGAAGATCGCAC
                ************************************************************

Fn1426_exp      GAAAGGGAAGGCGAAATTCAAGGCCATTGCGGCGTCCATCGGGAATGGTGGTGCCAGCGA
Fn1426_Seq      GAAAGGGAAGGCGAAATTCAAGGCCATTGCGGCGTCCATCGGGAATGGTGGTGCCAGCGA
                ************************************************************

Fn1426_exp      AACGAACAAATCGGTGAATCCCAATGTTAAGACGTATGAGAAAGTGTTTGAGCGCTTCAA
Fn1426_Seq      AACGAACAAATCGGTGAATCCCAATGTTAAGACGTATGAGAAAGTGTTTGAGCGCTTCAA
                ************************************************************

Fn1426_exp      CTTCAACCAGAAAGTGAAAGTCGTGAATCAGTCCTTTGGCGCAGACATCACTATCGAGGA
Fn1426_Seq      CTTCAACCAGAAAGTGAAAGTCGTGAATCAGTCCTTTGGCGCAGACATCACTATCGAGGA
                ************************************************************

Fn1426_exp      AGCCCCGTATACCAAGAACAACATTCGCAACTATGTATGGGCAGGCGATTCGAAACCATT
Fn1426_Seq      AGCCCCGTATACCAAGAACAACATTCGCAACTATGTATGGGCAGGCGATTCGAAACCATT
                ************************************************************

Fn1426_exp      TGCCACCTACTTCGAAGAAAAGGTCAACAACGATGGTGGCTTATTTGTTTGGGCCGCAGG
Fn1426_Seq      TGCCACCTACTTCGAAGAAAAGGTCAACAACGATGGTGGCTTATTTGTTTGGGCCGCAGG
                ************************************************************

Fn1426_exp      AAATCGGAAAGGCGCGACAGAAACCAACCCGGGACAGGATATGGATTCAGTTGGCATGGA
Fn1426_Seq      AAATCGGAAAGGCGCGACAGAAACCAACCCGGGACAGGATATGGATTCAGTTGGCATGGA
                ************************************************************

Fn1426_exp      AGCGGGCCTTCCGTATCTGGTGAATGACCTGGAGAAAGGTTGGATTGCAGTTGTTGGCAT
Fn1426_Seq      AGCGGGCCTTCCGTATCTGGTGAATGACCTGGAGAAAGGTTGGATTGCAGTTGTTGGCAT
                ************************************************************

Fn1426_exp      TCAACCCAAAGAAACGGTCCGTGTGGGTACGGCACCTGATGGCACGCCGATTGTGAACAT
Fn1426_Seq      TCAACCCAAAGAAACGGTCCGTGTGGGTACGGCACCTGATGGCACGCCGATTGTGAACAT
                ************************************************************

Fn1426_exp      CAAACCGAATGGGAAACTCAATATTCATCGCACTGGGACTGATCGCCTGGCATATGCTGG
Fn1426_Seq      CAAACCGAATGGGAAACTCAATATTCATCGCACTGGGACTGATCGCCTGGCATATGCTGG
                ************************************************************

Fn1426_exp      CGATAACGCCAAATACTGGAGCATCAGTGCCGATGATTCAGCGATTCCGACTGCTGGTCG
Fn1426_Seq      CGATAACGCCAAATACTGGAGCATCAGTGCCGATGATTCAGCGATTCCGACTGCTGGTCG
                ************************************************************

Fn1426_exp      TGCTGGTATCGGCTCCTCATATGCCGCCCCTCGTGTGTCTCGCGCTGCTGCGTTAGTTGC
Fn1426_Seq      TGCTGGTATCGGCTCCTCATATGCCGCCCCTCGTGTGTCTCGCGCTGCTGCGTTAGTTGC
                ************************************************************

Fn1426_exp      GGAGAAATTTGACTGGATGACCGCTGATCAGGTACGGCAGACCCTGTTTACCACAACTGA
Fn1426_Seq      GGAGAAATTTGACTGGATGACCGCTGATCAGGTACGGCAGACCCTGTTTACCACAACTGA
                ************************************************************
```

```
Fn1426_exp      CGATACAGAACTGGATGCCTCTTTGGCGGGAAACGCGAATGCGGAAAAGCGCCGTCGTGT
Fn1426_Seq      CGATACAGAACTGGATGCCTCTTTGGCGGGAAACGCGAATGCGGAAAAGCGCCGTCGTGT
                ************************************************************

Fn1426_exp      GAAAACGAGTCCGGATTACAAATATGGTTGGGGCATGCTGAACCAGGAACGTGCGCTGAA
Fn1426_Seq      GAAAACGAGTCCGGATTACAAATATGGTTGGGGCATGCTGAACCAGGAACGTGCGCTGAA
                ************************************************************

Fn1426_exp      AGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATC
Fn1426_Seq      AGTCGACAAACCTCCCACACCTCCCCCT---------------------------
                ****************************
```

Fn1449 (sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn1449_exp      ----------TCCAAAATCGGATCTGGTTCCGCGTGGATCCAGCAAAGACTCGAACAAG
Fn1449_Seq      CTCCCTATAGGTCCAAAATCGGATCTGGTTCCGCGTGGATCCAGCAAAGACTCGAACAAG
                          ************************************************

Fn1449_exp      GTCAAAACCTTTGGCATGAAAGGGGAATACAAAACCGATACTGCAGGAGTGATTGACTAC
Fn1449_Seq      GTCAAAACCTTTGGCATGAAAGGGGAATACAAAACCGATACTGCAGGAGTGATTGACTAC
                ************************************************************

Fn1449_exp      AAGTATAATGCGTATGGCGTAGCGTATGTCCATGAGAATGAAGATATCAAACTGGGCAAA
Fn1449_Seq      AAGTATAATGCGTATGGCGTAGCGTATGTCCATGAGAATGAAGATATCAAACTGGGCAAA
                ************************************************************

Fn1449_exp      GGTACAGGTTGGTATACCGGTATCGTGCACAACACCTTCAAGTTCAAAGACATCGGCAAT
Fn1449_Seq      GGTACAGGTTGGTATACCGGTATCGTGCACAACACCTTCAAGTTCAAAGACATCGGCAAT
                ************************************************************

Fn1449_exp      TCCAAAGAGAAACAGCTGCAAGCGAAAGTTGGCCTGTTTAAAAGTGTCCCGTTTGACGAA
Fn1449_Seq      TCCAAAGAGAAACAGCTGCAAGCGAAAGTTGGCCTGTTTAAAAGTGTCCCGTTTGACGAA
                ************************************************************

Fn1449_exp      AATAACTCTCTGAACTGGACTATTAGTGGCGACATTTTCATTGGACACAACAAACTCGAA
Fn1449_Seq      AATAACTCTCTGAACTGGACTATTAGTGGCGACATTTTCATTGGACACAACAAACTCGAA
                ************************************************************

Fn1449_exp      CGCAAGTTTCTTGTGGTTGATGAAATCTTTCATGCCAAAAGCAAGTACTACACGTATGGT
Fn1449_Seq      CGCAAGTTTCTTGTGGTTGATGAAATCTTTCATGCCAAAAGCAAGTACTACACGTATGGT
                ************************************************************

Fn1449_exp      ATCGGCATCAAGAACGAAATTGGCAAGGAGTTCCGTTTAAGCGAAGATTTTAGCATTCGC
Fn1449_Seq      ATCGGCATCAAGAACGAAATTGGCAAGGAGTTCCGTTTAAGCGAAGATTTTAGCATTCGC
                ************************************************************

Fn1449_exp      CCATATGGTGCCTTGAAAGTGGAATATGGTCGCGTGTCGAAAATCAAAGAGAAATCTGGG
Fn1449_Seq      CCATATGGTGCCTTGAAAGTGGAATATGGTCGCGTGTCGAAAATCAAAGAGAAATCTGGG
                ************************************************************

Fn1449_exp      GAAATGAAACTGGAGGTAAAAGAAAACGATTACCTCTCCATTCGTCCGGAAATTGGTACG
Fn1449_Seq      GAAATGAAACTGGAGGTAAAAGAAAACGATTACCTCTCCATTCGTCCGGAAATTGGTACG
                ************************************************************

Fn1449_exp      GAACTGGCGTATCGGCATTACTTTGGCACCAAAACCTTGCGTACCTCAGTTGGGGTTGCA
Fn1449_Seq      GAACTGGCGTATCGGCATTACTTTGGCACCAAAACCTTGCGTACCTCAGTTGGGGTTGCA
                ************************************************************

Fn1449_exp      TACGAAAATGAACTGGGTCGTGTGGCTAATGGCAAGAACAAAGCACGCGTTGCTGGTACA
Fn1449_Seq      TACGAAAATGAACTGGGTCGTGTGGCTAATGGCAAGAACAAAGCACGCGTTGCTGGTACA
                ************************************************************

Fn1449_exp      ACTGCCGATTGGTTCAACATTCGTGGGGAGAAAGAGGATCGCAAAGGCAATGTGAAAGTG
Fn1449_Seq      ACTGCCGATTGGTTCAACATTCGTGGGGAGAAAGAGGATCGCAAAGGCAATGTGAAAGTG
                ************************************************************
```

```
Fn1449_exp      GATCTGAACGTTGGCATTGATAACCAGCGCTTAGGGGTCACGGGTAATGTCGGCTATGAC
Fn1449_Seq      GATCTGAACGTTGGCATTGATAACCAGCGCTTAGGGGTCACGGGTAATGTCGGCTATGAC
                ************************************************************

Fn1449_exp      ACGAAAGGACACAATGTAGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAA
Fn1449_Seq      ACGAAAGGACACAATGTAGTCGACAAACCTCCCACACCTCCCCCG---------------
                *****************************************

Fn1449_exp      GCGGCCGCATC
Fn1449_Seq      -----------
```

Fn1526 (sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn1526_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCAACGTGAACAAGTTTGGCATCATTTATAC
Fn1526_Seq      ---------TCGATTGGTTCCGCGTGGATCCAACGTGAACAAGTTTGGCATCATTTATAC
                         *****************************************

Fn1526_exp      AGGCGGTGAACACAAAGATAGCACGCTTGGAGTGTCCGGGTATAAATACAAATCGACCGG
Fn1526_Seq      AGGCGGTGAACACAAAGATAGCACGCTTGGAGTGTCCGGGTATAAATACAAATCGACCGG
                ************************************************************

Fn1526_exp      TGTGCTCTATCTGAACGATCGCGAAGCCTTTACGTATGGTGGCAAATACGGCTGGTCAGC
Fn1526_Seq      TGTGCTCTATCTGAACGATCGCGAAGCCTTTACGTATGGTGGCAAATACGGCTGGTCAGC
                ************************************************************

Fn1526_exp      CGGTATTGTCGGGAGCAACTTCGAGTTTAACGGTGATACCAATAAAGGGTCTAAAGAACG
Fn1526_Seq      CGGTATTGTCGGGAGCAACTTCGAGTTTAACGGTGATACCAATAAAGGGTCTAAAGAACG
                ************************************************************

Fn1526_exp      CGTTGTTAGTGGTAAACTGGGCCTGCATTACCAGGCTCCCCTGAATAAAGAAGATGACAA
Fn1526_Seq      CGTTGTTAGTGGTAAACTGGGCCTGCATTACCAGGCTCCCCTGAATAAAGAAGATGACAA
                ************************************************************

Fn1526_exp      TGCGAAACTCAAATGGCTTACTCGCGGTGAAGTTACGGTCAACAACCATCGCACTAATCG
Fn1526_Seq      TGCGAAACTCAAATGGCTTACTCGCGGTGAAGTTACGGTCAACAACCATCGCACTAATCG
                ************************************************************

Fn1526_exp      GTACTCGCAAGTAGGCAAAGATACCTATCAGAACAAAGCCTCGTTTTATTCCACGGAATT
Fn1526_Seq      GTACTCGCAAGTAGGCAAAGATACCTATCAGAACAAAGCCTCGTTTTATTCCACGGAATT
                ************************************************************

Fn1526_exp      GAGCTGGAAGAACATTATCTCCTATGACTACGACATCAATACGAACTGGATGGTTAAACC
Fn1526_Seq      GAGCTGGAAGAACATTATCTCCTATGACTACGACATCAATACGAACTGGATGGTTAAACC
                ************************************************************

Fn1526_exp      GTATACCGGGATTGACATGAGCTATGGTCACATCTTCAACATCAAAGAGAAGAACGAAGG
Fn1526_Seq      GTATACCGGGATTGACATGAGCTATGGTCACATCTTCAACATCAAAGAGAAGAACGAAGG
                ************************************************************

Fn1526_exp      CTTACCGCTGGAAGTGAAAGGCAAAGATTACTTCGTCATTACCCCGAATGTAGGCGTGGA
Fn1526_Seq      CTTACCGCTGGAAGTGAAAGGCAAAGATTACTTCGTCATTACCCCGAATGTAGGCGTGGA
                ************************************************************

Fn1526_exp      AACCAAGTATGTACTGCCATTAGGCGCAACTCACCAGGTGTTTGCGAAAGCGGATACAGA
Fn1526_Seq      AACCAAGTATGTACTGCCATTAGGCGCAACTCACCAGGTGTTTGCGAAAGCGGATACAGA
                ************************************************************

Fn1526_exp      GTTCAGCTATGATGTGGCTAAACTGTACCATGGTGTCAATCAGGCGAAAATGAAGAATGC
Fn1526_Seq      GTTCAGCTATGATGTGGCTAAACTGTACCATGGTGTCAATCAGGCGAAAATGAAGAATGC
                ************************************************************

Fn1526_exp      GAGTTCTGGCTACTACGACCTGAGTAAGCCTGAACGTCGTCGTGCTCGTGTTGCCGTGGG
Fn1526_Seq      GAGTTCTGGCTACTACGACCTGAGTAAGCCTGAACGTCGTCGTGCTCGTGTTGCCGTGGG
                ************************************************************
```

```
Fn1526_exp      AGCAGAACTGGGTTTGGAGAAAGAGAATGCGTATGGCATTACCTTTCGCGCAGAATATCA
Fn1526_Seq      AGCAGAACTGGGTTTGGAGAAAGAGAATGCGTATGGCATTACCTTTCGCGCAGAATATCA
                ************************************************************

Fn1526_exp      GGGATATAAGAAATCACAACTGAATTACGTCGACAAACCTCCCACACCTCCCCCTGAACC
Fn1526_Seq      GGGATATAAGAAATCACAACTGAATTACGTCGACAAACCTCCCACACCTCCCCCTAAC--
                ************************************************** *

Fn1526_exp      TGAAACATAAGCGGCCGCATC
Fn1526_Seq      ---------------------
```

Fn1817_1 (sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn1817_1_exp    ----------------------TCCAAAATCGGATCTGGTTCCGCGTGGATCCGTTGTTAC
Fn1817_1_Seq    TTAATACGACTCACTATAGGGTCCAAAATCGGATCTGGTTCCGCGTGGATCCGTTGTTAC
                                      ************************************

Fn1817_1_exp    GAATGACCTGAAAGTGGTAACAGGGAGTAATAGCACCACTTCTACCAACAACATTGCCAT
Fn1817_1_Seq    GAATGACCTGAAAGTGGTAACAGGGAGTAATAGCACCACTTCTACCAACAACATTGCCAT
                ************************************************************

Fn1817_1_exp    TGATGCCAAAGAGTTAGGTGGCATGTATGCGAATCGTATTCGCATCATTAGCACGGATAA
Fn1817_1_Seq    TGATGCCAAAGAGTTAGGTGGCATGTATGCGAATCGTATTCGCATCATTAGCACGGATAA
                ************************************************************

Fn1817_1_exp    AGGTGCTGGAGTCAATTCAGACGCGTTTATCGTGTCGAAGAACTCCAAACTGGAAATTAC
Fn1817_1_Seq    AGGTGCTGGAGTCAATTCAGACGCGTTTATCGTGTCGAAGAACTCCAAACTGGAAATTAC
                ************************************************************

Fn1817_1_exp    CGCAGATGGCAAAATCAAGGTGAACAAAGTCGACAAACCTCCCACACCTCCCCCTGAACC
Fn1817_1_Seq    CGCAGATGGCAAAATCAAGGTGAACAAAGTCGACAAACCTCCCACACCTCCCCCGAAGT-
                ********************************************************  *

Fn1817_1_exp    TGAAACATAAGCGGCCGCATC
Fn1817_1_Seq    ---------------------
```

Fn1817_2 (sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn1817_2_exp    ----------------------TCCAAAATCGGATCTGGTTCCGCGTGGATCCGCCGCAGC
Fn1817_2_Seq    TTAATACGACTCACTATAGGGTCCAAAATCGGATCTGGTTCCGCGTGGATCCGCCGCAGC
                                      ************************************

Fn1817_2_exp    TGGTGATCTGACCTTAACTGCGACCAATAAGGTCGATAACAAAAGCGGGAAAACGATCTT
Fn1817_2_Seq    TGGTGATCTGACCTTAACTGCGACCAATAAGGTCGATAACAAAAGCGGGAAAACGATCTT
                ************************************************************

Fn1817_2_exp    TGCAGGCAATAAACTGACGGTAACAGCGAAAGAGATCAAGAACAACAAACGTGCTGAACT
Fn1817_2_Seq    TGCAGGCAATAAACTGACGGTAACAGCGAAAGAGATCAAGAACAACAAACGTGCTGAACT
                ************************************************************

Fn1817_2_exp    CTTGGGTACGAACATTGAACTGACTGCCGATAAAGTGCGCAATGAAGTTGGCACCATTAA
Fn1817_2_Seq    CTTGGGTACGAACATTGAACTGACTGCCGATAAAGTGCGCAATGAAGTTGGCACCATTAA
                ************************************************************

Fn1817_2_exp    AGCGTTCAATGACATCACCATTAAAGTCGACAAACCTCCCACACCTCCCCCTGAACCTGA
Fn1817_2_Seq    AGCGTTCAATGACATCACCATTAAAGTCGACAAACCTCCCACACCTCCCCCTAAC-----
                ************************************************** *

Fn1817_2_exp    AACATAAGCGGCCGCATC
Fn1817_2_Seq    ------------------
```

Fn1859 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn1859_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCCCGACACCAGCGCCTAAGAAAGTCGTCGA
Fn1859_Seq      ----------------------------CCCGACACCAGCGCCTAAGAAAGTCGTCGA
                                            ****************************

Fn1859_exp      ATACGTGGAGAAACCGGTGATTGTCTATCGTGATCGGGAAGTAGCGCCGGCTTGGCGTCC
Fn1859_Seq      ATACGTGGAGAAACCGGTGATTGTCTATCGTGATCGGGAAGTAGCGCCGGCTTGGCGTCC
                ************************************************************

Fn1859_exp      CAATGGTAGCGTTGATGTTCAGTATCGCTGGTATGGCAATGTGGAGAATCGCACCCCGAA
Fn1859_Seq      CAATGGTAGCGTTGATGTTCAGTATCGCTGGTATGGCAATGTGGAGAATCGCACCCCGAA
                ************************************************************

Fn1859_exp      GAAAGAAGATCCAGCCTCTCCGTGGCTTGGTGATAATGTGAATGCGGGGCGCTTACAGAC
Fn1859_Seq      GAAAGAAGATCCAGCCTCTCCGTGGCTTGGTGATAATGTGAATGCGGGGCGCTTACAGAC
                ************************************************************

Fn1859_exp      GCTCACCAAAGTGAACTTTACCGAGAAACAAACGCTTGAAATTCGCACACGGAATTATCA
Fn1859_Seq      GCTCACCAAAGTGAACTTTACCGAGAAACAAACGCTTGAAATTCGCACACGGAATTATCA
                ************************************************************

Fn1859_exp      TACTCTGATGAACCCCAAGGATTCACAAGCTGCTGATGACCAAGTCCGTGTTCGCCACTT
Fn1859_Seq      TACTCTGATGAACCCCAAGGATTCACAAGCTGCTGATGACCAAGTCCGTGTTCGCCACTT
                ************************************************************

Fn1859_exp      CTACAAATTCGGGAAACTGGGCAGTTCGAAAATCGATGTAACCAGTCGCTTGGAGTATAA
Fn1859_Seq      CTACAAATTCGGGAAACTGGGCAGTTCGAAAATCGATGTAACCAGTCGCTTGGAGTATAA
                ************************************************************

Fn1859_exp      AAAAAACAATGGAGATGCCGGACGCAAACAGGCTGAAGCGTCAGTACTGTTCGATTTTGC
Fn1859_Seq      AAAAAACAATGGAGATGCCGGACGCAAACAGGCTGAAGCGTCAGTACTGTTCGATTTTGC
                ************************************************************

Fn1859_exp      GGACTACATCTATTCTAACAACTTCTTCAAAGCCGACAAATTCGGCTTTCGTCTGGGGTA
Fn1859_Seq      GGACTACATCTATTCTAACAACTTCTTCAAAGCCGACAAATTCGGCTTTCGTCTGGGGTA
                ************************************************************

Fn1859_exp      TCAGCACAAATGGGCGGGTCATAACTCGGGTGTTGTGGGCCAGCCGTTTAACAAAGGTAC
Fn1859_Seq      TCAGCACAAATGGGCGGGTCATAACTCGGGTGTTGTGGGCCAGCCGTTTAACAAAGGTAC
                ************************************************************

Fn1859_exp      TCAGGATAACTACTTTATCAATTTCGAAAGTGAATACACGTTACCTTGGGGCTTTTCGGC
Fn1859_Seq      TCAGGATAACTACTTTATCAATTTCGAAAGTGAATACACGTTACCTTGGGGCTTTTCGGC
                ************************************************************

Fn1859_exp      CGAACTGAACGCCTACAACTATTACAATGTTCACAACAAGAAATTTGCCACCTATAACAA
Fn1859_Seq      CGAACTGAACGCCTACAACTATTACAATGTTCACAACAAGAAATTTGCCACCTATAACAA
                ************************************************************

Fn1859_exp      AGGCAACAAGAAAAGCCAGTTCTATGGCGAAATTGAGGCCTATTTGTACCAGCATACCCC
Fn1859_Seq      AGGCAACAAGAAAAGCCAGTTCTATGGCGAAATTGAGGCCTATTTGTACCAGCATACCCC
                ************************************************************

Fn1859_exp      ACTCTATAAAACAAACAATGTGGAACTGTCCTTTGACTTTGAAGGTGGCTATGATCCGTA
Fn1859_Seq      ACTCTATAAAACAAACAATGTGGAACTGTCCTTTGACTTTGAAGGTGGCTATGATCCGTA
                ************************************************************

Fn1859_exp      TACGTGGCATCAGTACAAAGTCGTTTCCGCAAAAGACAGCAATAAATACGAAGTGTACAT
Fn1859_Seq      TACGTGGCATCAGTACAAAGTCGTTTCCGCAAAAGACAGCAATAAATACGAAGTGTACAT
                ************************************************************

Fn1859_exp      GCTGCCTACGCTGCAAGTTAGCTACAAACCGACCGACTTTGTGAAACTGTATGCAGCGGC
Fn1859_Seq      GCTGCCTACGCTGCAAGTTAGCTACAAACCGACCGACTTTGTGAAACTGTATGCAGCGGC
                ************************************************************
```

```
Fn1859_exp      AGGCGCGGAATATCGCAATTGGGCAGTAACCGCAGAGAGCAAAGCGAAGAACTGGCGTTG
Fn1859_Seq      AGGCGCGGAATATCGCAATTGGGCAGTAACCGCAGAGAGCAAAGCGAAGAACTGGCGTTG
                ************************************************************

Fn1859_exp      GCAACCGACTGCATGGGCGGGTATGAAGGTGACCTTTGTCGACAAACCTCCCACACCTCC
Fn1859_Seq      GCAACCGACTGCATGGGCGGGTATGAAGGTGACCTTTGTCGACAAACCTCCCACACCTCC
                ************************************************************

Fn1859_exp      CCCTGAACCTGAAACATAAGCGGCCGCATC
Fn1859_Seq      CCT----------------------------
                **
```

Fn1893 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Fn1893_exp      --------TCCAAAATCGGATCTGGTTCCGCGTGGATCCAGCAAAGATAGCAACAAAATC
Fn1893_Seq      CCCTTAGGTCCAAAATCGGATCTGGTTCCGCGTGGATCCAGCAAAGATAGCAACAAAATC
                        **************************************************

Fn1893_exp      AAGATTTTCGGTATTAAGGGCGAATACAAAACCGATACTGCAGGCGTGATTGACTATAAG
Fn1893_Seq      AAGATTTTCGGTATTAAGGGCGAATACAAAACCGATACTGCAGGCGTGATTGACTATAAG
                ************************************************************

Fn1893_exp      AATGAAGCCTATGGAATGGCGTATGTGCACGAAAATGAGGACATTAAACTGGGTAAAGGG
Fn1893_Seq      AATGAAGCCTATGGAATGGCGTATGTGCACGAAAATGAGGACATTAAACTGGGTAAAGGG
                ************************************************************

Fn1893_exp      ATTGGCTGGTATACTGGGATTGTGGATAACACCTTTAAATTCAAAGATATCGGTAAATCC
Fn1893_Seq      ATTGGCTGGTATACTGGGATTGTGGATAACACCTTTAAATTCAAAGATATCGGTAAATCC
                ************************************************************

Fn1893_exp      AAAGAGGAACAGATTCAGGCCAAAGTCGGTCTGCTGAAATCAATCCCGTTTGATGACAAC
Fn1893_Seq      AAAGAGGAACAGATTCAGGCCAAAGTCGGTCTGCTGAAATCAATCCCGTTTGATGACAAC
                ************************************************************

Fn1893_exp      AATTCGCTCAATTGGACCATTTCTGGGGACATTTTCGTCGGCTATAACAAAATGCATCGC
Fn1893_Seq      AATTCGCTCAATTGGACCATTTCTGGGGACATTTTCGTCGGCTATAACAAAATGCATCGC
                ************************************************************

Fn1893_exp      AAGTACCTGGTTGTGAACGAAATCTTTAATGCGAAATCGAAGTATTACACGTATGGCATT
Fn1893_Seq      AAGTACCTGGTTGTGAACGAAATCTTTAATGCGAAATCGAAGTATTACACGTATGGCATT
                ************************************************************

Fn1893_exp      GGCATCAAGAATAAAATCAGCAAAGACTTTCGCTTGTCCGAAGATTTTAGTCTGGTACCA
Fn1893_Seq      GGCATCAAGAATAAAATCAGCAAAGACTTTCGCTTGTCCGAAGATTTTAGTCTGGTACCA
                ************************************************************

Fn1893_exp      TACGGCTCTCTGAACTTAGAGTATGGCCGCGTGAACAAAATTAAGGAGAAAGTTGGTGAA
Fn1893_Seq      TACGGCTCTCTGAACTTAGAGTATGGCCGCGTGAACAAAATTAAGGAGAAAGTTGGTGAA
                ************************************************************

Fn1893_exp      ATCCGGTTGGAAGTCAAAGAAAACTACTACGTAAGCGTGAATCCGGAAATCGGAGCGGAA
Fn1893_Seq      ATCCGGTTGGAAGTCAAAGAAAACTACTACGTAAGCGTGAATCCGGAAATCGGAGCGGAA
                ************************************************************

Fn1893_exp      TTAACCTACAAACACCTCCTTGCCAGTCGCAAAACGTTCCGTATGGGTCTGGGAATTGCC
Fn1893_Seq      TTAACCTACAAACACCTCCTTGCCAGTCGCAAAACGTTCCGTATGGGTCTGGGAATTGCC
                ************************************************************

Fn1893_exp      TACGAAAACGAGCTTGGTAAAGTTGCGAATGGCAAGAACAAAGCTCGTGTTGCGTATACC
Fn1893_Seq      TACGAAAACGAGCTTGGTAAAGTTGCGAATGGCAAGAACAAAGCTCGTGTTGCGTATACC
                ************************************************************

Fn1893_exp      AATGCTGACTGGTTTAACATTCGTGGTGAGAAAGAAGATCGCAAAGGCAACATCAAATTC
Fn1893_Seq      AATGCTGACTGGTTTAACATTCGTGGTGAGAAAGAAGATCGCAAAGGCAACATCAAATTC
                ************************************************************
```

```
Fn1893_exp      GATCTGAACATCGGTCTGGATAACCAACGTGTAGGTGTGACAGCAAATGCAGGCTATGAT
Fn1893_Seq      GATCTGAACATCGGTCTGGATAACCAACGTGTAGGTGTGACAGCAAATGCAGGCTATGAT
                ************************************************************

Fn1893_exp      ACGAAAGGGCATAATGTCGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAA
Fn1893_Seq      ACGAAAGGGCATAATGTCGTCGACAAACCTCCCACACCTCCCCCTAA--------------
                ******************************************* *

Fn1893_exp      GCGGCCGCATC
Fn1893_Seq      -----------
```

Gallo0112A (forward sequencing of miniPrep DNA via primer pGEXfor, reverse sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo0112A_exp      ATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA
Gallo0112A_Seq      ----------------------------------------GCGACATCCTCCAAA
                                                           *     *  * ***

Gallo0112A_exp      TCGGATCTGGTTCCGCGTGGATCCGATGAAGCTGTGTCTAGTCCGATGGAACTCGTAACG
Gallo0112A_Seq      TCGGATCTGGTTCCGCGTGGATCCGATGAAGCTGTGTCTAGTCCGATGGAACTCGTAACG
                    ************************************************************

Gallo0112A_exp      GTCGCGGATGTTAACGCGGATAACAATACCAATACCGATCAGAAAACCGAGGTGGACCAG
Gallo0112A_Seq      GTCGCGGATGTTAACGCGGATAACAATACCAATACCGATCAGAAAACCGAGGTGGACCAG
                    ************************************************************

Gallo0112A_exp      GAAACGCAGCAGCCTGCGAACCAAGTTGAAACGCCGGAGAACCAGACCCCGATTGAGCAA
Gallo0112A_Seq      GAAACGCAGCAGCCTGCGAACCAAGTTGAAACGCCGGAGAACCAGACCCCGATTGAGCAA
                    ************************************************************

Gallo0112A_exp      GGTGTCGTTGGCGAGCAGAATCAGAAAGTGACGGAGGAAAACCAGGTGACGGAGAACCAA
Gallo0112A_Seq      GGTGTCGTTGGCGAGCAGAATCAGAAAGTGACGGAGGAAAACCAGGTGACGGAGAACCAA
                    ************************************************************

Gallo0112A_exp      GATGTAACCCAACAGAATCAGGTCACCGAAAATCAAGAACCTGCGACCAAAACCCAGGAT
Gallo0112A_Seq      GATGTAACCCAACAGAATCAGGTCACCGAAAATCAAGAACCTGCGACCAAAACCCAGGAT
                    ************************************************************

Gallo0112A_exp      GATGCCCAGAAAACGGAAACGACAGATGCGGAAGAGAAAGTCGAAGTAACGGATAGCCTG
Gallo0112A_Seq      GATGCCCAGAAAACGGAAACGACAGATGCGGAAGAGAAAGTCGAAGTAACGGATAGCCTG
                    ************************************************************

Gallo0112A_exp      AAACAGAAAGCTGATCAGCCAAACGAATCGACCGAGAAAGCGCGTAAGGCTCTGTCAACG
Gallo0112A_Seq      AAACAGAAAGCTGATCAGCCAAACGAATCGACCGAGAAAGCGCGTAAGGCTCTGTCAACG
                    ************************************************************

Gallo0112A_exp      AATCTGACGACGAAGAAGGAATCGTCCTATAATACGAACCTGCAGGGGTTGTCGTATGAC
Gallo0112A_Seq      AATCTGACGACGAAGAAGGAATCGTCCTATAATACGAACCTGCAGGGGTTGTCGTATGAC
                    ************************************************************

Gallo0112A_exp      GCCAACGTTTGGGAGGTCCGGGAGGATGGCCTGTATAGCAATGCGATTGGCGAAGGTGAT
Gallo0112A_Seq      GCCAACGTTTGGGAGGTCCGGGAGGATGGCCTGTATAGCAATGCGATTGGCGAAGGTGAT
                    ************************************************************

Gallo0112A_exp      TCCTTTCTGCTCTCGACATCCGCAGGTAAGAA--TTTTGTGTTCCAGACAGATGTGACGT
Gallo0112A_Seq      TCCTTTCTGCTCTCGACATCCGCAGGTAAAGAAATTTTGTGTTCCAGACAGATGTGACGT
                    ****************************  *  **************************

Gallo0112A_exp      TTCTGCAAAACACCGGTGCAGCCTCACTGGTATTTCGCAGCACTGGGGACGCACAGAACC
Gallo0112A_Seq      TTCTGCAAAACACCGGTGCAC-------------------------------------
                    ********************

Gallo0112A_exp    TTAAAGGATACGTGGTGAACCTGGACGGCAATAGCCACAAAATCAAATTCATGCGTTGGGGC
Gallo0112A_Seq    ------------------------------------------------------------
```

```
Gallo0112A_exp      GAGGCTAACCTGATTGACGAAAAGGAAATTGAAGCAACTAGCGATAACAAATACAGTCTG
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      AAAGTTGTTGCAGCGAATGGTTGGATCTCCTATTACATTAACGGGATTTTGGTGGCAAAT
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      CTGTCGGATTATACCATCCAACGCGATGATCGTGGCCAAACGACCTATATTAAGGATGGC
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      AATTTCAGCCTTCTGAACTGGAACGGGGAAATGATTTTCCAGAACACCTTCTATCGCGAA
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      CTTACTGACGCTGAACTGCCAATTCTGAAAGATGTAACCGTGTCGTCAAAGAATGGTCCA
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      GTCGAACCCAAAGGGCAATTCTTTCCGGAGGGTGCGGTTTATATCCAGTATGTCAGTCAC
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      GATGCCTCTACTGTGGACTTGTCGTTCGTTCCGAATAATCAAGACGCAGTCATCAAAGTG
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      ACCGATGACCAAGGCAACGTTTATAGCGACCCGAGCAACATTCCCGTAAGCGTTGGTGCC
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      AATTACCTGACCGTGACCTCTACCTACACAGTTGATGGCTACGAAGTGACATCCACCTAT
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      CGCATCAATGTTCATCGCCGTCAAAGCGCCGAAGTCTACTATAACGAGAACTTTCGCGAT
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      CAGTACCACTATAGTGTGAAAGACGGTTGGGCCAATGATCCGAACGGTTTAGTGTACTAC
Gallo0112A_Seq      ------------------------------------------------------------

Gallo0112A_exp      AATGGCGTATATCACATGTTTTATCAGTTCTATGATGACATTCAATGGGGTCCGATGCAT
Gallo0112A_Seq      ----------------TGTTTTATCAGTTCTATGATGACATTCAATGGGGTCCGATGCAT
                                    ****************************************

Gallo0112A_exp      TGGGCACATGCGACTTCCACTGATCTGATCCATTGGGAGGATCAACCCATCGCGTTTTAT
Gallo0112A_Seq      TGGGCACATGCGACTTCCACTGATCTGATCCATTGGGAGGATCAACCCATCGCGTTTTAT
                    ************************************************************

Gallo0112A_exp      CCGGATTACAACGGAGCCATGTTTAGTGGTTGCATTGTGGCGGACCCCAATAACACCTCT
Gallo0112A_Seq      CCGGATTACAACGGAGCCATGTTTAGTGGTTGCATTGTGGCGGACCCCAATAACACCTCT
                    ************************************************************

Gallo0112A_exp      GGACTGTTCGAAGGCGACAAAGGCGGTCTGGTGGCACTGATTACGGCCGATGGCGAGGGT
Gallo0112A_Seq      GGACTGTTCGAAGGCGACAAAGGCGGTCTGGTGGCACTGATTACGGCCGATGGCGAGGGT
                    ************************************************************

Gallo0112A_exp      CAGCGGATCAAAGTTGCGTACTCTAAGGATGAAGGCAAAACATGGCAGAAATTAGACGAA
Gallo0112A_Seq      CAGCGGATCAAAGTTGCGTACTCTAAGGATGAAGGCAAAACATGGCAGAAATTAGACGAA
                    ************************************************************

Gallo0112A_exp      GTCGCGGCGGACTGGTCTACCGATCCGCTGCAGAATCGCGACTTTCGTGATCCTAAAGTG
Gallo0112A_Seq      GTCGCGGCGGACTGGTCTACCGATCCGCTGCAGAATCGCGACTTTCGTGATCCTAAAGTG
                    ************************************************************
```

```
Gallo0112A_exp      TTTCGCTGGGAAGGCAAATGGTTCATGGTCCTTGCCGGAGGACCACTGCGCATCTATAGC
Gallo0112A_Seq      TTTCGCTGGGAAGGCAAATGGTTCATGGTCCTTGCCGGAGGACCACTGCGCATCTATAGC
                    ************************************************************

Gallo0112A_exp      TCCGACAACTTACTTGATTGGTCAGTTGAAAGCACCTATCCTGACCTGCATACTGAATGT
Gallo0112A_Seq      TCCGACAACTTACTTGATTGGTCAGTTGAAAGCACCTATCCTGACCTGCATACTGAATGT
                    ************************************************************

Gallo0112A_exp      CCGGACTTGTATCCGATTATGGCCGAAGGAAACACCGTTAAATGGGTCTTGAGTCGTGGT
Gallo0112A_Seq      CCGGACTTGTATCCGATTATGGCCGAAGGAAACACCGTTAAATGGGTCTTGAGTCGTGGT
                    ************************************************************

Gallo0112A_exp      GGGCGTTATTACAAGGTAGGTGACCTGAAACAGGTGGATGGCCATTGGAAATTCGTTGCA
Gallo0112A_Seq      GGGCGTTATTACAAGGTAGGTGACCTGAAACAGGTGGATGGCCATTGGAAATTCGTTGCA
                    ************************************************************


Gallo0112A_exp      GATGCCGATTACCAGGAATCAGATGGCATCATGAATTTTGGCAAAGATAGTTATGCCGCT
Gallo0112A_Seq      GATGCCGATTACCAGGAATCAGATGGCATCATGAATTTTGGCAAAGATAGTTATGCCGCT
                    ************************************************************

Gallo0112A_exp      ATGACTTACTATGTGCAAGATTTTGGTACAAAAGACAACCCGACCATTCCGCAGATTATT
Gallo0112A_Seq      ATGACTTACTATGTGCAAGATTTTGGTACAAAAGACAACCCGACCATTCCGCAGATTATT
                    ************************************************************

Gallo0112A_exp      GAACTCAACTGGATGAACACTTGGGATAACTACTGCAATCTCGTAGCTGAACGCACAGGT
Gallo0112A_Seq      GAACTCAACTGGATGAACACTTGGGATAACTACTGCAATCTCGTAGCTGAACGCACAGGT
                    ************************************************************

Gallo0112A_exp      CAGAAATTCAATGGGACCTTTAATCTCAACTTGACGCTGGGCTTAGTGAAAGATGGCGAC
Gallo0112A_Seq      CAGAAATTCAATGGGACCTTTAATCTCAACTTGACGCTGGGCTTAGTGAAAGATGGCGAC
                    ************************************************************

Gallo0112A_exp      AAATATGTGTTAACCCAGACTCCAATCAAGGCGTACGAAAGCTTACGTGATGTAGACCAC
Gallo0112A_Seq      AAATATGTGTTAACCCAGACTCCAATCAAGGCGTACGAAAGCTTACGTGATGTAGACCAC
                    ************************************************************

Gallo0112A_exp      AAGGTTGAATACAAAGACGTCGTGGTCGGCAAAGATAACAATCTGTTTAAAGACTTTTCT
Gallo0112A_Seq      AAGGTTGAATACAAAGACGTCGTGGTCGGCAAAGATAACAATCTGTTTAAAGACTTTTCT
                    ************************************************************

Gallo0112A_exp      GGGGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATC
Gallo0112A_Seq      GGGGTCGACAAACCTCCCACACCTCCCCCT-------------------------
                    ****************************
```

Gallo0112B (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo0112B_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCATCAAAGCGTATGAATCGCTTCGCGATGT
Gallo0112B_Seq      ----------------GTTCGCGTGGATCCATCAAAGCGTATGAATCGCTTCGCGATGT
                                    * ***********************************

Gallo0112B_exp      GGATCACAAAGTTGAGTACAAAGACGTCGTAGTGGGCAAAGACAACAACTTGTTCAAAGA
Gallo0112B_Seq      GGATCACAAAGTTGAGTACAAAGACGTCGTAGTGGGCAAAGACAACAACTTGTTCAAAGA
                    ************************************************************

Gallo0112B_exp      CTTTAGTGGTGATACCTATGAAATCGTCGCGCATTTCAAGCCGTCCGATCGTACGACGAA
Gallo0112B_Seq      CTTTAGTGGTGATACCTATGAAATCGTCGCGCATTTCAAGCCGTCCGATCGTACGACGAA
                    ************************************************************

Gallo0112B_exp      AGTGGGTTTTAACCTCCGCGTTGGGCAAGGCGAAGTGACAAAAGTCTACTACGACCTTCA
Gallo0112B_Seq      AGTGGGTTTTAACCTCCGCGTTGGGCAAGGCGAAGTGACAAAAGTCTACTACGACCTTCA
                    ************************************************************

Gallo0112B_exp      GACCGGTCGCATTGCTATCGATCGCAGCCAATCAGGCATTATTCTGACCGAACTCTTTCG
Gallo0112B_Seq      GACCGGTCGCATTGCTATCGATCGCAGCCAATCAGGCATTATTCTGACCGAACTCTTTCG
                    ************************************************************
```

```
Gallo0112B_exp      CAACGTCGATTCTCAAGCCGTGACGCGCAATGCGGACGGTTCCATTGATCTGCACATCTT
Gallo0112B_Seq      CAACGTCGATTCTCAAGCCGTGACGCGCAATGCGGACGGTTCCATTGATCTGCACATCTT
                    ************************************************************

Gallo0112B_exp      TGTAGATCGTGCGAGCGTTGAAGTGTTCACCAAGGGCGGTACAGTGACGGGTGCCAACCA
Gallo0112B_Seq      TGTAGATCGTGCGAGCGTTGAAGTGTTCACCAAGGGCGGTACAGTGACGGGTGCCAACCA
                    ************************************************************

Gallo0112B_exp      GATTTTCACGAGCCCGCAATCTCTCGGCTTAGGCGTGTTTGCGGAAGGGTATGAAGCAAA
Gallo0112B_Seq      GATTTTCACGAGCCCGCAATCTCTCGGCTTAGGCGTGTTTGCGGAAGGGTATGAAGCAAA
                    ************************************************************

Gallo0112B_exp      AGCTGATATCGCCCTTTATCCGTTAAAATCTATTTGGAAAGATAAAGTTGAAACCACCAA
Gallo0112B_Seq      AGCTGATATCGCCCTTTATCCGTTAAAATCTATTTGGAAAGATAAAGTTGAAACCACCAA
                    ************************************************************

Gallo0112B_exp      ACCCCAGAGTATTGTTCCCGCAAGCGCGAAGAACGTTCGCATGAACGTTGGCGATTCGAC
Gallo0112B_Seq      ACCCCAGAGTATTGTTCCCGCAAGCGCGAAGAACGTTCGCATGAACGTTGGCGATTCGAC
                    ************************************************************

Gallo0112B_exp      CGTAGTAAAAGCGTATGTTTCGCCTGCCGTTGTTAATCAGGATCTGTTGTGGAGCATCCT
Gallo0112B_Seq      CGTAGTAAAAGCGTATGTTTCGCCTGCCGTTGTTAATCAGGATCTGTTGTGGAGCATCCT
                    ************************************************************

Gallo0112B_exp      GAACAATGGGAATGTTAGCACGGAAATTAGCGGTAATCAAGTCTTTGTGAAAGCCCTGAA
Gallo0112B_Seq      GAACAATGGGAATGTTAGCACGGAAATTAGCGGTAATCAAGTCTTTGTGAAAGCCCTGAA
                    ************************************************************

Gallo0112B_exp      GAAGGGTCAGGTCATTGTCCGGGCACAGTCCAAAACAGACCCGTCAGTCTATCAGGACTT
Gallo0112B_Seq      GAAGGGTCAGGTCATTGTCCGGGCACAGTCCAAAACAGACCCGTCAGTCTATCAGGACTT
                    ************************************************************

Gallo0112B_exp      CGTCCTGGATATTCTGGAGGACAATTTTAAAACCAACGTGAAGAACGTAAAAGTGTTTGC
Gallo0112B_Seq      CGTCCTGGATATTCTGGAGGACAATTTTAAAACCAACGTGAAGAACGTAAAAGTGTTTGC
                    ************************************************************

Gallo0112B_exp      TGGGGACTGGCATGCCGATGGTGAATCGCTGAAAGTGGAAAATCACAACAGTAATGACAT
Gallo0112B_Seq      TGGGGACTGGCATGCCGATGGTGAATCGCTGAAAGTGGAAAATCACAACAGTAATGACAT
                    ************************************************************

Gallo0112B_exp      CTATATGGCAGCTGATAAAATGCCGTACGAGAATTACCAGATGGATCTGGATATCAAATA
Gallo0112B_Seq      CTATATGGCAGCTGATAAAATGCCGTACGAGAATTACCAGATGGATCTGGATATCAAATA
                    ************************************************************

Gallo0112B_exp      TGGCCGTGGAGTCGTTAACATTTTCTTTGCTAGTGGCAACCCAGATGCGAACAATGCGTA
Gallo0112B_Seq      TGGCCGTGGAGTCGTTAACATTTTCTTTGCTAGTGGCAACCCAGATGCGAACAATGCGTA
                    ************************************************************

Gallo0112B_exp      CTCAATCCAGTTTGGAGGAGATAATTCGGTGCGTCTGTTTCGGTTTTATAGCGACACCAT
Gallo0112B_Seq      CTCAATCCAGTTTGGAGGAGATAATTCGGTGCGTCTGTTTCGGTTTTATAGCGACACCAT
                    ************************************************************

Gallo0112B_exp      TTCCGAATCTCAAATGACGGCCGCAATCAACGATAACCAATTTCATCATGTGCGTCTGGT
Gallo0112B_Seq      TTCCGAATCTCAAATGACGGCCGCAATCAACGATAACCAATTTCATCATGTGCGTCTGGT
                    ************************************************************

Gallo0112B_exp      AAAGAGCGCCAATGCCATCCAGGTTTTCGTAGACAATCAGCTGGCCATGTCATATACCTT
Gallo0112B_Seq      AAAGAGCGCCAATGCCATCCAGGTTTTCGTAGACAATCAGCTGGCCATGTCATATACCTT
                    ************************************************************

Gallo0112B_exp      TGATCAGGTGGAAGATTTCTTCAACAATCCGTACATTGGCTTAGGCTTATGGGACGGCGA
Gallo0112B_Seq      TGATCAGGTGGAAGATTTCTTCAACAATCCGTACATTGGCTTAGGCTTATGGGACGGCGA
                    ************************************************************

Gallo0112B_exp      ATTGGAGGTGCAGAATTTCTTCGTGGTAGACCTGGACGCAAAGGAACCGACCCAGAACGA
Gallo0112B_Seq      ATTGGAGGTGCAGAATTTCTTCGTGGTAGACCTGGACGCAAAGGAACCGACCCAGAACGA
                    ************************************************************
```

```
Gallo0112B_exp      AGAGAAAGTGGAAGTCGTTCCGACCGATCCTCAGACTCCGGCTGAACAGGTCGTGACGAC
Gallo0112B_Seq      AGAGAAAGTGGAAGTCGTTCCGACCGATCCTCAGACTCCGGCTGAACAGGTCGTGACGAC
                    ************************************************************

Gallo0112B_exp      CACAACTCTGGCGGCGAAAGCGCCAGCAAAATCTGAGAAAGCGACCGATGCGAAAGCCCC
Gallo0112B_Seq      CACAACTCTGGCGGCGAAAGCGCCAGCAAAATCTGAGAAAGCGACCGATGCGAAAGCCCC
                    ************************************************************

Gallo0112B_exp      AGTAATTCCGAAAACTGCACTGGTGAGTGAGACTGTGTTGCCTCAAACTGGTGAGAAAGA
Gallo0112B_Seq      AGTAATTCCGAAAACTGCACTGGTGAGTGAGACTGTGTTGCCTCAAACTGGTGAGAAAGA
                    ************************************************************

Gallo0112B_exp      TTCCCATGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATC
Gallo0112B_Seq      TTCCCATGTCGACAAACCTCCCACACCTCCCCCTA-----------------------
                    *********************************
```

Gallo0272 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo0272_exp       TCCAAAATCGGATCTGGTTCCGCGTGGATCCGAAGAGATTATCAATGCGCAGAATAGCGT
Gallo0272_Seq       ----------------GGGTCGCGTGGATCCGAAGAGATTATCAATGCGCAGAATAGCGT
                                    *   ************************************

Gallo0272_exp       GAATAAACAACTGCAAGACCTCATGGCCTCTCTGAATGCGGTTACGCAGACGATTACCGG
Gallo0272_Seq       GAATAAACAACTGCAAGACCTCATGGCCTCTCTGAATGCGGTTACGCAGACGATTACCGG
                    ************************************************************

Gallo0272_exp       CAATAAAGTGACTGTATCTAGTATTGAAGAAGCCAACAAGAAACTGGCCGAAATTAAAGC
Gallo0272_Seq       CAATAAAGTGACTGTATCTAGTATTGAAGAAGCCAACAAGAAACTGGCCGAAATTAAAGC
                    ************************************************************

Gallo0272_exp       AAAGATTCAGGCTGTGGACAAGTTAAACGCACAGCTGAAAGCAGAGTATGACGCTGAAGT
Gallo0272_Seq       AAAGATTCAGGCTGTGGACAAGTTAAACGCACAGCTGAAAGCAGAGTATGACGCTGAAGT
                    ************************************************************

Gallo0272_exp       CCAACGCGTTAATGAGCATAACGCACAACTGAAAGCGGACTACGAAAAGAAACTTGCACA
Gallo0272_Seq       CCAACGCGTTAATGAGCATAACGCACAACTGAAAGCGGACTACGAAAAGAAACTTGCACA
                    ************************************************************

Gallo0272_exp       GTATGAAGCCGATAAAGCCGAATACGATAAGAAACTCGCCGAATACGAAGCCAACAAAGG
Gallo0272_Seq       GTATGAAGCCGATAAAGCCGAATACGATAAGAAACTCGCCGAATACGAAGCCAACAAAGG
                    ************************************************************

Gallo0272_exp       GAAAGATGGTTATCTGAACCAAACGTACGTCCAAGGACTGATCTTTAAATCAGAAGCGGA
Gallo0272_Seq       GAAAGATGGTTATCTGAACCAAACGTACGTCCAAGGACTGATCTTTAAATCAGAAGCGGA
                    ************************************************************

Gallo0272_exp       TGCTCATGTTACCATTGTCAAGTCGGATGGTGCTATCATCGTCAATGACAGTACAGACTC
Gallo0272_Seq       TGCTCATGTTACCATTGTCAAGTCGGATGGTGCTATCATCGTCAATGACAGTACAGACTC
                    ************************************************************

Gallo0272_exp       ACATCGTGTAGTGCTGCATCAGGGCGAATCCGTAACGGTGACCTACACCAATCTGAAGAA
Gallo0272_Seq       ACATCGTGTAGTGCTGCATCAGGGCGAATCCGTAACGGTGACCTACACCAATCTGAAGAA
                    ************************************************************

Gallo0272_exp       TTCGTACTACAACGGTGTGAAAATCGACAAAGTGGTTTACGTATACACGGCTAAGGATGC
Gallo0272_Seq       TTCGTACTACAACGGTGTGAAAATCGACAAAGTGGTTTACGTATACACGGCTAAGGATGC
                    ************************************************************

Gallo0272_exp       GGTCAATGGTTTGCACATCTCTAACAACCCGAACATCACCGTGACCTTTATGAGTTCGGA
Gallo0272_Seq       GGTCAATGGTTTGCACATCTCTAACAACCCGAACATCACCGTGACCTTTATGAGTTCGGA
                    ************************************************************

Gallo0272_exp       CTTCGATACAGATGATAAGAACGGTGAGCAAAATGGGTCACAATCCAGCCATATTGGCAT
Gallo0272_Seq       CTTCGATACAGATGATAAGAACGGTGAGCAAAATGGGTCACAATCCAGCCATATTGGCAT
                    ************************************************************
```

```
Gallo0272_exp      GTCTATTCAGTTCTTTGACGAGAAAGGACAGGTCATCACATTCAACGAGAAGAATCCGGC
Gallo0272_Seq      GTCTATTCAGTTCTTTGACGAGAAAGGACAGGTCATCACATTCAACGAGAAGAATCCGGC
                   ************************************************************

Gallo0272_exp      GTTAATTGCCTTCAATAGCCTGAACAAAACTGAAGTGTATGCGGGTTCAGGGTATGGCGA
Gallo0272_Seq      GTTAATTGCCTTCAATAGCCTGAACAAAACTGAAGTGTATGCGGGTTCAGGGTATGGCGA
                   ************************************************************

Gallo0272_exp      AAGCATCCACAACCTGAGCTCGAATATCAAAATCGAAACGATTGCGGGTAGTAGTGTCAT
Gallo0272_Seq      AAGCATCCACAACCTGAGCTCGAATATCAAAATCGAAACGATTGCGGGTAGTAGTGTCAT
                   ************************************************************

Gallo0272_exp      CTATAAAGACGGCGTGTTATATGCGGGCAATTACAACGATTATGTTTCCAATGGTAGTCG
Gallo0272_Seq      CTATAAAGACGGCGTGTTATATGCGGGCAATTACAACGATTATGTTTCCAATGGTAGTCG
                   ************************************************************

Gallo0272_exp      CTTTGATGCTAATCCAGCGACAGATCCGAATTCATATTGGGACGGTGATACCCAGGCGAA
Gallo0272_Seq      CTTTGATGCTAATCCAGCGACAGATCCGAATTCATATTGGGACGGTGATACCCAGGCGAA
                   ************************************************************

Gallo0272_exp      TCGCTGGTATGGAGCCGCAGTTGGGGTTGTGAGCTCCGGCGATACCATTAGCTTTGATGT
Gallo0272_Seq      TCGCTGGTATGGAGCCGCAGTTGGGGTTGTGAGCTCCGGCGATACCATTAGCTTTGATGT
                   ************************************************************

Gallo0272_exp      GGTAATGGATGCTGGTGCCGATGCCAAACGTCACGAATACGGCAAATTTTGGTTCGCGTT
Gallo0272_Seq      GGTAATGGATGCTGGTGCCGATGCCAAACGTCACGAATACGGCAAATTTTGGTTCGCGTT
                   ************************************************************

Gallo0272_exp      TTCGAGCGATGTTGCAGCTCCAGTGTTAACCCCGCCGACTCCTCCGGAAGTCCCCAACTA
Gallo0272_Seq      TTCGAGCGATGTTGCAGCTCCAGTGTTAACCCCGCCGACTCCTCCGGAAGTCCCCAACTA
                   ************************************************************

Gallo0272_exp      CAAGAAGGACCCTACGACCCCACCGGATTACCAGAAAGTAAACGTCCCGACTATTCAGAT
Gallo0272_Seq      CAAGAAGGACCCTACGACCCCACCGGATTACCAGAAAGTAAACGTCCCGACTATTCAGAT
                   ************************************************************

Gallo0272_exp      TAAAACCGATGTGCATGAAGTTGGCATTAACAAGACGACCAGTATTGATGTGCAGACCCC
Gallo0272_Seq      TAAAACCGATGTGCATGAAGTTGGCATTAACAAGACGACCAGTATTGATGTGCAGACCCC
                   ************************************************************

Gallo0272_exp      GCAGTTAGAGACAACTGTTCACGAAGTTGGGGTTAACAAAACCACGGAAATGGAAGTTGA
Gallo0272_Seq      GCAGTTAGAGACAACTGTTCACGAAGTTGGGGTTAACAAAACCACGGAAATGGAAGTTGA
                   ************************************************************

Gallo0272_exp      GACTCCCCAACTTGAAACGGATGTACACGAAGTGGGTATCAACAAAACGACGGAGATGAA
Gallo0272_Seq      GACTCCCCAACTTGAAACGGATGTACACGAAGTGGGTATCAACAAAACGACGGAGATGAA
                   ************************************************************

Gallo0272_exp      AGTCGAAACTCCACAGTTGAAAATGGACATGCACACCGTTGCCTATGATAAACCGGCAAC
Gallo0272_Seq      AGTCGAAACTCCACAGTTGAAAATGGACATGCACACCGTTGCCTATGATAAACCGGCAAC
                   ************************************************************

Gallo0272_exp      GCCTCAGGTGGTCAAGTCAAGCATCGTCGACAAACCTCCCACACCTCCCCCTGAACCTGA
Gallo0272_Seq      GCCTCAGGTGGTCAAGTCAAGCATCGTCGACAAACCTCCCACACCTCCCCCTA-------
                   **************************************************

Gallo0272_exp      AACATAAGCGGCCGCATC
Gallo0272_Seq      ------------------
```

Gallo0577 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo0577_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCGATACAGTGGACATTACGGTGAGCAATAC
Gallo0577_Seq      ----------------GGTCGCGTGGATCCGATACAGTGGACATTACGGTGAGCAATAC
                                   ************************************

Gallo0577_exp      CTCGTTAAGTACAAATGCTATCAATGGTGGTACGAGTACAGAATTCTCGTTCGATTTTGC
Gallo0577_Seq      CTCGTTAAGTACAAATGCTATCAATGGTGGTACGAGTACAGAATTCTCGTTCGATTTTGC
                   ************************************************************

Gallo0577_exp      CGTTCCGAATAGTGCGAAATCCGGTGATACGACCGTTATCTCGTTGCCGGACGAACTGAA
Gallo0577_Seq      CGTTCCGAATAGTGCGAAATCCGGTGATACGACCGTTATCTCGTTGCCGGACGAACTGAA
                   ************************************************************

Gallo0577_exp      TTTCCAACGCAACCAGACCTTCAACGTGTATGCCTCTGATGGTACAACGGTCGTGGCAAC
Gallo0577_Seq      TTTCCAACGCAACCAGACCTTCAACGTGTATGCCTCTGATGGTACAACGGTCGTGGCAAC
                   ************************************************************

Gallo0577_exp      CGCCGTGATTGACACCACAACTAAAACCCTGACACTGACTTACACGGACTATGTTGATAC
Gallo0577_Seq      CGCCGTGATTGACACCACAACTAAAACCCTGACACTGACTTACACGGACTATGTTGATAC
                   ************************************************************

Gallo0577_exp      GCACGATGATGTCACGGGGCATCTCTCAATGAACGTAGTCGTGGATCGCACCGTTGTGAC
Gallo0577_Seq      GCACGATGATGTCACGGGGCATCTCTCAATGAACGTAGTCGTGGATCGCACCGTTGTGAC
                   ************************************************************

Gallo0577_exp      GGAAGCGACGACTGTTCCAGCCACTGTTACCATTAACGGCACTACCACGATTACGATTTC
Gallo0577_Seq      GGAAGCGACGACTGTTCCAGCCACTGTTACCATTAACGGCACTACCACGATTACGATTTC
                   ************************************************************

Gallo0577_exp      TTCCGGCGGAATTAACTACACCGTTTCTACAGGCGATAGCGATGACATCGATTTCTGGAA
Gallo0577_Seq      TTCCGGCGGAATTAACTACACCGTTTCTACAGGCGATAGCGATGACATCGATTTCTGGAA
                   ************************************************************

Gallo0577_exp      ATACGGCGTAAGCTATTCCGATGATGAAGTCATGTACCTGATTAACGTGAACACTTCCGC
Gallo0577_Seq      ATACGGCGTAAGCTATTCCGATGATGAAGTCATGTACCTGATTAACGTGAACACTTCCGC
                   ************************************************************

Gallo0577_exp      TGCGACGGTATCGAATGTGGTGATCTCAGATACGATCAATTCAACTGGACTGGAGTACGT
Gallo0577_Seq      TGCGACGGTATCGAATGTGGTGATCTCAGATACGATCAATTCAACTGGACTGGAGTACGT
                   ************************************************************

Gallo0577_exp      TGACGGGTCTTTTGAAATCTTTGAGGGTACCTGGTATAAGAATGCGCAGAACTACTGGGC
Gallo0577_Seq      TGACGGGTCTTTTGAAATCTTTGAGGGTACCTGGTATAAGAATGCGCAGAACTACTGGGC
                   ************************************************************

Gallo0577_exp      ATTGGGAGGCAGTACCAACGTGACGTCGAACTACAACATCGAGCTGTCAGCAGACAATAC
Gallo0577_Seq      ATTGGGAGGCAGTACCAACGTGACGTCGAACTACAACATCGAGCTGTCAGCAGACAATAC
                   ************************************************************

Gallo0577_exp      GTCGTTTAGCATTAATCTGGGTACCATTTCGAAAGGCTACATGATTCGGTATCGTGTCAA
Gallo0577_Seq      GTCGTTTAGCATTAATCTGGGTACCATTTCGAAAGGCTACATGATTCGGTATCGTGTCAA
                   ************************************************************

Gallo0577_exp      AGCGAATTACACCCTCATTAATGGCGAACAGCTGTCCAATAGCGCGACTTATTACAGCGA
Gallo0577_Seq      AGCGAATTACACCCTCATTAATGGCGAACAGCTGTCCAATAGCGCGACTTATTACAGCGA
                   ************************************************************

Gallo0577_exp      AAACACCGCCCTGAACAACGCCGACAATACCTTTACGTATCAAGGCGCGAGCGGTACGGC
Gallo0577_Seq      AAACACCGCCCTGAACAACGCCGACAATACCTTTACGTATCAAGGCGCGAGCGGTACGGC
                   ************************************************************

Gallo0577_exp      CAGTGGCTATAATTACTCCCTCACCGTACAGAAAGTGAACGAAGCAGGCGAAGCATTAGC
Gallo0577_Seq      CAGTGGCTATAATTACTCCCTCACCGTACAGAAAGTGAACGAAGCAGGCGAAGCATTAGC
                   ************************************************************
```

```
Gallo0577_exp      AGGCGCGGAGTTCACCGTTACGCGTGAAAGCACTGGACAAGTGGTCGGGACGATCACGAC
Gallo0577_Seq      AGGCGCGGAGTT-------------------------------------------------
                   ************

Gallo0577_exp      CGGCTCAGACGGTACAGCCACCATTTCAGGTTTACTGAAAGACAATTACATCATTACCGA
Gallo0577_Seq      ------------------------------------------------------------


Gallo0577_exp      AACGAAAGCTCCTACTGGGTACGCCATTGCCGATCCAGTGACGGCTGAAGCCGATAACAG
Gallo0577_Seq      ------------------------------------GACGGCTGAAGCCGATAACAG
                                                       ********************

Gallo0577_exp      TACGGTCACCGTTACTGACAAGAAAGCGACCGTGGAAGTAACCGGTACCAAAACGTGGGA
Gallo0577_Seq      TACGGTCACCGTTACTGACAAGAAAGCGACCGTGGAAGTAACCGGTACCAAAACGTGGGA
                   ************************************************************

Gallo0577_exp      TGACAACAACGATCAAGATGGTAAGCGTCCCGATTCCATCACTGTTAATCTGTTAGCGAA
Gallo0577_Seq      TGACAACAACGATCAAGATGGTAAGCGTCCCGATTCCATCACTGTTAATCTGTTAGCGAA
                   ************************************************************

Gallo0577_exp      CGGTACAGTAGTTGATACCAAAACAGTCACAGCGGATGACAATTGGACTTATGCGTTTAG
Gallo0577_Seq      CGGTACAGTAGTTGATACCAAAACAGTCACAGCGGATGACAATTGGACTTATGCGTTTAG
                   ************************************************************

Gallo0577_exp      CGACCTGGATCAGTATGACGCTGATGGTAACGAAATTGCCTACACTGTGTCGGAGGAAAT
Gallo0577_Seq      CGACCTGGATCAGTATGACGCTGATGGTAACGAAATTGCCTACACTGTGTCGGAGGAAAT
                   ************************************************************

Gallo0577_exp      GGTTGATGGGTATACGACAGTCGTCGATGGCTATAACATCACCAATACCCACGCATCAGA
Gallo0577_Seq      GGTTGATGGGTATACGACAGTCGTCGATGGCTATAACATCACCAATACCCACGCATCAGA
                   ************************************************************

Gallo0577_exp      AACCACCGAAGTTTCAGGCACTAAAACATGGGATGATAACGACGACCAAGATGGCAAACG
Gallo0577_Seq      AACCACCGAAGTTTCAGGCACTAAAACATGGGATGATAACGACGACCAAGATGGCAAACG
                   ************************************************************

Gallo0577_exp      CCCGGATTCCATCACGGTGAACCTGCTGGCAAATGGCACGGTCGTGGATACGAAAACGGT
Gallo0577_Seq      CCCGGATTCCATCACGGTGAACCTGCTGGCAAATGGCACGGTCGTGGATACGAAAACGGT
                   ************************************************************

Gallo0577_exp      AACAGCCGATGATAATTGGTCTTATAGCTTTACCGATTTGCCGAAATACGATAATGGAAA
Gallo0577_Seq      AACAGCCGATGATAATTGGTCTTATAGCTTTACCGATTTGCCGAAATACGATAATGGAAA
                   ************************************************************

Gallo0577_exp      CGAGATCACATACACCGTAACCGAAGATACAGTCGCTGACTATACAACTACGTATGACGG
Gallo0577_Seq      CGAGATCACATACACCGTAACCGAAGATACAGTCGCTGACTATACAACTACGTATGACGG
                   ************************************************************

Gallo0577_exp      GTACAACATTACCAACAGTTACACCCCGGGTGAAACCAGTATCACCGTCACCAAAGTGTG
Gallo0577_Seq      GTACAACATTACCAACAGTTACACCCCGGGTGAAACCAGTATCACCGTCACCAAAGTGTG
                   ************************************************************

Gallo0577_exp      GGACGACAATAATGATCAGGACGGTATTCGCCCTGATGCGATTCAGGTGCAGCTGTATGC
Gallo0577_Seq      GGACGACAATAATGATCAGGACGGTATTCGCCCTGATGCGATTCAGGTGCAGCTGTATGC
                   ************************************************************

Gallo0577_exp      GAATGGCGAGAAAAGCGGTGATGTGATCACTCTTACGGTCGCAGACAACTGGACCTATAC
Gallo0577_Seq      GAATGGCGAGAAAAGCGGTGATGTGATCACTCTTACGGTCGCAGACAACTGGACCTATAC
                   ************************************************************

Gallo0577_exp      CTGGACTGGTTTGGCTGAGAAAGCGAACAAGAAAACTATCACTTACACGGTAGAAGAGGT
Gallo0577_Seq      CTGGACTGGTTTGGCTGAGAAAGCGAACAAGAAAACTATCACTTACACGGTAGAAGAGGT
                   ************************************************************

Gallo0577_exp      TAGTGCAGTTGACGGGTATACCGCGACAGTAGGCGAGGTCGAAAATGGCAATGTGACAAT
Gallo0577_Seq      TAGTGCAGTTGACGGGTATACCGCGACAGTAGGCGAGGTCGAAAATGGCAATGTGACAAT
                   ************************************************************
```

```
Gallo0577_exp      CACCAACACCCATACTCCTACGACCCCAGAAACTCCGAGCAGCGATGAACCGACAACCCC
Gallo0577_Seq      CACCAACACCCATACTCCTACGACCCCAGAAACTCCGAGCAGCGATGAACCGACAACCCC
                   ************************************************************

Gallo0577_exp      GTCGCAAAGCAACAAGAAATCTGATAAAGAGCAGGATAAGAACATTATCGCTGCGCTTGT
Gallo0577_Seq      GTCGCAAAGCAACAAGAAATCTGATAAAGAGCAGGATAAGAACATTATCGCTGCGCTTGT
                   ************************************************************

Gallo0577_exp      CGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATC
Gallo0577_Seq      CGACAAACCTCCCACACCTCCCCCTA-------------------------
                   *************************
```

Gallo0748 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo0748_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCGATGAACTCTCCAAAGCTGCGGGTGTGAG
Gallo0748_Seq      ---------------------------TCCGATGAACTCTCCAAAGCTGCGGGTGTGAG
                                              ********************************

Gallo0748_exp      CCAGACCGATCCGGCGTCGAACATTGAGCAAGTGGTGCAAGCAACCGAATCCTCTAGCAC
Gallo0748_Seq      CCAGACCGATCCGGCGTCGAACATTGAGCAAGTGGTGCAAGCAACCGAATCCTCTAGCAC
                   ************************************************************

Gallo0748_exp      AGCTGATTTTGCCCAGGTGGCATCCGTTGAAGCGACCACAGAAGTGAGCGGAGTGGAAAG
Gallo0748_Seq      AGCTGATTTTGCCCAGGTGGCATCCGTTGAAGCGACCACAGAAGTGAGCGGAGTGGAAAG
                   ************************************************************

Gallo0748_exp      CACAGCTACTGTTTCGGTAACAGCGGACGAAGTTGCTGTGGTAAGCAAAACTCAAGAAAT
Gallo0748_Seq      CACAGCTACTGTTTCGGTAACAGCGGACGAAGTTGCTGTGGTAAGCAAAACTCAAGAAAT
                   ************************************************************

Gallo0748_exp      TGTATCGGAAGAGTTGAGTAGTCCGGCCGCAACGTCTGATGCGACCGCTGTTGGGAACGT
Gallo0748_Seq      TGTATCGGAAGAGTTGAGTAGTCCGGCCGCAACGTCTGATGCGACCGCTGTTGGGAACGT
                   ************************************************************

Gallo0748_exp      AGCTAACGCACAGAATTCGGGCGTTTCTAGTGAAGTCGCGGAAGAGATTGCGCAAGACGT
Gallo0748_Seq      AGCTAACGCACAGAATTCGGGCGTTTCTAGTGAAGTCGCGGAAGAGATTGCGCAAGACGT
                   ************************************************************

Gallo0748_exp      TGAAGCATCTGCCACCAGTGTGAGCTCAGAAGTTGTCACGGAAGTTACGGAGAAAGCCCA
Gallo0748_Seq      TGAAGCATCTGCCACCAGTGTGAGCTCAGAAGTTGTCACGGAAGTTACGGAGAAAGCCCA
                   ************************************************************

Gallo0748_exp      GTCTGAGGAACAGACGTTAGATTCCGCCACCCCGCAGTCTATCGACTCGGACGAATTGAT
Gallo0748_Seq      GTCTGAGGAACAGACGTTAGATTCCGCCACCCCGCAGTCTATCGACTCGGACGAATTGAT
                   ************************************************************

Gallo0748_exp      CACGGTACCGGAAGCGTGGGAATCGGGCTATAAAGGCCAGGGCACCATTGTGGCTATCAT
Gallo0748_Seq      CACGGTACCGGAAGCGTGGGAATCGGGCTATAAAGGCCAGGGCACCATTGTGGCTATCAT
                   ************************************************************

Gallo0748_exp      TGACTCAGGGCTGGATGTAGAACATGATGTGCTGCACATTAGCGACTTAAGTACCGCCAA
Gallo0748_Seq      TGACTCAGGGCTGGATGTAGAACATGATGTGCTGCACATTAGCGACTTAAGTACCGCCAA
                   ************************************************************

Gallo0748_exp      ATATGGGTCGGAGGAAGAAATTGAGGCGGCGAAAGCAGCCGCGGGTATTACGTATGGCAA
Gallo0748_Seq      ATATGGGTCGGAGGAAGAAATTGAGGCGGCGAAAGCAGCCGCGGGTATTACGTATGGCAA
                   ************************************************************

Gallo0748_exp      ATGGTTCAATGATAAAGTCGTGTTTGGTTACAACTACGTGGACGGGAATACCATCCTGAA
Gallo0748_Seq      ATGGTTCAATGATAAAGTCGTGTTTGGTTACAACTACGTGGACGGGAATACCATCCTGAA
                   ************************************************************

Gallo0748_exp      AGAGGGAGAAGAAGCGTCCCATGGCATGCACGTCACCGGGATCGCTACCGGGAATCCGAC
Gallo0748_Seq      AGAGGGAGAAGAAGCGTCCCATGGCATGCACGTCACCGGGATCGCTACCGGGAATCCGAC
                   ************************************************************
```

```
Gallo0748_exp      CAAAGCATTGGGAGATGAATACATCTACGGTGTAGCGCCGGAGGCACAGGTCATCTTCCT
Gallo0748_Seq      CAAAGCATTGGGAGATGAATACATCTACGGTGTAGCGCCGGAGGCACAGGTCATCTTCCT
                   ************************************************************

Gallo0748_exp      GCGTGTCTTTAGTGATCTGAAATCCTATACCGGCCCTGCGCTGTATGTCCGTGCGATCGA
Gallo0748_Seq      GCGTGTCTTTAGTGATCTGAAATCCTATACCGGCCCTGCGCTGTATGTCCGTGCGATCGA
                   ************************************************************

Gallo0748_exp      GGATGCAGTGAAACTGGGTGCTGACAGCATCAACCTGAGTCTGGGCTCGACAACTGGCAG
Gallo0748_Seq      GGATGCAGTGAAACTGGGTGCTGACAGCATCAACCTGAGTCTGGGCTCGACAACTGGCAG
                   ************************************************************

Gallo0748_exp      CGAGGTCAACATGGATGAAACCTTAATTGCAGCCATCAAAGCAGCACAGAAAGCGGGTGT
Gallo0748_Seq      CGAGGTCAACATGGATGAAACCTTAATTGCAGCCATCAAAGCAGCACAGAAAG-------
                   *************************************************

Gallo0748_exp      AAACGTGGCTATTAGCGCGGGCAATGATGGCGTATTTGGCGATAGCATTAATCCGAGCGC
Gallo0748_Seq      ------------------------------------------------------------

Gallo0748_exp      AGAAAATCCCGATTATGGCCTGGTAGGTAACCCCAGCACGACGCAGGATGTTATTAGCGT
Gallo0748_Seq      ------------------------------------------------------------

Gallo0748_exp      TGCGTCGTACAATAACTCAATCACTCGCAGCAATGTTGTGACGTTTGTTGGTATGGAAGA
Gallo0748_Seq      ------------------------------------------------------------

Gallo0748_exp      TAACGCTGAACTGAACAATGGCAAATCTTCCTTCACCAACCCGGACAAAAGCGACAAGAA
Gallo0748_Seq      ------------------------------------------------------------

Gallo0748_exp      ATTCGAAAATGGAAAGGCGTATGATTATGTGTACGTTGGCACGGGGACTGCCGAGGAACT
Gallo0748_Seq      ------------------------------------------------------------

Gallo0748_exp      TGAAGGTGTGGACTTGACCGGGAAGCTGGCTCTGATTCAACGCGGTGGTCTTACGTTTTC
Gallo0748_Seq      ------------------------------------------------------------

Gallo0748_exp      GGAAAAGATTGCGAACGCGACTGCACATGGCGCCGAGGGTGTGATTATTTTCAACAACGA
Gallo0748_Seq      ---------------------------TGGCGCCGAGGGTGTGATTATTTTCAACAACGA
                                              *********************************

Gallo0748_exp      TCCAGATGGAAGTAATGTTTCTATGGCCATTGACGATACTGCCATTGCAATTCCTTCTGC
Gallo0748_Seq      TCCAGATGGAAGTAATGTTTCTATGGCCATTGACGATACTGCCATTGCAATTCCTTCTGC
                   ************************************************************

Gallo0748_exp      GTTTATCCCGTACAAGTTCGGTATTGAGCTGGCCAAAGGCGGTTACCAGATCAAGTTCTC
Gallo0748_Seq      GTTTATCCCGTACAAGTTCGGTATTGAGCTGGCCAAAGGCGGTTACCAGATCAAGTTCTC
                   ************************************************************

Gallo0748_exp      CGATGTCGCCGAGAAATTCGATAATCCCGGAGCGGGCAAGTTCAGTAGTTTCAGCTCATG
Gallo0748_Seq      CGATGTCGCCGAGAAATTCGATAATCCCGGAGCGGGCAAGTTCAGTAGTTTCAGCTCATG
                   ************************************************************

Gallo0748_exp      GGGACTGACCGCCGATGGCGAACTGAAGCCAGATGTGGCGGCACCAGGCGGGTCAATCTA
Gallo0748_Seq      GGGACTGACCGCCGATGGCGAACTGAAGCCAGATGTGGCGGCACCAGGCGGGTCAATCTA
                   ************************************************************

Gallo0748_exp      TTCGTCTTACAACAACGACAAATACGGCTCTATGTCCGGTACCTCAATGGCCTCACCGCA
Gallo0748_Seq      TTCGTCTTACAACAACGACAAATACGGCTCTATGTCCGGTACCTCAATGGCCTCACCGCA
                   ************************************************************

Gallo0748_exp      TGTTGCCGGTGTGATCGCGCTTGTGAAACAGTACCTGAAAGAGAACTTTCCAGAGAAATC
Gallo0748_Seq      TGTTGCCGGTGTGATCGCGCTTGTGAAACAGTACCTGAAAGAGAACTTTCCAGAGAAATC
                   ************************************************************
```

```
Gallo0748_exp      CGATGAGGAAGTCGGCTATCTCGTTAAAGCCTTAATTATGAGCACCGCCAAAGCGCACTA
Gallo0748_Seq      CGATGAGGAAGTCGGCTATCTCGTTAAAGCCTTAATTATGAGCACCGCCAAAGCGCACTA
                   ************************************************************

Gallo0748_exp      TGACAAAGAAGCCCAAGCCTATACTAGTCCTCGTCAGCAAGGTGCGGGATTAGTCGATAC
Gallo0748_Seq      TGACAAAGAAGCCCAAGCCTATACTAGTCCTCGTCAGCAAGGTGCGGGATTAGTCGATAC
                   ************************************************************

Gallo0748_exp      TGCGTCAGCTGTCTCAACGGGCCTGTACGTGACGGGTGATGATGGCTACGGTAGTGTCAC
Gallo0748_Seq      TGCGTCAGCTGTCTCAACGGGCCTGTACGTGACGGGTGATGATGGCTACGGTAGTGTCAC
                   ************************************************************

Gallo0748_exp      TCTGGGGAACGTGGGTGATACCTTCACCTTTGACGTCACCATCCACAATATTGGTGACCA
Gallo0748_Seq      TCTGGGGAACGTGGGTGATACCTTCACCTTTGACGTCACCATCCACAATATTGGTGACCA
                   ************************************************************


Gallo0748_exp      AGATAAAACTCTGACGTATGAAACGAACTTAGGCACAGACACAGTTGAAAATGGCGAAAT
Gallo0748_Seq      AGATAAAACTCTGACGTATGAAACGAACTTAGGCACAGACACAGTTGAAAATGGCGAAAT
                   ************************************************************

Gallo0748_exp      CACCCTTGCACCTCGGCAGTTGTCCACGACAACCGGTCATACCATTACCGTTAAGGCGAA
Gallo0748_Seq      CACCCTTGCACCTCGGCAGTTGTCCACGACAACCGGTCATACCATTACCGTTAAGGCGAA
                   ************************************************************

Gallo0748_exp      TAGCTCGGAAACCATCACAATTACCGTGGACGCATCCCAGTTTGCGGAACTGCTCAGCAA
Gallo0748_Seq      TAGCTCGGAAACCATCACAATTACCGTGGACGCATCCCAGTTTGCGGAACTGCTCAGCAA
                   ************************************************************

Gallo0748_exp      AGAAATTCCGAATGGCTATTATCTGGAGGGCTTTGTGCGCTTTCTCGATCCGACGGATCT
Gallo0748_Seq      AGAAATTCCGAATGGCTATTATCTGGAGGGCTTTGTGCGCTTTCTCGATCCGACGGATCT
                   ************************************************************

Gallo0748_exp      GGCCGAAGTCATCAGCATTCCGTATGTGGGTTTTCGCGGTGACTTTGTCGACAAACCTCC
Gallo0748_Seq      GGCCGAAGTCATCAGCATTCCGTATGTGGGTTTTCGCGGTGACTTTGTCGACAAACCTCC
                   ************************************************************

Gallo0748_exp      CACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATC
Gallo0748_Seq      CACACCTCCCC----------------------------
                   ***********
```

Gallo0933 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo0933_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCCTGACTGCGTGCAGCAGCTCTTCCAACTC
Gallo0933_Seq      ---------------TTGTCGCGTGGATCCCTGACTGCGTGCAGCAGCTCTTCCAACTC
                                  *  ************************************

Gallo0933_exp      ATCAACTAGCAGCAGTAGTAGTCAGAATACGACAGCGTCAACCAGCTCTTTAAGCAGCGG
Gallo0933_Seq      ATCAACTAGCAGCAGTAGTAGTCAGAATACGACAGCGTCAACCAGCTCTTTAAGCAGCGG
                   ************************************************************

Gallo0933_exp      CGAAGTCTCCACAACCCTGGATAAAGTGGACAACTCTAAATGGCAGTATAATGCGGATGA
Gallo0933_Seq      CGAAGTCTCCACAACCCTGGATAAAGTGGACAACTCTAAATGGCAGTATAATGCGGATGA
                   ************************************************************

Gallo0933_exp      CAATGTGTACTACCAGATCGGGATTTCGTACGCTGCAAACCCGACAGATGCTGAACAGCA
Gallo0933_Seq      CAATGTGTACTACCAGATCGGGATTTCGTACGCTGCAAACCCGACAGATGCTGAACAGCA
                   ************************************************************

Gallo0933_exp      GACGTTATCCATTTTCGTGCCAGGCGATTATATGACCGCGACGGATAACGGTAATGGTAC
Gallo0933_Seq      GACGTTATCCATTTTCGTGCCAGGCGATTATATGACCGCGACGGATAACGGTAATGGTAC
                   ************************************************************

Gallo0933_exp      CTATACGTGCGAAATTAACACGTCGGCCACAGTCGGAAACTACACTAGCGAAACCGCGCC
Gallo0933_Seq      CTATACGTGCGAAATTAACACGTCGGCCACAGTCGGAAACTACACTAGCGAAACCGCGCC
                   ************************************************************
```

```
Gallo0933_exp      GATTGTGATTCCCATCAACACCCCGGGCTATTCCGCCATGTCGGCCTTAACAGAGTATAC
Gallo0933_Seq      GATTGTGATTCCCATCAACACCCCGGGCTATTCCGCCATGTCGGCCTTAACAGAGTATAC
                   ************************************************************

Gallo0933_exp      CTCAGATGCGACCGACTATACCTCGCAAGGCATGATTTACGTTAGCGCCGGATTACGTGG
Gallo0933_Seq      CTCAGATGCGACCGACTATACCTCGCAAGGCATGATTTACGTTAGCGCCGGATTACGTGG
                   ************************************************************

Gallo0933_exp      ACGCGATAGTGGCGCACCTAGCGGTGTTACCGATGCCAAAGCAGCGATTCGCTATCTCCG
Gallo0933_Seq      ACGCGATAGTGGCGCACCTAGCGGTGTTACCGATGCCAAAGCAGCGATTCGCTATCTCCG
                   ************************************************************

Gallo0933_exp      CTATAATCAGGGTAACATTTCCGGCAATACCGACAGCATCTTCGTGTTCGGCATGAGTGG
Gallo0933_Seq      CTATAATCAGGGTAACATTTCCGGCAATACCGACAGCATCTTCGTGTTCGGCATGAGTGG
                   ************************************************************

Gallo0933_exp      TGGAGGTGCACAATCTGCGATTATTGGCAGCAGTGGGGACAGTTCCTTGTATGACGACTA
Gallo0933_Seq      TGGAGGTGCACAATCTGCGATTATTGGCAGCAGTGGGGACAGTTCCTTGTATGACGACTA
                   ************************************************************

Gallo0933_exp      CCTGACGGAGATCGGGGCTGTTGAGGGCGTTAGCGACAGTGTAGCTGGTGTAATGGCCTG
Gallo0933_Seq      CCTGACGGAGATCGGGGCTGTTGAGGGCGTTAGCGACAGTGTAGCTGGTGTAATGGCCTG
                   ************************************************************

Gallo0933_exp      GTGTCCGATTACTAATCTGGACACGGCCAACGAAGCCTATGAATGGAACATGGGTAGTAC
Gallo0933_Seq      GTGTCCGATTACTAATCTGGACACGGCCAACGAAGCCTATGAATGGAACATGGGTAGTAC
                   ************************************************************

Gallo0933_exp      CCGTTCTGACTTGAGTGACGAGGAACAGACCATCTCAGATGGATTGGCTACCGCCTTTGC
Gallo0933_Seq      CCGTTCTGACTTGAGTGACGAGGAACAGACCATCTCAGATGGATTGGCTACCGCCTTTGC
                   ************************************************************

Gallo0933_exp      CAAATACATCAACAAACTTGGGCTTCAGGATGAAGATGGGAACAAACTGACCCTGAAGAA
Gallo0933_Seq      CAAATACATCAACAAACTTGGGCTTCAGGATGAAGATGGGAACAAACTGACCCTGAAGAA
                   ************************************************************

Gallo0933_exp      ATCGGACGACGGAATCTATCAAGCAGGCTCGTACTACAATTACCTGAAATCCGTGATCGA
Gallo0933_Seq      ATCGGACGACGGAATCTATCAAGCAGGCTCGTACTACAATTACCTGAAATCCGTGATCGA
                   ************************************************************

Gallo0933_exp      ATCGGACGACGGAATCTATCAAGCAGGCTCGTACTACAATTACCTGAAATCCGTGATCGA
Gallo0933_Seq      ATCGGACGACGGAATCTATCAAGCAGGCTCGTACTACAATTACCTGAAATCCGTGATCGA
                   ************************************************************

Gallo0933_exp      AGATAGTCTGAACACCTTTCTCGCGAATACCACCTTTCCGTACGATGCAAGCTCATCAAG
Gallo0933_Seq      AGATAGTCTGAACACCTTTCTCGCGAATACCACCTTTCCGTACGATGCAAGCTCATCAAG
                   ************************************************************

Gallo0933_exp      CCAAGGCGGTCTTGGCGGTGGGGATATGCCAACTGGCGAAGCACCTACGGATCTGGGTAC
Gallo0933_Seq      CCAAGGCGGTCTTGGCGGTGGGGATATGCCAACTGGCGAAGCACCTACGGATCTGGGTAC
                   ************************************************************

Gallo0933_exp      GACGGATGACACGACCTCTATTGAGGACGTTGATGATATCAATCGCACGAGCTCTTCGAG
Gallo0933_Seq      GACGGATGACACGACCTCTATTGAGGACGTTGATGATATCAATCGCACGAGCTCTTCGAG
                   ************************************************************

Gallo0933_exp      CATCACTATTGATCTGTCTGGTACTTACGAGACTGCAGCCGACTACATTGCAGCATTGAA
Gallo0933_Seq      CATCACTATTGATCTGTCTGGTACTTACGAGACTGCAGCCGACTACATTGCAGCATTGAA
                   ************************************************************

Gallo0933_exp      CGCCGATTCCACGTGGGTCACGTATGACGAAGATACCAATACGGCTTCAATTAGCAGCAT
Gallo0933_Seq      CGCCGATTCCACGTGGGTCACGTATGACGAAGATACCAATACGGCTTCAATTAGCAGCAT
                   ************************************************************

Gallo0933_exp      TGCGGATTTCGTGAAGTACATGAAGTCGAGCACGAAATCCCTGGGTGCGTTTGATGCGCT
Gallo0933_Seq      TGCGGATTTCGTGAAGTACATGAAGTCGAGCACGAAATCCCTGGGTGCGTTTGATGCGCT
                   ************************************************************
```

```
Gallo0933_exp      CGATCTGAGCCAGGGCGAAAACCAACTGTTTGGTTATGGCGATGGCAATTCCGTGCATTG
Gallo0933_Seq      CGATCTGAGCCAGGGCGAAAACCAACTGTTTGGTTATGGCGATGGCAATTCCGTGCATTG
                   ************************************************************

Gallo0933_exp      GGATTCTACCCTGGGCGATCTGTTTAAAGGCACTGATTATGAAGAAGCGTTTACAACAGA
Gallo0933_Seq      GGATTCTACCCTGGGCGATCTGTTTAAAGGCACTGATTATGAAGAAGCGTTTACAACAGA
                   ************************************************************

Gallo0933_exp      CCTCGTTAAGACGGATAGTCTGGGTAATGATTTAACTACCCGCATCAACATGTATACCCC
Gallo0933_Seq      CCTCGTTAAGACGGATAGTCTGGGTAATGATTTAACTACCCGCATCAACATGTATACCCC
                   ************************************************************

Gallo0933_exp      GCTGTATTATCTGACCGATTACTATGGTGGGGAAAATTCCTCGAACGTCGCGTCGTATTG
Gallo0933_Seq      GCTGTATTATCTGACCGATTACTATGGTGGGGAAAATTCCTCGAACGTCGCGTCGTATTG
                   ************************************************************

Gallo0933_exp      GCGGATTCGTACAGGGTTATCCCAAGGCGATACAGCGCTGACCACTGAGGTAAATCTGGC
Gallo0933_Seq      GCGGATTCGTACAGGGTTATCCCAAGGCGATACAGCGCTGACCACTGAGGTAAATCTGGC
                   ************************************************************

Gallo0933_exp      CCTGGCGCTTGAAAACTATGGTGTGAAAGATCTGGATTTCGCTACCGTATGGGGCGAACA
Gallo0933_Seq      CCTGGCGCTTGAAAACTATGGTGTGAAAGATCTGGATTTCGCTACCGTATGGGGCGAACA
                   ************************************************************

Gallo0933_exp      GCACACCGAAGCTGAGATCTCTGGCGACTCAACCTCGAACTTCATCGATTGGGTCAATCA
Gallo0933_Seq      GCACACCGAAGCTGAGATCTCTGGCGACTCAACCTCGAACTTCATCGATTGGGTCAATCA
                   ************************************************************

Gallo0933_exp      GTCTTTGGCGGACAACTCGGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATA
Gallo0933_Seq      GTCTTTGGCGGACAACTCGGTCGACAAACCTCCCACACCTCCCCCT--------------
                   *******************************************

Gallo0933_exp      AGCGGCCGCATC
Gallo0933_Seq      ------------
```

Gallo1570 (forward sequencing of miniPrep DNA via primer pGEXfor, reverse sequencing of PCR product via T3 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo1570_exp      ATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA
Gallo1570_Seq      ----------------------------------------------CATCCTCCAAA
                                                                 *  * ***

Gallo1570_exp      TCGGATCTGGTTCCGCGTGGATCCAAAGCCGAAGAGGATGTGTACTATACCGGCTATACC
Gallo1570_Seq      TCGGATCTGGTTCCGCGTGGATCCAAAGCCGAAGAGGATGTGTACTATACCGGCTATACC
                   ************************************************************

Gallo1570_exp      TCGGATATTTCCCTGAACAGTAGTTACATCAATCCGGACCCAGGGCCATATGCGATTGAC
Gallo1570_Seq      TCGGATATTTCCCTGAACAGTAGTTACATCAATCCGGACCCAGGGCCATATGCGATTGAC
                   ************************************************************

Gallo1570_exp      GAAGGCGGGGAGTCGAAATTAGCCTATTGCTTTAACCGGAACAAATCGCGTCCTCCTGCA
Gallo1570_Seq      GAAGGCGGGGAGTCGAAATTAGCCTATTGCTTTAACCGGAACAAATCGCGTCCTCCTGCA
                   ************************************************************

Gallo1570_exp      AAGTCAGAACCGGAGGACGGTGAAGCGAAATACCGCAAAATTGCAGATGTGGATTACGTC
Gallo1570_Seq      AAGTCAGAACCGGAGGACGGTGAAGCGAAATACCGCAAAATTGCAGATGTGGATTACGTC
                   ************************************************************

Gallo1570_exp      CGCCTTAAAGAGAACTGTTCGTCTGACATGGAAGGCCGTGAATTGTACGATGCCATCATG
Gallo1570_Seq      CGCCTTAAAGAGAACTGTTCGTCTGACATGGAAGGCCGTGAATTGTACGATGCCATCATG
                   ************************************************************

Gallo1570_exp      AAAGTGATCTACAACGGGTATCCGAACAATTGTAGCGGCATCAATGGCAAATATCGCCTG
Gallo1570_Seq      AAAGTGATCTACAACGGGTATCCGAACAATTGTAGCGGCATCAATGGCAAATATCGCCTG
                   ************************************************************
```

```
Gallo1570_exp      AAAGACGGCGACTTTTGCGCGATTACCCAGTGGGCTATCTGGCACTTTACGGATGGCGCG
Gallo1570_Seq      AAAGACGGCGACTTTTGCGCGATTACCCAGTGGGCTATCTGGCACTTTACGGATGGCGCG
                   ************************************************************

Gallo1570_exp      GATAGCGATGGTACCGGCAATCTGCCGTATTATGGGAAAGAAAGCATGTGGAACCGCTCA
Gallo1570_Seq      GATAGCGATGGTACCGGCAATCTGCCGTATTATGGG----------------------
                   *********************************

Gallo1570_exp      GACGTGAAAGAAGCCTATCTCGAGCTCATCGATGTTGCGAACCTGTCTTACCCAGCAGAC
Gallo1570_Seq      ------------------------------------------------------------

Gallo1570_exp      GCAAAACTGAACCTGTATATTTACGATCATGGTGCCGAACACGATCGCCAGAATCTGCTT
Gallo1570_Seq      ------------------------------------------------------------

Gallo1570_exp      ACCACGGACGTAGGCTATACAAATCTGTCTGTCGAGAAAGTGTGGAATGACAGCGATGAT
Gallo1570_Seq      ------------------------------------------------------------

Gallo1570_exp      CAGGATGGTATTCGTCCGGCTTTTATCGATGTACAGCTGTTAGCGAATGGAGTGGAAGTT
Gallo1570_Seq      ------------------------------------------------------------

Gallo1570_exp      GAGGGACAGAAAATCGAACTGTCCAAATTTCTGAATTCGAACTGGCAAGGTGTATTCCGT
Gallo1570_Seq      ------------------------------------------------------------

Gallo1570_exp      GGTCTTAGTCTCTACGATAGTGACGGTAATCCTATCGAATATTCCGTGAAGGAAGTTGAG
Gallo1570_Seq      ------AGTCTCTACGATAGTGACGGTAATCCTATCGAATATTCCGTGAAGGAAGTTGAG
                         **************************************************

Gallo1570_exp      AAGTACCGCGGACAGTTGGATGGTTACCAGTCTACTGTGACGAAAAGCGACAGCGGCTAT
Gallo1570_Seq      AAGTACCGCGGACAGTTGGATGGTTACCAGTCTACTGTGACGAAAAGCGACAGCGGCTAT
                   ************************************************************

Gallo1570_exp      TCCTATACCATCACCAATACACACGTTCCGGAAACAACCGAAATTAGCGGTACTAAAACG
Gallo1570_Seq      TCCTATACCATCACCAATACACACGTTCCGGAAACAACCGAAATTAGCGGTACTAAAACG
                   ************************************************************

Gallo1570_exp      TGGGATGATAAAGACGATCAAGACGGGAAACGTCCCTCTAGCATTACGGTGAAATTACTG
Gallo1570_Seq      TGGGATGATAAAGACGATCAAGACGGGAAACGTCCCTCTAGCATTACGGTGAAATTACTG
                   ************************************************************

Gallo1570_exp      GCTGATGATGAGGAAATCGATAGTCAAGAGGTGACGGCAGATACGGACTGGAAGTACAGC
Gallo1570_Seq      GCTGATGATGAGGAAATCGATAGTCAAGAGGTGACGGCAGATACGGACTGGAAGTACAGC
                   ************************************************************

Gallo1570_exp      TTTAAAGATCTGCCGAAATATAAGAACGAAGGCGTCGAAATTAACTATTCAGTCGCCGAA
Gallo1570_Seq      TTTAAAGATCTGCCGAAATATAAGAACGAAGGCGTCGAAATTAACTATTCAGTCGCCGAA
                   ************************************************************

Gallo1570_exp      GAATCAGTGAGCGATTATGAAACCACCATCAGCGGTACGGATATTACGAACACTCATGTC
Gallo1570_Seq      GAATCAGTGAGCGATTATGAAACCACCATCAGCGGTACGGATATTACGAACACTCATGTC
                   ************************************************************

Gallo1570_exp      CCGGAAACAACAGAAATTTCGGGAACTAAAACCTGGGACGATAACGATGACCAAGATGGC
Gallo1570_Seq      CCGGAAACAACAGAAATTTCGGGAACTAAAACCTGGGACGATAACGATGACCAAGATGGC
                   ************************************************************

Gallo1570_exp      AAACGGCCGACGGCGATTACAGTCAACTTGCTGGCTGATGGCGTTAAAGTAGATTCCAAG
Gallo1570_Seq      AAACGGCCGACGGCGATTACAGTCAACTTGCTGGCTGATGGCGTTAAAGTAGATTCCAAG
                   ************************************************************

Gallo1570_exp      AAAGTTACGGCAGCCGACGATTGGAAATATGAATTCAAAGACTTGCCGAAGTACAAGGCG
Gallo1570_Seq      AAAGTTACGGCAGCCGACGATTGGAAATATGAATTCAAAGACTTGCCGAAGTACAAGGCG
                   ************************************************************
```

```
Gallo1570_exp      GGTCAGGAAATCAAGTATTCTGTAACCGAAGAAGCCGTGAAAGACTATGAGACAAAAGTT
Gallo1570_Seq      GGTCAGGAAATCAAGTATTCTGTAACCGAAGAAGCCGTGAAAGACTATGAGACAAAAGTT
                   ************************************************************

Gallo1570_exp      TCCGGTACTGACATTACCAACATTCATACTCCGGAAACCACCGACATTACCGTTACGAAA
Gallo1570_Seq      TCCGGTACTGACATTACCAACATTCATACTCCGGAAACCACCGACATTACCGTTACGAAA
                   ************************************************************

Gallo1570_exp      ATCTGGGATGATCGCAACGATAAAGAAAAGAAACGCCCCGATAGTATCAAAGTCACCCTG
Gallo1570_Seq      ATCTGGGATGATCGCAACGATAAAGAAAAGAAACGCCCCGATAGTATCAAAGTCACCCTG
                   ************************************************************

Gallo1570_exp      AAAGCGAATGACAAAGATCTGCAAACCGTGACTATTACGGCGGAGGATGATTGGAAATAC
Gallo1570_Seq      AAAGCGAATGACAAAGATCTGCAAACCGTGACTATTACGGCGGAGGATGATTGGAAATAC
                   ************************************************************

Gallo1570_exp      GAGTTCAAAGATCTGCCCAAATACGAAAATGGCAAACAGATTAAGTATTCAGTCACTGAG
Gallo1570_Seq      GAGTTCAAAGATCTGCCCAAATACGAAAATGGCAAACAGATTAAGTATTCAGTCACTGAG
                   ************************************************************

Gallo1570_exp      GAAGAAGTTACGGGGTATACCACCACCATTGAAGAGGACGAGAGCGGCAACTTCGAAATT
Gallo1570_Seq      GAAGAAGTTACGGGGTATACCACCACCATTGAAGAGGACGAGAGCGGCAACTTCGAAATT
                   ************************************************************

Gallo1570_exp      ACCAATAAGATTCCACGTGACTACTTATTCGTCGACAAACCTCCCACACCTCCCCCTGAA
Gallo1570_Seq      ACCAATAAGATTCCACGTGACTACTTATTCGTCGACAAACCTCCCACACCTCCCCCTA--
                   ************************************************************

Gallo1570_exp      CCTGAAACATAAGCGGCCGCATC
Gallo1570_Seq      -----------------------
```

Gallo1675 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo1675_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCGCCGTGGTTCCGGATGGTACCGACGTACC
Gallo1675_Seq      --------------------GCGTGGATCCGCCGTGGTTCCGGATGGTACCGACGTACC
                                       ************************************

Gallo1675_exp      AGTCGTGGCGGAAGCAAGTCAGACGATCGTTGAACCAGCGAGCGATGAGTTAAACACAGC
Gallo1675_Seq      AGTCGTGGCGGAAGCAAGTCAGACGATCGTTGAACCAGCGAGCGATGAGTTAAACACAGC
                   ************************************************************

Gallo1675_exp      GATTAGCGATGCGGAAAATGCGGGTGTGACGGTATCTCAAACCACATCTGAAACTGTGGT
Gallo1675_Seq      GATTAGCGATGCGGAAAATGCGGGTGTGACGGTATCTCAAACCACATCTGAAACTGTGGT
                   ************************************************************

Gallo1675_exp      TAACCAGGAAGAAGCTCAAGCAGATTATGCCACCCAGGCAGAATCACTGGAAGCCGTGAC
Gallo1675_Seq      TAACCAGGAAGAAGCTCAAGCAGATTATGCCACCCAGGCAGAATCACTGGAAGCCGTGAC
                   ************************************************************

Gallo1675_exp      TGCCCAGCAGGAGCAGATTAATACGGAAAATGCGCAGATTACCGCCGATAATCAGGCTCT
Gallo1675_Seq      TGCCCAGCAGGAGCAGATTAATACGGAAAATGCGCAGATTACCGCCGATAATCAGGCTCT
                   ************************************************************

Gallo1675_exp      CAACGAAGCTTACGAATCGGCCAAAGCTCAGGCCGAATCCACTAACCAGGCAGTCTCGGA
Gallo1675_Seq      CAACGAAGCTTACGAATCGGCCAAAGCTCAGGCCGAATCCACTAACCAGGCAGTCTCGGA
                   ************************************************************

Gallo1675_exp      AGCCCAAAGCACGTATGGCGCCACGGTGACCGAAACAACGGTGGACTATGGAGATGGTAC
Gallo1675_Seq      AGCCCAAAGCACGTATGGCGCCACGGTGACCGAAACAACGGTGGACTATGGAGATGGTAC
                   ************************************************************

Gallo1675_exp      TCTGACCACTGACTATCAAGCGGGTCAGGCGCAGGCAGAGTCCATTGCTGAAGCTAACGA
Gallo1675_Seq      TCTGACCACTGACTATCAAGCGGGTCAGGCGCAGGCAGAGTCCATTGCTGAAGCTAACGA
                   ************************************************************
```

```
Gallo1675_exp      GCAGGCAGTCTCAGACTACCTGACGGAGAAAGCGGCAGTAGATGCGTATAACGCGCAAGT
Gallo1675_Seq      GCAGGCAGTCTCAGACTACCTGACGGAGAAAGCGGCAGTAGATGCGTATAACGCGCAAGT
                   ************************************************************

Gallo1675_exp      GAAAGCACGTGAGGATGCACTTAAGAGCAACAACATTGCATCGGATGAAGCGAACTACCT
Gallo1675_Seq      GAAAGCACGTGAGGATGCACTTAAGAGCAACAACATTGCATCGGATGAAGCGAACTACCT
                   ************************************************************

Gallo1675_exp      CTATGTAACTGGCGAGTTTGACACTAACGCGACCGGACTGGCTTACTACCAGAACATCAA
Gallo1675_Seq      CTATGTAACTGGCGAGTTTGACACTAACGCGACCGGACTGGCTTACTACCAGAACATCAA
                   ************************************************************

Gallo1675_exp      AGTAGTTACGCTTGACCCCAATGCGAAAACCGCCCAGTCTCTGGGGTGGCAGGATAACAC
Gallo1675_Seq      AGTAGTTACGCTTGACCCCAATGCGAAAACCGCCCAGTCTCTGGGGTGGCAGGATAACAC
                   ************************************************************

Gallo1675_exp      CACTATTAGCAACGCGAATGGCGTCACGGTAACGAGCCATGATACGGCCAATGACCCTGC
Gallo1675_Seq      CACTATTAGCAACGCGAATGGCGTCACGGTAACGAGCCATGATACGGCCAATGACCCTGC
                   ************************************************************

Gallo1675_exp      CATTTATGGCACCACCTCTGACTTCTTGTACAAAGTCACGGAAGCTACGGTGGGCGATAC
Gallo1675_Seq      CATTTATGGCACCACCTCTGACTTCTTGTACAAAGTCACGGAAGCTACGGTGGGCGATAC
                   ************************************************************

Gallo1675_exp      GTTCACGTTAAACAACATTGGCAAAGCCACCGACGGCACAAACATCAACGCTATCGTGAC
Gallo1675_Seq      GTTCACGTTAAACAACATTGGCAAAGCCACCGACGGCACAAACATCAACGCTATCGTGAC
                   ************************************************************

Gallo1675_exp      CATCACCAAAGCATCAGCGTTAACGGATAAGGAAGATAGCTGGTTCGTTATCGGGAAAAC
Gallo1675_Seq      CATCACCAAAGCATCAGCGTTAACGGATAAGGAAGATAGCTGGTTCGTTATCGGGAAAAC
                   ************************************************************

Gallo1675_exp      CGCGGATAACGGTATTGCCGTTGATTACTGGAACTATGACAATCTGGGCTTGAGCTTCCA
Gallo1675_Seq      CGCGGATAACGGTATTGCCGTTGATTACTGGAACTATGACAATCTGGGCTTGAGCTTCCA
                   ************************************************************

Gallo1675_exp      GTTTGTTGACGATTCGGGCAACGCTGTAAAACTGGTGGTCGCGAGTGTTGTCGGTGATGT
Gallo1675_Seq      GTTTGTTGACGA------------------------------------------------
                   ************

Gallo1675_exp      GGACAACGATCAGACGTCCAAGATTGAATTCGACGGGAATACTCTGAACTACGTGAATCC
Gallo1675_Seq      ------------------------------------------------------------


Gallo1675_exp      GGATGGGAGCGGTCTTATCGCCAATGCCGATAAATCACTGACCGGCCTGGGCTTTGCGGT
Gallo1675_Seq      ----------------------------ATAAATCACTGACCGGCCTGGGCTTTGCGGT
                                               ******************************

Gallo1675_exp      TGACGGTTACCAACAAGCGCCACAAGGTACCTATCTGATGGTGGGCTCTTCCACCACGGT
Gallo1675_Seq      TGACGGTTACCAACAAGCGCCACAAGGTACCTATCTGATGGTGGGCTCTTCCACCACGGT
                   ************************************************************

Gallo1675_exp      GAATTATACCCATACGAGTGACGATAATGTCGTGGACGGTAATGGCAATATCGTGAACTA
Gallo1675_Seq      GAATTATACCCATACGAGTGACGATAATGTCGTGGACGGTAATGGCAATATCGTGAACTA
                   ************************************************************

Gallo1675_exp      TATCGAATTCGACCTGTTTGGTACCACCAGTATGGTTACCACAGAAGAATTCAAGTACTT
Gallo1675_Seq      TATCGAATTCGACCTGTTTGGTACCACCAGTATGGTTACCACAGAAGAATTCAAGTACTT
                   ************************************************************

Gallo1675_exp      GCCCGATCCGACCTTAACCCTGACAAGTGTCACACTGCCGACTTCGCCTGTTGAGACACC
Gallo1675_Seq      GCCCGATCCGACCTTAACCCTGACAAGTGTCACACTGCCGACTTCGCCTGTTGAGACACC
                   ************************************************************

Gallo1675_exp      TCTGAAAGACAATTTGACCGCAACCTACCACCTCAATGAGTACGACGTAGCATTAACCAC
Gallo1675_Seq      TCTGAAAGACAATTTGACCGCAACCTACCACCTCAATGAGTACGACGTAGCATTAACCAC
                   ************************************************************
```

```
Gallo1675_exp      CGTTAAAGACGTACTGAATGATCAGGGTATCAGCATTGACGGTGGAGAGCTCCAAATTGG
Gallo1675_Seq      CGTTAAAGACGTACTGAATGATCAGGGTATCAGCATTGACGGTGGAGAGCTCCAAATTGG
                   ************************************************************

Gallo1675_exp      AGAGACAGGTCACTATACCCTGGAAGGTGCCAAAGTGCTGGCTAATGGAAAAGATACCTT
Gallo1675_Seq      AGAGACAGGTCACTATACCCTGGAAGGTGCCAAAGTGCTGGCTAATGGAAAAGATACCTT
                   ************************************************************

Gallo1675_exp      GGTCAAGTATGACTTCGAAGATTATCTGGATATCGAACATGATGAGTACCAGGGCTATTC
Gallo1675_Seq      GGTCAAGTATGACTTCGAAGATTATCTGGATATCGAACATGATGAGTACCAGGGCTATTC
                   ************************************************************

Gallo1675_exp      GATTTACGCGTTTGTACCGATTACGTTAAAAGATGGCACCGTGATCCAGTCTGGCGAAGA
Gallo1675_Seq      GATTTACGCGTTTGTACCGATTACGTTAAAAGATGGCACCGTGATCCAGTCTGGCGAAGA
                   ************************************************************

Gallo1675_exp      TCTGAAGGCATATGCGCAAGCGGTCTATGATGATGTAACTGGGCACTTTTATGTCAGCCT
Gallo1675_Seq      TCTGAAGGCATATGCGCAAGCGGTCTATGATGATGTAACTGGGCACTTTTATGTCAGCCT
                   ************************************************************

Gallo1675_exp      GAATAGCGATTTTCTTGCTCAGGTTGCGAAAGATTCCGATTTTCAGGCCAAAGTGGACAT
Gallo1675_Seq      GAATAGCGATTTTCTTGCTCAGGTTGCGAAAGATTCCGATTTTCAGGCCAAAGTGGACAT
                   ************************************************************

Gallo1675_exp      TGAATTTGTGCGCATTGCCGCAGGCGATGTCTATAACGACTTTACGAACCATCTGGCCTT
Gallo1675_Seq      TGAATTTGTGCGCATTGCCGCAGGCGATGTCTATAACGACTTTACGAACCATCTGGCCTT
                   ************************************************************

Gallo1675_exp      TGAGGATGAGGATGGGAACGTTACTGAAGTTCCGGTTCCGTCAAATGAAGTCGTGACTCA
Gallo1675_Seq      TGAGGATGAGGATGGGAACGTTACTGAAGTTCCGGTTCCGTCAAATGAAGTCGTGACTCA
                   ************************************************************

Gallo1675_exp      TACAGTGGAACCGCCGGTGGAAGAAGTTCCCGAAGAGCCGCAAGCGCCGACCGATGTGCA
Gallo1675_Seq      TACAGTGGAACCGCCGGTGGAAGAAGTTCCCGAAGAGCCGCAAGCGCCGACCGATGTGCA
                   ************************************************************

Gallo1675_exp      AACCCCGGAAGTCGCGGAGGATGTGCCAGTGGTTTCCCAGAGTGTTGTCGACAAACCTCC
Gallo1675_Seq      AACCCCGGAAGTCGCGGAGGATGTGCCAGTGGTTTCCCAGAGTGTTGTCGACAAACCTCC
                   ************************************************************

Gallo1675_exp      CACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATC
Gallo1675_Seq      CACACCTCCCCC-------------------------
                   ************
```

Gallo2018 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo2018_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCGACGATGAACTGGTTCCAACGACAGAAAC
Gallo2018_Seq      --------------------------TCCGACGATGAACTGGTTCCAACGACAGAAAC
                                             ******************************

Gallo2018_exp      CACCGAAGTAGTTGATAACGGGGATAACGTGACCAAGAATCTTGCGACTGACATCATTGA
Gallo2018_Seq      CACCGAAGTAGTTGATAACGGGGATAACGTGACCAAGAATCTTGCGACTGACATCATTGA
                   ************************************************************

Gallo2018_exp      ACCGTCCAATGATATCTCCGAATCTCAAAGCGAGAAAACCGAAGAGGAGTCGTCAATCGA
Gallo2018_Seq      ACCGTCCAATGATATCTCCGAATCTCAAAGCGAGAAAACCGAAGAGGAGTCGTCAATCGA
                   ************************************************************

Gallo2018_exp      AACTGCCGATAACAGTTCCGTGATTATGGAGAGCACCGAAGCGACTGAAACGATTGCGAG
Gallo2018_Seq      AACTGCCGATAACAGTTCCGTGATTATGGAGAGCACCGAAGCGACTGAAACGATTGCGAG
                   ************************************************************

Gallo2018_exp      TGACACATCGGATGAACCGGAAGAAGCGGAGGTAACGATCCCGCAGTATGAAGAGAATGT
Gallo2018_Seq      TGACACATCGGATGAACCGGAAGAAGCGGAGGTAACGATCCCGCAGTATGAAGAGAATGT
                   ************************************************************
```

```
Gallo2018_exp      TGCCGACTTTAACCATGTCCCGATGACCGATGTCTACGTGATGTTCACCGAGGATGGCAA
Gallo2018_Seq      TGCCGACTTTAACCATGTCCCGATGACCGATGTCTACGTGATGTTCACCGAGGATGGCAA
                   ************************************************************

Gallo2018_exp      AGAACACGTTATCTATGTAGGTCGTCCAACGTGCTATTATTGTCGCCAGTTTAGTCCTGC
Gallo2018_Seq      AGAACACGTTATCTATGTAGGTCGTCCAACGTGCTATTATTGTCGCCAGTTTAGTCCTGC
                   ************************************************************

Gallo2018_exp      GTTGAAAGAGTTCAATACGCTGATGGACAATCGCCTCGAATACTACAATACCGATTCACA
Gallo2018_Seq      GTTGAAAGAGTTCAATACGCTGATGGACAATCGCCTCGAATACTACAATACCGATTCACA
                   ************************************************************

Gallo2018_exp      GGACTTTGATGAAGAAGCAGCGAACTTCCTGTTTGGCACAATTGGCATTCCTGGAACACC
Gallo2018_Seq      GGACTTTGATGAAGAAGCAGCGAACTTCCTGTTTGGCACAATTGGCATTCCTGGAACACC
                   ************************************************************

Gallo2018_exp      GACGATTATTCGCTTACAGAATGGCCAAATTGTGTCTGCGTGGATTGGAGGTGGCATCTC
Gallo2018_Seq      GACGATTATTCGCTTACAGAATGGCCAAATTGTGTCTGCGTGGATTGGAGGTGGCATCTC
                   ************************************************************

Gallo2018_exp      TGGTCAGGAGCTGTATGACTACCTGTTCTATGGGAAAATTCCCGTGGCCATGGCTGCAGC
Gallo2018_Seq      TGGTCAGGAGCTGTATGACTACCTGTTCTATGGGAAAATTCCCGTGGCCATGGCTGCAGC
                   ************************************************************

Gallo2018_exp      AATGGCGGAACAGAGCAATGAAGATAACACTGAAACCATTGCCTTTGACGCCAAAGAGAT
Gallo2018_Seq      AATGGCGGAACAGAGCAATGAAGATAACACTGAAACCATTGCCTTTGACGCCAAAGAGAT
                   ************************************************************

Gallo2018_exp      CAAGACCGATAGCAACATCCAGAATGTCGTCTTTCTGCCGCAAAACGATGTGAAAACGGC
Gallo2018_Seq      CAAGACCGATAGCAACATCCAGAATGTCGTCTTTCTGCCGCAAAACGATGTGAAAACGGC
                   ************************************************************

Gallo2018_exp      AGAACGTGCACTGATTGTGCCCGAAAGCCCACAAGCTTTCAGCAAGAACGAAACCAAAAC
Gallo2018_Seq      AGAACGTGCACTGATTGTGCCCGAAAGCCCACAAGCTTTCAGCAAGAACGAAACCAAAAC
                   ************************************************************

Gallo2018_exp      CAACAATTCGAACGCTTTACCGAAATTGGGTATCAAAGCCAACAACGTCGACAAACCTCC
Gallo2018_Seq      CAACAATTCGAACGCTTTACCGAAATTGGGTATCAAAGCCAACAACGTCGACAAACCTCC
                   ************************************************************

Gallo2018_exp      CACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATC
Gallo2018_Seq      CACACCTCCCCC-------------------------
                   ************
```

Gallo2178 (sequencing of PCR product via T3 and T7 primer (PCR primer pGEXs+T7 and pGEXas+T3))

```
Gallo2178_exp      TCCAAAATCGGATCTGGTTCCGCGTGGATCCTATGATATTACCGTGGAAAATGGCGGAAG
Gallo2178_Seq      ------------------TCGCGTGGATCCTATGATATTACCGTGGAAAATGGCGGAAG
                                     ************************************

Gallo2178_exp      TGGTACCTACGAGAGCTATCAGATCTTTACTGGGACTTTAAGCGAGGATGGCAAAACCCT
Gallo2178_Seq      TGGTACCTACGAGAGCTATCAGATCTTTACTGGGACTTTAAGCGAGGATGGCAAAACCCT
                   ************************************************************

Gallo2178_exp      GTCCAATATCGAATGGGGTAACGGCATTACGACGGCAGGCCAAACGGCATTACAGGAGAA
Gallo2178_Seq      GTCCAATATCGAATGGGGTAACGGCATTACGACGGCAGGCCAAACGGCATTACAGGAGAA
                   ************************************************************

Gallo2178_exp      ATATGGTGTCAGTTCAGCCGCGGGTCTGGCCGAAGTTTTGGGCGCTGACGATTTTACTGC
Gallo2178_Seq      ATATGGTGTCAGTTCAGCCGCGGGTCTGGCCGAAGTTTTGGGCGCTGACGATTTTACTGC
                   ************************************************************

Gallo2178_exp      GAGTCAAGCCGAGGAATTCGCGAAAGTAGTTGGGCAGTATCTTCAGAATGCGGGTGGTTT
Gallo2178_Seq      GAGTCAAGCCGAGGAATTCGCGAAAGTAGTTGGGCAGTATCTTCAGAATGCGGGTGGTTT
                   ************************************************************
```

```
Gallo2178_exp      GACCGGATTAGCTGCGGGGTATTACCTGGTCCAGAATGCCTCAGTGGGCAATAACGAAGC
Gallo2178_Seq      GACCGGATTAGCTGCGGGGTATTACCTGGTCCAGAATGCCTCAGTGGGCAATAACGAAGC
                   ************************************************************

Gallo2178_exp      GCATACCAACTATATTCTCCAGGTGGTGAAAGACGTTATTGTGGAACCCAAGACAAGTGT
Gallo2178_Seq      GCATACCAACTATATTCTCCAGGTGGTGAAAGACGTTATTGTGGAACCCAAGACAAGTGT
                   ************************************************************

Gallo2178_exp      ACCAACGGTAGAAAAGAAACTGAAAGACACGAACGATACGACGGGCGAAACGACCGATTG
Gallo2178_Seq      ACCAACGGTAGAAAAGAAACTGAAAGACACGAACGATACGACGGGCGAAACGACCGATTG
                   ************************************************************

Gallo2178_exp      GCAGGATAGCGCCGACTACGATATTAACGATTCAGTGCCTTTCCAACTCACCGCAACTCT
Gallo2178_Seq      GCAGGATAGCGCCGACTACGATATTAACGATTCAGTGCCTTTCCAACTCACCGCAACTCT
                   ************************************************************

Gallo2178_exp      TCCGGATAATCTGGCTTCTTACGACGAATACTATCTGGAGCTGAGTGACACCTTGTCGGC
Gallo2178_Seq      TCCGGATAATCTGGCTTCTTACGACGAATACTATCTGGAGCTGAGTGACACCTTGTCGGC
                   ************************************************************

Gallo2178_exp      TGGTTTGACGTACAACAAAGACGCCAAAGTCTATCTCGTTAATGGCACCACCAAAACCGA
Gallo2178_Seq      TGGTTTGACGTACAACAAAGACGCCAAAGTCTATCTCGTTAATGGCACCACCAAAACCGA
                   ************************************************************

Gallo2178_exp      TGTTACCTCGAGTTTCACCATTGCAGATGATGGCTCGTCTTTCAAAATCAACAACCTGAA
Gallo2178_Seq      TGTTACCTCGAGTTTCACCATTGCAGATGATGGCTCGTCTTTCAAAATCAACAACCTGAA
                   ************************************************************

Gallo2178_exp      AAGCTTAGATGGGGTTACCAGCAGCACCAAAGTTGTGGTCGAGTATACTGCCACACTGAA
Gallo2178_Seq      AAGCTTAGATGGGGTTACCAGCAGCACCAAAGTTGTGGTCGAGTATACTGCCACACTGAA
                   ************************************************************

Gallo2178_exp      CTCTAATGCAGTAATTGGCCTGGAAGGGAACCCGAACACAGTGAAACTGATCTATTCCAA
Gallo2178_Seq      CTCTAATGCAGTAATTGGCCTGGAAGGGAACCCGAACACAGTGAAACTGATCTATTCCAA
                   ************************************************************

Gallo2178_exp      CAACCCGAATTATACAGGTTCCGGCGAAACGTCGCCAACAGGCGAAACACCGGAGGACAA
Gallo2178_Seq      CAACCCGAATTATACAGGTTCCGGCGAAACGTCGCCAACAGGCGAAACACCGGAGGACAA
                   ************************************************************

Gallo2178_exp      AGTCATCGTGTTCACCTACAAAGTAGTGGTAAACAAAGTGGATCAATCCGGCAATGCGCT
Gallo2178_Seq      AGTCATCGTGTTCACCTACAAAGTAGTGGTAAACAAAGTGGATCAATCCGGCAATGCGCT
                   ************************************************************

Gallo2178_exp      TGCAGGAGCCGGTTTTACGCTGTACAAGAAAGATTCCTCTGGCAATTGGAACGCGGTTAG
Gallo2178_Seq      TGCAGGAGCCGGTTTTACGCTGTACAAGAAAGATTCCTCTGGCAATTGGAACGCGGTTAG
                   ************************************************************

Gallo2178_exp      CGACGAAATTACTGGTGTCACCACCTTTACCTTTTCCGGCCTGGATGATGGAGATTACAA
Gallo2178_Seq      CGACGAAATTACTGGTGTCACCACCTTTACCTTTTCCGGCCTGGATGATGGAGATTACAA
                   ************************************************************

Gallo2178_exp      GCTGTCTGAAACCACTACCCCGAATGGGTATAATACCATTGACGATATCACCTTTACGGT
Gallo2178_Seq      GCTGTCTGAAACCACTACCCCGAATGGGTATAATACCATTGACGATATCACCTTTACGGT
                   ************************************************************

Gallo2178_exp      CACTGCGGATCACGACGTGAAAAGCGATTCACCGGCGCTGAATAGCCTGAGCGGTGACGT
Gallo2178_Seq      CACTGCGGATCACGACGTGAAAAGCGATTCACCGGCGCTGAATAGCCTGAGCGGTGACGT
                   ************************************************************

Gallo2178_exp      GACCACCGGTAGCCTGACGTTTGCGTCGAACATCACGGAAGATGACGCATCGCTCACTAC
Gallo2178_Seq      GACCACCGGTAGCCTGACGTTTGCGTCGAACATCACGGAAGATGACGCATCGCTCACTAC
                   ************************************************************

Gallo2178_exp      GAACGTTGTCAACAAGAAGGGTGCTACTCTGCCTTCAACAGTCGACAAACCTCCCACACC
Gallo2178_Seq      GAACGTTGTCAACAAGAAGGGTGCTACTCTGCCTTCAACAGTCGACAAACCTCCCACACC
                   ************************************************************
```

```
Gallo2178_exp      TCCCCCTGAACCTGAAACATAAGCGGCCGCATC
Gallo2178_Seq      TCCCCCT-------------------------
                   *******
```

Gallo2179 (sequencing of miniPrep DNA via primer pGEXfor and pGEXrev)

```
Gallo2179_exp      ATAGCATGGCCTTTGCAGGGCTGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAA
Gallo2179_Seq      ----------------------------------------------------TCCTCCAAA
                                                                        *  * ***

Gallo2179_exp      TCGGATCTGGTTCCGCGTGGATCCGCCGATGTATCTAACCGGGTAACCTCACTTACAGTG
Gallo2179_Seq      TCGGATCTGGTTCCGCGTGGATCCGCCGATGTATCTAACCGGGTAACCTCACTTACAGTG
                   ************************************************************

Gallo2179_exp      GCCACAACGGAACTCCAAGACGGTGGTCGTACCACTGTACGTGTCGAGTTTAACGATCGT
Gallo2179_Seq      GCCACAACGGAACTCCAAGACGGTGGTCGTACCACTGTACGTGTCGAGTTTAACGATCGT
                   ************************************************************

Gallo2179_exp      GCAGGCAAAATTCATAGCGGCGATACGATCGAAGTCACCTGGAGTATCTCAAACAGCATT
Gallo2179_Seq      GCAGGCAAAATTCATAGCGGCGATACGATCGAAGTCACCTGGAGTATCTCAAACAGCATT
                   ************************************************************

Gallo2179_exp      TATCTGAACGGTTACACAAAATCGATTCCTCTGACCATCCAGGGTGTGAACGTTGGGACG
Gallo2179_Seq      TATCTGAACGGTTACACAAAATCGATTCCTCTGACCATCCAGGGTGTGAACGTTGGGACG
                   ************************************************************

Gallo2179_exp      TTAGAAGTCACCGAACATAACGCGATCTTCAAATTCAACTCCAATATTGAAACGATGGAA
Gallo2179_Seq      TTAGAAGTCACCGAACATAACGCGATCTTCAAATTCAACTCCAATATTGAAACGATGGAA
                   ************************************************************

Gallo2179_exp      AATGTTTCTGGCTGGGGTGAGTTTGAAGTAATTGGCCGCAATGTGACGAATACTAGCAGC
Gallo2179_Seq      AATGTTTCTGGCTGGGGTGAGTTTGAAGTAATTGGCCGCAATGTGACGAATACTAGCAGC
                   ************************************************************

Gallo2179_exp      GAGAATACGGGAACCGCCGTGGTGCAAGTGGGCGGCTACTCTCAGAACATCTCAATCACT
Gallo2179_Seq      GAGAATACGGGAACCGCCGTGGTGCAAGTGGGCGGCTACTCTCAGAACATCTCAATCACT
                   ************************************************************

Gallo2179_exp      AAACCCCAAAGTGGGACGGGCACCTCAAGCTTCTACTATAAAACTGGGGATATTCAGCCG
Gallo2179_Seq      AA------------------------------------------------------------
                   **

Gallo2179_exp      TCAGATACCAATCGGGTTCGCTGGTTTCTGTTGGTGAATAATAACAAAGAGTATGTCGAA
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      AGTGATGTGACGATCGAAGATGACATCCAAAGCGGGCAAACCCTGGATATGTCCTCGTTC
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      GATATCACCATTTCAGGGTATCAGAACAAGCGCTTTGTTGGCGAATCTGCACTCGAGGAA
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      TTCAAACGTTCGTGCCCAAATTCTAGCATCGAAATTACGCAGAAGAGCGAAGGTGGTCAC
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      ATCTCCATTCGCCTGAGCCGCGACGATGTCATCTTGAACACCATTTCGATCCACTACAAA
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      ACGAAAATTCTGGACTTTGATCAGGAGAAATTTGCGAATAATAGTAATATTACCTACAAA
Gallo2179_Seq      ------------------------------------------------------------
```

```
Gallo2179_exp      CCCTTGTATAAAGACTGGGTAACTAACAAAGAAAGCAACTATGAAGTGGTCAATGTTAAC
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      GCTAACGGTGGCGTCGATGGTTCCCGCTATACGTCGGTTACAGTTAACAAGGTGTGGAAT
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      GATAAAGACAACCAAGATGGCAAACGCTCTGACAAAGTGGTGATTCAGCTTTTGGCGGAT
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      GGTCAGGAGATCAGCGGTAAACAGCTTGAGCTGAGCGAAGAAAACGGTTGGAGTGGTACC
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      TTTGAGAAGCTGAACAAATATCACTCGGATAATACGCTGATTACCTATACTGTGAAAGAA
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      GTCACTGATTTACCGGACTATCAGACGACCGTTTCTGAAAACTCGAAGAACAACTACACC
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      ATTACCAATACCCACATTCCTGAAGTGATTGACCTCTCGGGCAAGAAAATCTGGGATGAC
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      AATAATAATCAAGATGGAATTCGCCCAGAAACCATTACCGTTCATCTGTTAGCTAACGGC
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      GTTGATACCGGACAGGTGAAAACGGTGTCCAAAAGCGACAACTGGGAATACCAGTTTAAA
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      GATCTGCCGAAGTATCAGAATGGCGAAAAGGTTGTGTACACCGTAAGTGAAGATGTTGTA
Gallo2179_Seq      ------------------------------------------------------------

Gallo2179_exp      GTGGGGTATGAGATGAGTGTGTCTGGCATGAACCTGACTAATACCCATACACCAGAAGTC
Gallo2179_Seq      ----------------------TGGCATGAACCTGACTAATACCCATACACCAGAAGTC
                                          **********************************

Gallo2179_exp      ACGAATATCCTGATTAGCAAATATTGGGATGACAACGACGACAAGCTGAAGAAACGTCCG
Gallo2179_Seq      ACGAATATCCTGATTAGCAAATATTGGGATGACAACGACGACAAGCTGAAGAAACGTCCG
                   ************************************************************

Gallo2179_exp      GAAAGCATTCAAATCACGCTGCATGCCAACGGAAAAGAGTACCAGACTGTAACCTTAACT
Gallo2179_Seq      GAAAGCATTCAAATCACGCTGCATGCCAACGGAAAAGAGTACCAGACTGTAACCTTAACT
                   ************************************************************

Gallo2179_exp      GCATCCAATCAGTGGCAATATGAGTTCAAAGACCTCCCGAAATACAAAGATGGTGAGAAA
Gallo2179_Seq      GCATCCAATCAGTGGCAATATGAGTTCAAAGACCTCCCGAAATACAAAGATGGTGAGAAA
                   ************************************************************

Gallo2179_exp      ATCGCGTACACAGTCACAGAAGCGGATGTTCCGAACTATCAGCTGATTTCCATTGAAGAA
Gallo2179_Seq      ATCGCGTACACAGTCACAGAAGCGGATGTTCCGAACTATCAGCTGATTTCCATTGAAGAA
                   ************************************************************

Gallo2179_exp      GATGAGTCCGGCAACTGGAAAATTACCAACAAAGTCGAAGAAAGTTACCTGTTTCCGAAT
Gallo2179_Seq      GATGAGTCCGGCAACTGGAAAATTACCAACAAAGTCGAAGAAAGTTACCTGTTTCCGAAT
                   ************************************************************

Gallo2179_exp      ACCGGCGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATCG
Gallo2179_Seq      ACCGGCGTCGACAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAGCGGCCGCATCG
                   ************************************************************
```

```
Gallo2179_exp      TGACTGACTGACGATCTGCCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACAT
Gallo2179_Seq      TGACTGACTGACGATCTGCCTCG-----------------------------------
                   **********************

Gallo2179_exp      GCAGCTC
Gallo2179_Seq      -------
```

III. Alignment of *S. gallolyticus* proteins Gallo0577, Gallo1570 and Gallo2179

```
Gallo2179      --------------ADVSNRVTSLTVATTELQDGGRTTVRVEFNDRA-------------
Gallo0577      ----------DTVDITVS----NTSLSTNAINGGTSTEFSFDFAVPNSAKSGDTTVISLP
Gallo1570      KAEEDVYYTGYTSDISLNSSYINPDPGPYAIDEGGESKLAYCFNRNKSRPPA----KSEP
                                :.    .   .   :: *  : .   *

Gallo2179      ------------------------------------------------GKIHSGDT---
Gallo0577      DELNFQRNQTFNVYASDGTTVVATAVIDTTTKTLTLTYTDYVDTHDDVTGHLSMNVVVDR
Gallo1570      -------------------------EDGEAKYRKIADVDYVRLKENCSSDMEGRELYDA
                                                                  ..:

Gallo2179      --IEVTWSISNSIYLNGYTKSIPLTIQGVNV-----------------------------
Gallo0577      TVVTEATTVPATVTINGTTTI-TISSGGINYTVSTGDSDDIDFWKYGVSYSDDEVMYLIN
Gallo1570      IM---------KVIYNGYP-------------------------------------
                          .:  **

Gallo2179      -GTLEVTEHNAIFKFNSNIETMENVSGWGEFEVIGRNVTNTSSENTGTAVVQVGGYSQNI
Gallo0577      VNTSAATVSNVVISDTINSTGLEYVD--GSFEIFEGTWYK----NA-QNYWALGGSTNVT
Gallo1570      ------------------NNCSGIN--GKYRLKDGDFCA----ITQWAIWHFTDGADSD
                                . :.  *.:.:           :      . . ::

Gallo2179      SI-TKPQSGTGTSSFYYKTGDIQPS-----DTNRVR--------WFLLVNN-----NKEY
Gallo0577      SNYNIELSA-DNTSFSINLGTISKGYMI-------RYRVKANYTL---INGEQLSNSATY
Gallo1570      GTGNLPYYG-KESMW--NRSDVKEAYLELIDVANLSYPADAKLNLYIYDHGAEHDR-QNL
               .  .    .   : :  : . :. .                        :.

Gallo2179      VESDVTIEDDIQSGQTLDMSSFDITISGYQNKRFVGESALEEFKRSCPNSSIEITQ----
Gallo0577      YSENTALNNA-DN--TFTYQGASGTASGYNYSLTV-QK-VNEAGEALAGAEFTVTRE---
Gallo1570      LTTDVGYTNL-SV--EKVWN-DSDDQDGIRPAFID-VQ-LLANGVEVEGQKIELSKFLNS
                  :.   :  .       .  .   .* .       . :        . .: :::

Gallo2179      KSEGGHISIRLSRDDVILNTISIHYKTKILDFDQEKFANNSNITYKPLYKDWVTNKESNY
Gallo0577      -STGQVVGTITTGSDG----TAT-------------ISG----LLKDNYIITETKAPTGY
Gallo1570      NWQGVFRGLSLYDSDG----NPIEYSVKEVE----KYRG----Q-LDGYQSTVTKSDSGY
                  *   .     .*                       .         *    *:  :.*

Gallo2179      EVVNVNA-------NGGVDGSRYTSVTVNKVWNDKDNQDGKRSDKVVIQLLADGQEISGK
Gallo0577      AIADPVTAEADNSTVTVTDKKATVEVTGTKTWDDNNDQDGKRPDSITVNLLANGTVVDTK
Gallo1570      S-----------YTITNTHVPETTEISGTKTWDDKDDQDGKRPSSITVKLLADDEEIDSQ
                              ..    ..:: .*.*:*:::***** ..:.::***:.  :. :

Gallo2179      QLELSEENGWSGTFEKLNKYHSDNTLITYTVKEVTDLPDYQTTVSENSKNNYTITNTHIP
Gallo0577      T--VTADDNWTYAFSDLDQYDADGNEIAYTVSEEM-VDGYTTV-----VDGYNITNTHAS
Gallo1570      E--VTADTDWKYSFKDLPKYKNEGVEINYSVAEES-VSDYETT-----ISGTDITNTHVP
                  :: : .*. :*..* :*. :.  * *:* *   : .* *.      ..  *****

Gallo2179      EVIDLSGKKIWDDNNNQDGIRPETITVHLLANGVDTGQVKTVSKSDNWEYQFKDLPKYQN
Gallo0577      ETTEVSGTKTWDDNDDQDGKRPDSITVNLLANGTVVD-TKTVTADDNWSYSFTDLPKYDN
Gallo1570      ETTEISGTKTWDDNDDQDGKRPTAITVNLLADGVKVD-SKKVTAADDWKYEFKDLPKYKA
               *. ::**.* ****::*** ** :***:***:*. ..  *.*:  *:*.*.*.*****.

Gallo2179      GEKVVYTVSEDVVVGYEMSVSGMNLTNTHTPEVTNILISKYWDDNDDKLKKRPESIQITL
Gallo0577      GNEITYTVTEDTVADYTTTYDGYNITNSYTPGETSITVTKVWDDNNDQDGIRPDAIQVQL
Gallo1570      GQEIKYSVTEEAVKDYETKVSGTDITNIHTPETTDITVTKIWDDRNDKEKKRPDSIKVTL
               *::: *:*:*:.* .*  . .* ::** :**  *.* ::* ***.:*:   **::*:: *

Gallo2179      HANGKE-YQTVTLTASNQWQYEFKDLPKYKDGEKIAYTVTEADVPNYQLISIEEDESGNW
Gallo0577      YANGEKSGDVITLTVADNWTYTWTGLAEKANKKTITYTVEEVSAVDGYTATVGEVENGNV
Gallo1570      KANDKD-LQTVTITAEDDWKYEFKDLPKYENGKQIKYSVTEE-EVTGYTTTIEEDESGNF
                **.:.  :.:*:*. ::* * :..* :  : : * *:* *         :: * *.**

Gallo2179      KITNKVEESYLFPNTG------------------------
Gallo0577      TITNTHTPTTPETPSSDEPTTPSQSNKKSDKEQDKNIIAAL
Gallo1570      EITNKIPRDYLF----------------------------
                ***.
```

IV. Alignment of *F. nucleatum* proteins Fn0387, Fn1449 and Fn1893

```
Fn0387      SKDSNKIKAFGARGEYKTNTAGVIDYKNYAYGVAYIHENESVKLGKDIGWYTGFVHNTFR
Fn1449      SKDSNKVKTFGMKGEYKTDTAGVIDYKYNAYGVAYVHENEDIKLGKGTGWYTGIVHNTFK
Fn1893      SKDSNKIKIFGIKGEYKTDTAGVIDYKNEAYGMAYVHENEDIKLGKGIGWYTGIVDNTFK
            ******:* ** :*****:********  ***:**:****.:****. *****:*.***:

Fn0387      FEDIGKSKEEMLLGKIGMFKSIPFDDDNSLNWTVSGNVFVGRNKMHRKFLIVDEIFNAKS
Fn1449      FKDIGNSKEKQLQAKVGLFKSVPFDENNSLNWTISGDIFIGHNKLERKFLVVDEIFHAKS
Fn1893      FKDIGKSKEEQIQAKVGLLKSIPFDDNNSLNWTISGDIFVGYNKMHRKYLVVNEIFNAKS
            *:***:***: : .*:*::**:***::******:**::*:* **:.**:*:*:***:***

Fn0387      KYYAYGIGVKNEIGKEFRLSEDFSIRPYGALKLEYGRISKIKEKTGEIRLEVKSNDYVSI
Fn1449      KYYTYGIGIKNEIGKEFRLSEDFSIRPYGALKVEYGRVSKIKEKSGEMKLEVKENDYLSI
Fn1893      KYYTYGIGIKNKISKDFRLSEDFSLVPYGSLNLEYGRVNKIKEKVGEIRLEVKENYYVSV
            ***:****:**:*.*:********: ***:*::****:.***** **::****.* *:*:

Fn0387      KPEIGTELKYKYLFTNRKTLTVGLGVAYENELGKVANPKNKARVAYTAADWYNLRGEKED
Fn1449      RPEIGTELAYRHYFGT-KTLRTSVGVAYENELGRVANGKNKARVAGTTADWFNIRGEKED
Fn1893      NPEIGAELTYKHLLASRKTFRMGLGIAYENELGKVANGKNKARVAYTNADWFNIRGEKED
            .****:** *:: : . **:  .:*:*******:*** ******* * ***:*:******

Fn0387      RRGNIKTDLTIGLENTRFGATANVGYDTKGHNV
Fn1449      RKGNVKVDLNVGIDNQRLGVTGNVGYDTKGHNV
Fn1893      RKGNIKFDLNIGLDNQRVGVTANAGYDTKGHNV
            *:**:* **.:*::* *.*.*.*.*********
```

V. Different models for estimating the association of antibody responses to *F. nucleatum* with CRC in BliTz/DACHSplus, SCCS and EPIC

Supplementary table 1: Antibody responses to *F. nucleatum* proteins in relation to prevalence of CRC in the DACHSplus study compared to BliTz controls

	Positive n (%)		Unadjusted model[1]			Adjusted model 1[2]			Adjusted model 2[3]		
	Controls n=228	Cases n=318	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Fn0131	22 (10)	36 (11)	1.20	0.68-2.09	0.532	1.14	0.64-2.04	0.658	1.23	0.67-2.26	0.504
Fn0253	22 (10)	33 (10)	1.08	0.61-1.91	0.782	0.85	0.47-1.54	0.592	0.90	0.49-1.64	0.723
Fn0264	22 (10)	32 (10)	1.05	0.59-1.86	0.874	0.95	0.53-1.73	0.873	0.90	0.49-1.64	0.719
Fn0387	22 (10)	31 (10)	1.01	0.57-1.80	0.969	0.95	0.52-1.74	0.878	0.91	0.50-1.68	0.770
Fn1426	22 (10)	25 (8)	0.80	0.44-1.46	0.463	0.70	0.37-1.31	0.262	0.73	0.38-1.40	0.345
Fn1449	22 (10)	27 (8)	0.87	0.48-1.57	0.641	0.82	0.45-1.53	0.539	0.77	0.41-1.43	0.404
Fn1526	19 (8)	35 (11)	1.36	0.76-2.45	0.304	1.28	0.70-2.36	0.420	1.43	0.76-2.70	0.271
Fn1817_1	22 (10)	26 (8)	0.83	0.46-1.51	0.549	0.86	0.46-1.61	0.643	0.94	0.50-1.77	0.845
Fn1817_2	22 (10)	25 (8)	0.80	0.44-1.46	0.463	0.69	0.37-1.29	0.240	0.76	0.40-1.44	0.394
Fn1859	22 (10)	21 (7)	0.66	0.36-1.24	0.196	0.58	0.30-1.11	0.100	0.53	0.27-1.04	0.064
Fn1893	22 (10)	31 (10)	1.01	0.57-1.80	0.969	0.94	0.52-1.72	0.844	0.92	0.49-1.71	0.785
Any *F. nucleatum* protein	129 (57)	175 (55)	0.94	0.67-1.32	0.720	0.81	0.56-1.16	0.254	0.86	0.59-1.25	0.436

[1]Logistic regression model without further adjustment; [2]Logistic regression model with adjustment for age (continuous variable) and sex; [3]Logistic regression model with adjustment for age (continuous variable), sex, BMI, education and smoking with exclusion of samples with missing baseline information in any of these variables (n(controls)=218, n(CRC)=303); Significant associations are marked in bold font

Supplementary table 2: Antibody responses to *F. nculeatum* proteins in relation to CRC risk in a nested case-control study within SCCS

	Positive n (%)		Unadjusted model[1]			Adjusted model 1[2]			Adjusted model 2[3]		
	Controls n=348	Cases n=181	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Fn0131	34 (10)	15 (8)	0.82	0.42-1.58	0.551	0.89	0.45-1.77	0.741	0.86	0.44-1.70	0.664
Fn0253	34 (10)	15 (8)	0.81	0.42-1.55	0.520	0.89	0.45-1.78	0.743	0.92	0.47-1.82	0.819
Fn0264	34 (10)	18 (10)	1.05	0.57-1.92	0.877	0.86	0.45-1.62	0.636	0.92	0.49-1.71	0.788
Fn0387	34 (10)	17 (9)	0.97	0.53-1.77	0.919	1.04	0.55-1.95	0.913	1.06	0.56-1.98	0.864
Fn1426	34 (10)	30 (17)	**1.92**	**1.11-3.31**	**0.019**	**1.85**	**1.04-3.28**	**0.037**	**1.85**	**1.04-3.29**	**0.035**
Fn1449	34 (10)	12 (7)	0.67	0.34-1.31	0.240	0.69	0.34-1.43	0.323	0.77	0.38-1.55	0.457
Fn1526	34 (10)	15 (8)	0.80	0.41-1.56	0.510	0.80	0.40-1.61	0.534	0.82	0.41-1.64	0.579
Fn1817_1	34 (10)	19 (11)	1.09	0.61-1.96	0.763	1.21	0.65-2.24	0.554	1.13	0.61-2.08	0.700
Fn1817_2	34 (10)	16 (9)	0.91	0.48-1.75	0.785	1.06	0.54-2.05	0.872	1.02	0.53-1.97	0.951
Fn1859	34 (10)	12 (7)	0.64	0.32-1.28	0.209	0.65	0.32-1.32	0.234	0.65	0.32-1.31	0.227
Fn1893	34 (10)	19 (11)	1.11	0.62-2.01	0.721	1.19	0.64-2.20	0.586	1.21	0.66-2.23	0.545
Any *F. nucleatum* protein	205 (59)	99 (55)	0.85	0.59-1.21	0.358	0.85	0.57-1.25	0.398	0.87	0.59-1.27	0.460

[1]Conditional logistic regression model; [2]Conditional logistic regression model with adjustment for BMI, education and smoking status as categorical variables, missings in the variables are excluded from the analyses; [3]Conditional logistic regression model with adjustment for BMI, education and smoking status as categorical variables, missings in the variables are considered as individual category; Significant associations are marked in bold font

Supplementary table 3: Antibody responses to *F. nucleatum* proteins in relation to CRC risk in a nested case-control study within SCCS in cases diagnosed after more than two years from blood draw

	Positive n (%)		Unadjusted model[1]			Adjusted model 1[2]			Adjusted model 2[3]		
	Controls n=239	Cases n=124	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Fn0131	18 (8)	12 (10)	1.31	0.60-2.85	0.499	1.40	0.61-3.21	0.426	1.39	0.61-3.17	0.431
Fn0253	24 (10)	11 (9)	0.84	0.39-1.81	0.658	0.85	0.37-1.95	0.699	0.88	0.39-1.99	0.758
Fn0264	26 (11)	8 (6)	0.58	0.25-1.31	0.188	0.56	0.24-1.30	0.177	0.56	0.24-1.30	0.177
Fn0387	25 (10)	14 (11)	1.08	0.55-2.13	0.817	1.11	0.53-2.30	0.787	1.12	0.55-2.30	0.756
Fn1426	27 (11)	22 (18)	1.76	0.94-3.28	0.076	1.70	0.88-3.29	0.114	1.71	0.89-3.30	0.109
Fn1449	27 (11)	11 (9)	0.77	0.37-1.59	0.476	0.73	0.33-1.63	0.446	0.83	0.39-1.80	0.641
Fn1526	24 (10)	11 (9)	0.83	0.37-1.84	0.641	0.80	0.34-1.89	0.616	0.84	0.36-1.94	0.681
Fn1817_1	27 (11)	11 (9)	0.78	0.37-1.64	0.507	0.82	0.37-1.82	0.626	0.74	0.34-1.64	0.463
Fn1817_2	21 (9)	7 (6)	0.62	0.24-1.56	0.306	0.68	0.26-1.76	0.424	0.67	0.26-1.73	0.410
Fn1859	21 (9)	10 (8)	0.92	0.40-2.08	0.836	0.88	0.37-2.08	0.776	0.91	0.39-2.13	0.830
Fn1893	24 (10)	16 (13)	1.35	0.69-2.65	0.379	1.27	0.62-2.59	0.517	1.29	0.64-2.63	0.475
Any *F. nucleatum* protein	139 (58)	68 (55)	0.88	0.57-1.36	0.550	0.94	0.58-1.53	0.816	0.96	0.60-1.54	0.868

[1]Conditional logistic regression model; [2]Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are excluded from the analyses; [3]Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are considered as individual category; Significant associations are marked in bold font

Supplementary table 4: Antibody responses to *F. nculeatum* proteins in relation to CRC risk in a nested case-control study within EPIC

	Positive n (%)		Unadjusted model[1]			Adjusted model 1[2]			Adjusted model 2[3]		
	Controls n=485	Cases n=485	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Fn0131	46 (9)	31 (6)	0.66	0.41-1.05	0.081	0.65	0.39-1.10	0.107	**0.59**	**0.36-0.95**	**0.030**
Fn0253	15 (3)	10 (2)	0.67	0.30-2.48	0.321	0.58	0.24-1.43	0.240	0.69	0.30-1.60	0.391
Fn0264	24 (5)	32 (7)	1.36	0.79-2.36	0.269	0.97	0.51-1.84	0.928	1.37	0.78-2.43	0.276
Fn0387	38 (8)	42 (9)	1.11	0.71-1.76	0.642	1.16	0.70-1.91	0.562	1.05	0.66-1.69	0.828
Fn1426	49 (10)	53 (11)	1.10	0.72-1.65	0.673	0.90	0.56-1.43	0.651	1.04	0.68-1.61	0.844
Fn1449	47 (10)	41 (8)	0.86	0.55-1.34	0.503	0.86	0.53-1.40	0.547	0.83	0.52-1.30	0.414
Fn1526	20 (4)	15 (3)	0.74	0.37-1.47	0.386	0.72	0.34-1.52	0.395	0.71	0.35-1.44	0.337
Fn1817_1	49 (10)	40 (8)	0.80	0.51-1.24	0.312	0.87	0.54-1.40	0.565	0.77	0.49-1.23	0.276
Fn1817_2	48 (10)	46 (9)	0.96	0.64-1.44	0.835	0.92	0.58-1.44	0.705	0.96	0.63-1.47	0.864
Fn1859	46 (9)	34 (7)	0.71	0.45-1.14	0.159	**0.56**	**0.34-0.94**	**0.029**	0.69	0.43-1.12	0.134
Fn1893	47 (10)	45 (9)	0.95	0.62-1.47	0.825	1.01	0.63-1.63	0.966	0.91	0.58-1.43	0.692
Any *F. nucleatum* protein	255 (53)	230 (47)	0.81	0.62-1.04	0.101	**0.73**	**0.54-0.99**	**0.040**	0.79	0.60-1.04	0.087

[1]Conditional logistic regression model; [2]Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are excluded from the analyses; [3]Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are considered as individual category; Significant associations are marked in bold font

Supplementary table 5: Antibody responses to *F. nucleatum* proteins in relation to CRC risk in a nested case-control study within EPIC in cases diagnosed after more than two years from blood draw

	Positive n (%)		Unadjusted model[1]			Adjusted model 1[2]			Adjusted model 2[3]		
	Controls n=355	Cases n=355	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Fn0131	33 (9)	21 (6)	0.61	0.35-1.09	0.093	0.67	0.36-1.27	0.222	0.60	0.33-1.08	0.086
Fn0253	12 (3)	8 (2)	0.67	0.27-1.63	0.374	0.57	0.21-1.59	0.283	0.71	0.28-1.80	0.468
Fn0264	19 (5)	25 (7)	1.33	0.72-2.46	0.356	1.08	0.54-2.19	0.826	1.42	0.75-2.68	0.279
Fn0387	28 (8)	25 (7)	0.89	0.51-1.55	0.668	0.91	0.49-1.68	0.752	0.86	0.48-1.53	0.596
Fn1426	34 (10)	42 (12)	1.28	0.79-2.07	0.326	1.09	0.63-1.89	0.764	1.27	0.76-2.11	0.359
Fn1449	35 (10)	25 (7)	0.70	0.41-1.19	0.184	0.66	0.36-1.20	0.171	0.68	0.39-1.19	0.174
Fn1526	16 (5)	11 (3)	0.67	0.30-1.48	0.321	0.66	0.28-1.55	0.337	0.69	0.30-1.56	0.370
Fn1817_1	32 (9)	29 (8)	0.89	0.51-1.54	0.675	1.03	0.57-1.87	0.914	0.88	0.50-1.54	0.642
Fn1817_2	34 (10)	34 (10)	1.00	0.62-1.61	1.000	0.95	0.56-1.62	0.843	1.01	0.62-1.66	0.966
Fn1859	38 (11)	27 (8)	0.69	0.41-1.15	0.155	0.57	0.32-1.01	0.053	0.67	0.40-1.15	0.147
Fn1893	36 (10)	30 (8)	0.82	0.49-1.36	0.439	0.86	0.49-1.52	0.608	0.79	0.47-1.34	0.386
Any *F. nucleatum* protein	184 (52)	173 (49)	0.88	0.65-1.19	0.395	0.80	0.56-1.13	0.197	0.88	0.64-1.21	0.426

[1]Conditional logistic regression model; [2]Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are excluded from the analyses; [3]Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are considered as individual category; Significant associations are marked in bold font

VI. Different models for estimating the association of antibody responses to *S. gallolyticus* with CRC in BliTz/DACHSplus, SCCS and EPIC

Supplementary table 6: Antibody responses to *S. gallolyticus* proteins and protein combinations in relation to prevalence of CRC in the DACHSplus study compared to BliTz controls

	Positive n (%)		Unadjusted model[1]			Adjusted model 1[2]			Adjusted model 2[3]		
	Controls n=228	Cases n=318	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Gallo0112A	22 (10)	31 (10)	1.01	0.57-1.80	0.969	0.89	0.49-1.63	0.714	0.89	0.48-1.64	0.702
Gallo0112B	22 (10)	28 (9)	0.90	0.50-1.63	0.736	0.82	0.45-1.51	0.531	0.84	0.45-1.56	0.583
Gallo0272	22 (10)	39 (12)	1.31	0.75-2.28	0.340	1.14	0.64-2.02	0.658	1.13	0.63-2.04	0.686
Gallo0577	22 (10)	32 (10)	1.05	0.59-1.86	0.874	1.03	0.57-1.88	0.915	1.18	0.64-2.18	0.602
Gallo0748	22 (10)	37 (12)	1.23	0.71-2.15	0.462	1.18	0.66-2.11	0.569	1.32	0.72-2.41	0.373
Gallo0933	22 (10)	40 (13)	1.35	0.78-2.34	0.289	1.59	0.89-2.85	0.117	1.56	0.86-2.84	0.148
Gallo1570	22 (10)	16 (5)	**0.50**	**0.25-0.97**	**0.040**	**0.45**	**0.22-0.91**	**0.027**	0.51	0.25-1.06	0.069
Gallo1675	22 (10)	33 (10)	1.08	0.61-1.91	0.782	0.96	0.54-1.74	0.904	1.12	0.60-2.10	0.725
Gallo2018	22 (10)	38 (12)	1.27	0.73-2.21	0.398	1.17	0.65-2.08	0.602	1.23	0.68-2.23	0.491
Gallo2178	11 (5)	55 (17)	**4.13**	**2.11-8.08**	**<0.0001**	**4.30**	**2.14-8.65**	**<0.0001**	**4.50**	**2.22-9.11**	**<0.0001**
Gallo2179	22 (10)	34 (11)	1.12	0.64-1.97	0.692	1.22	0.68-2.19	0.512	1.35	0.73-2.51	0.342
Any *S. gallolyticus* protein	131 (57)	213 (67)	**1.50**	**1.06-2.14**	**0.023**	**1.47**	**1.02-2.12**	**0.039**	**1.65**	**1.13-2.41**	**0.010**
Gallo2178-Gallo2179 DP	0 (0)	14 (4)	-	-	**0.001**	-	-	-	-	-	-
≥ 2 of 6-marker panel[4]	24 (11)	60 (19)	**1.98**	**1.19-3.28**	**0.009**	**1.81**	**1.07-3.06**	**0.028**	**1.99**	**1.15-3.45**	**0.014**

[1]Logistic regression model without further adjustment; [2]Logistic regression model with adjustment for age (continuous variable) and sex; [3]Logistic regression model with adjustment for age (continuous variable), sex, BMI, education and smoking with exclusion of samples with missing baseline information in any of these variables (n(controls) = 218, n(CRC) = 303); [4]Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178, Gallo2179; DP = double-positive; Significant associations are marked in bold font; Significant associations with Bonferroni-correction (p<0.0036) are underlined

Supplementary table 7: Antibody responses to *S. gallolyticus* proteins and protein combinations in relation to CRC risk in a nested case-control study within SCCS

	Positive n (%)		Unadjusted model[1]			Adjusted model 1[2]			Adjusted model 2[3]		
	Controls n=348	Cases n=181	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Gallo0112A	34 (10)	20 (11)	1.14	0.62-2.08	0.679	1.26	0.66-2.40	0.480	1.33	0.71-2.52	0.373
Gallo0112B	34 (10)	21 (12)	1.20	0.67-2.14	0.547	1.32	0.72-2.44	0.372	1.25	0.68-2.29	0.473
Gallo0272	34 (10)	19 (11)	1.08	0.60-1.95	0.799	0.95	0.51-1.80	0.881	1.04	0.56-1.95	0.895
Gallo0577	34 (10)	26 (14)	1.50	0.88-2.56	0.135	1.25	0.72-2.18	0.437	1.33	0.77-2.31	0.309
Gallo0748	34 (10)	20 (11)	1.09	0.61-1.95	0.766	1.12	0.61-2.04	0.713	1.11	0.61-2.02	0.729
Gallo0933	34 (10)	13 (7)	0.71	0.37-1.38	0.315	0.78	0.38-1.57	0.478	0.81	0.41-1.61	0.547
Gallo1570	34 (10)	19 (11)	1.08	0.60-1.95	0.799	0.90	0.48-1.70	0.749	0.91	0.48-1.71	0.769
Gallo1675	34 (10)	19 (11)	1.10	0.61-1.97	0.763	1.11	0.60-2.06	0.741	1.07	0.58-1.97	0.826
Gallo2018	34 (10)	20 (11)	1.12	0.63-2.00	0.691	1.18	0.65-2.14	0.598	1.13	0.63-2.05	0.681
Gallo2178	34 (10)	24 (13)	1.40	0.82-2.41	0.218	1.31	0.74-2.29	0.355	1.31	0.74-2.29	0.352
Gallo2179	34 (10)	16 (9)	0.89	0.48-1.65	0.719	0.83	0.43-1.58	0.565	0.81	0.42-1.54	0.516
Any *S. gallolyticus* protein	206 (59)	112 (62)	1.11	0.77-1.58	0.584	1.13	0.77-1.65	0.534	1.11	0.76-1.60	0.599
Gallo2178-Gallo2179 DP	5 (1)	4 (2)	1.60	0.43-5.96	0.484	1.17	0.27-5.04	0.837	1.18	0.27-5.11	0.823
≥ 2 of 6-marker panel[4]	41 (12)	25 (14)	1.19	0.70-2.04	0.517	1.11	0.63-1.95	0.716	1.12	0.64-1.97	0.692

[1]Conditional logistic regression model; [2]Conditional logistic regression model with adjustment for BMI, education and smoking status as categorical variables, missings in the variables are excluded from the analyses; [3] Conditional logistic regression model with adjustment for BMI, education and smoking status as categorical variables, missings in the variables are considered as individual category; [4]Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178, Gallo2179; DP = double-positive; Significant associations are marked in bold font

Supplementary table 8: Antibody responses to *S. gallolyticus* proteins and protein combinations in relation to CRC risk in a nested case-control study within SCCS in cases diagnosed after more than two years from blood draw

	Positive n (%)		Unadjusted model[1]			Adjusted model 1[2]			Adjusted model 2[3]		
	Controls n=239	Cases n=124	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Gallo0112A	23 (9)	17 (14)	1.48	0.75-2.92	0.261	1.78	0.85-3.74	0.126	1.88	0.91-3.90	0.091
Gallo0112B	24 (10)	13 (10)	1.02	0.49-2.12	0.951	1.14	0.52-2.49	0.744	1.03	0.48-2.23	0.935
Gallo0272	23 (10)	13 (10)	1.12	0.54-2.32	0.755	0.91	0.41-2.03	0.825	1.08	0.49-2.35	0.854
Gallo0577	22 (9)	18 (15)	1.59	0.82-3.06	0.169	1.24	0.62-2.47	0.544	1.38	0.70-2.72	0.358
Gallo0748	27 (11)	12 (10)	0.79	0.38-1.62	0.516	0.78	0.37-1.66	0.515	0.77	0.36-1.64	0.498
Gallo0933	22 (9)	6 (5)	0.51	0.20-1.29	0.153	0.48	0.17-1.36	0.164	0.56	0.21-1.49	0.245
Gallo1570	22 (9)	10 (8)	0.87	0.41-1.84	0.709	0.78	0.35-1.73	0.536	0.79	0.35-1.74	0.552
Gallo1675	24 (10)	13 (10)	1.04	0.52-2.09	0.905	0.92	0.44-1.92	0.817	0.92	0.44-1.91	0.823
Gallo2018	23 (10)	13 (10)	1.07	0.53-2.16	0.857	1.09	0.52-2.29	0.811	1.09	0.52-2.27	0.819
Gallo2178	23 (10)	16 (13)	1.38	0.72-2.64	0.332	1.36	0.69-2.68	0.381	1.35	0.69-2.66	0.386
Gallo2179	21 (9)	13 (10)	1.17	0.57-2.40	0.664	1.01	0.47-2.17	0.974	1.03	0.48-2.20	0.943
Any *S. gallolyticus* protein	142 (59)	74 (60)	1.00	0.65-1.53	1.000	0.97	0.61-1.53	0.886	1.00	0.64-1.56	0.985
Gallo2178-Gallo2179 DP	4 (2)	4 (3)	2.00	0.50-8.00	0.327	1.48	0.32-6.81	0.614	1.50	0.33-6.88	0.604
≥ 2 of 6-marker panel[4]	29 (12)	17 (14)	1.13	0.60-2.15	0.700	0.97	0.49-1.91	0.921	0.98	0.50-1.93	0.952

[1]Conditional logistic regression model; [2]Conditional logistic regression model with adjustment for BMI, education and smoking status as categorical variables, missings in the variables are excluded from the analyses; [3] Conditional logistic regression model with adjustment for BMI, education and smoking status as categorical variables, missings in the variables are considered as individual category; [4]Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178, Gallo2179; DP = double-positive; Significant associations are marked in bold font

Supplementary table 9: Antibody responses to *S. gallolyticus* proteins and protein combinations in relation to CRC risk in a nested case-control study within EPIC

	Positive n (%)		Unadjusted model[1]			Adjusted model 1[2]			Adjusted model 2[3]		
	Controls n=485	Cases n=485	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Gallo0112A	33 (7)	37 (8)	1.14	0.69-1.90	0.606	1.26	0.71-2.24	0.426	1.08	0.63-1.82	0.788
Gallo0112B	28 (6)	26 (5)	0.93	0.54-1.60	0.782	0.90	0.49-1.21	0.730	0.95	0.55-1.66	0.861
Gallo0272	47 (10)	67 (14)	**1.49**	**1.00-2.21**	**0.049**	1.45	0.93-2.26	0.100	**1.52**	**1.01-2.29**	**0.044**
Gallo0577	47 (10)	49 (10)	1.05	0.69-1.59	0.831	1.10	0.67-1.21	0.715	1.04	0.68-1.61	0.844
Gallo0748	50 (10)	74 (15)	**1.51**	**1.05-2.18**	**0.028**	**1.60**	**1.06-2.42**	**0.025**	**1.49**	**1.02-2.17**	**0.039**
Gallo0933	49 (10)	44 (9)	0.89	0.58-1.36	0.583	0.99	0.62-1.58	0.971	0.91	0.59-1.43	0.689
Gallo1570	47 (10)	52 (11)	1.13	0.73-1.74	0.583	1.07	0.65-1.78	0.781	1.17	0.75-1.84	0.491
Gallo1675	48 (10)	51 (11)	1.07	0.70-1.63	0.748	1.21	0.75-1.94	0.435	1.12	0.72-1.72	0.618
Gallo2018	47 (10)	54 (11)	1.16	0.77-1.74	0.473	1.44	0.91-2.28	0.118	1.24	0.81-1.88	0.325
Gallo2178	12 (2)	31 (6)	**2.58**	**1.33-5.03**	**0.005**	**2.78**	**1.33-5.80**	**0.007**	**2.74**	**1.39-5.40**	**0.004**
Gallo2179	47 (10)	64 (13)	1.43	0.95-2.14	0.086	1.50	0.95-2.37	0.085	1.44	0.95-2.19	0.090
Any *S. gallolyticus* protein	273 (56)	306 (63)	**1.32**	**1.02-1.71**	**0.033**	**1.43**	**1.07-1.89**	**0.015**	**1.36**	**1.05-1.78**	**0.022**
Gallo2178-Gallo2179 DP	2 (0)	12 (2)	**6.00**	**1.34-26.81**	**0.019**	**4.72**	**1.01-22.13**	**0.049**	**7.02**	**1.52-32.51**	**0.013**
≥ 2 of 6-marker panel[4]	45 (9)	83 (17)	**2.03**	**1.37-3.01**	**<u>0.0004</u>**	**2.34**	**1.50-3.65**	**<u>0.0002</u>**	**2.10**	**1.40-3.14**	**<u>0.0004</u>**

[1]Conditional logistic regression model; [2]Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are excluded from the analyses; [3] Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are considered as individual category; [4]Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178, Gallo2179; DP = double-positive; Significant associations are marked in bold font; Significant associations with Bonferroni-correction ($p<0.0036$) are underlined

Supplementary table 10: Antibody responses to *S. gallolyticus* proteins and protein combinations in relation to CRC risk in a nested case-control study within EPIC in cases diagnosed after more than two years from blood draw.

	Positive n (%)		Unadjusted model[1]			adjusted model 1[2]			adjusted model 2[3]		
	Controls n=355	Cases n=355	OR	95% CI	p-value	OR	95% CI	p-value	OR	95% CI	p-value
Gallo0112A	22 (6)	23 (6)	1.06	0.55-2.01	0.869	1.21	0.58-2.55	0.615	1.10	0.56-2.16	0.785
Gallo0112B	15 (4)	16 (5)	1.07	0.53-2.16	0.858	1.02	0.48-2.15	0.963	1.17	0.57-2.42	0.668
Gallo0272	32 (9)	51 (14)	1.68	**1.05-2.68**	**0.030**	**1.72**	**1.01-2.91**	**0.045**	**1.74**	**1.07-2.82**	**0.026**
Gallo0577	34 (10)	36 (10)	1.06	0.66-1.72	0.806	1.21	0.69-2.12	0.499	1.10	0.67-1.81	0.713
Gallo0748	37 (10)	51 (14)	1.40	0.91-2.16	0.128	1.50	0.92-2.44	0.103	1.38	0.89-2.15	0.151
Gallo0933	37 (10)	38 (11)	1.03	0.63-1.68	0.901	1.10	0.65-1.85	0.729	1.08	0.65-1.78	0.772
Gallo1570	36 (10)	41 (12)	1.17	0.72-1.90	0.536	1.18	0.66-2.10	0.581	1.25	0.75-2.07	0.391
Gallo1675	38 (11)	39 (11)	1.03	0.64-1.65	0.904	1.23	0.72-2.10	0.443	1.06	0.65-1.72	0.813
Gallo2018	38 (11)	43 (12)	1.14	0.73-1.78	0.569	1.34	0.81-2.23	0.255	1.19	0.75-1.89	0.466
Gallo2178	7 (2)	17 (5)	2.43	**1.01-5.86**	**0.048**	**2.93**	**1.04-8.30**	**0.043**	**2.64**	**1.07-6.51**	**0.035**
Gallo2179	34 (10)	44 (12)	1.33	0.83-2.14	0.234	1.46	0.85-2.51	0.171	1.46	0.89-2.39	0.132
Any *S. gallolyticus* protein	201 (57)	224 (63)	1.30	0.96-1.74	0.087	**1.41**	**1.01-1.95**	**0.041**	**1.37**	**1.01-1.86**	**0.043**
Gallo2178-Gallo2179 DP	1 (0)	7 (2)	7.00	0.86-56.89	0.069	4.45	0.48-41.55	0.191	**9.92**	**1.03-77.45**	**0.047**
≥ 2 of 6-marker panel[4]	36 (10)	60 (17)	1.80	**1.15-2.81**	**0.010**	**2.24**	**1.34-3.74**	**0.002**	**1.91**	**1.21-3.03**	**0.006**

[1]Conditional logistic regression model; [2]Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are excluded from the analyses; [3] Conditional logistic regression model with adjustment for BMI, education, smoking and alcohol status as categorical variables, missings in the variables are considered as individual category; [4]Gallo0272, Gallo0748, Gallo1675, Gallo2018, Gallo2178, Gallo2179; DP = double-positive; Significant associations are marked in bold font; Significant associations with Bonferroni-correction ($p<0.0036$) are underlined

VII. Abbreviations

°C	Degrees Celsius
%CV	Coefficient of variation
amp	ampicillin
APC	*Adenomatous polyposis coli*
APS	Ammoniumperoxodisulfate
ASR	Age-standardized incidence rate
ATCC	American type culture collection
BliTz	"Begleitende Evaluierung innovativer Testverfahren zur Darmkrebs-Früherkennung"
BMI	Body mass index
bp	Base pairs
BSA	Bovine serum albumin
CagA	Cytotoxin-associated antigen A
CBS-K	Superchemiblock
CD3	Cluster of differentiation 3
CHC	Community health care centers
CI	Confidence interval
CIMP	CpG island methylator phenotype
cm	centimeter
CRC	Colorectal cancer
C-terminus	Carboxy-terminus
CW	Cell wall
DACHSplus	"Darmkrebs: Chancen der Verhütung durch Screening"
ddH_2O	Double-distilled water
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DP	Double-positive
DTT	1,4-Dithiothreitol
EC	extracellular
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetate
e.g.	Exempli gratia
ELISA	Enzyme-linked immunosorbent assay
EPIC	European prospective investigation into cancer and nutrition
et al.	Et alii
FIT	Fecal immunochemical test
F. nucleatum	*Fusobacterium nucleatum*
g	gramm
GC	Gastric cancer
gFOBT	Guaiac-based fecal occult blood test

GST	Glutathione-S-transferase
HC	Healthy control
HNPCC	Hereditary non-polyposis colorectal cancer
H. pylori	*Helicobacter pylori*
HRP	Horse radish peroxidase
IBD	Inflammatory bowel disease
ICD-O	International classification of diseases for oncology
i.e.	Id est
Ig	Immunoglobulin
inHg	Inch Hg
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kD	kilo Dalton
KRAS	*Kirsten rat sarcoma*
kV	kilovolt
LB	Lysogeny broth
µF	microfarad
µl	microliter
M	Marker
MAPK	Mitogen-activated protein kinase
MFI	Median fluorescence intensity
min	minute
ml	milliliter
mm	millimeter
mM	millimolar
MSI-H	Microsatellite instability-high
n	Number
neg	Negative
NFκB	Nuclear factor kappa B
nm	nanometer
NSAID	Non-steroidal anti-inflammatory drugs
nt	nucleotide
N-terminus	Amino-terminus
OD	Optical density
OM	Outer membrane
Omp	Outer membrane protein
OR	Odds ratio
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PBS-T	PBS-Tween
PCR	Polymerase chain reaction
pos	Positive
PVA	Polyvinylalcohol
PVP	Polyvinylpyrrolidone

PVX	Synonym for PVA plus PVP
qPCR	Quantitative PCR
R^2	R-square correlation coefficient
ref	reference
RNA	Ribonucleic acid
rpm	Rounds per minute
S. bovis	*Streptococcus bovis*
SCCS	Southern community cohort study
SDS	Sodium dodecyl sulfate
sec	seconds
SEER	Surveillance, epidemiology, and end results program
S. equinus	*Streptococcus equinus*
S. gallolyticus	*Streptococcus gallolyticus* subsp. *gallolyticus*
SGG-pos	*S. gallolyticus* positive
SGMB	*S. gallolyticus* member bacteria
S. infantarius	*Streptococcus infantarius* subsp. *infantarius*
S. macedonicus	*Streptococcus gallolyticus* subsp. *macedonicus*
S. pasteurianus	*Streptococcus gallolyticus* subsp. *pasteurianus*
Strep-PE	Streptavidin-R-Phycoerythrin
subsp.	Subspecies
SV40	Simian virus 40
TEMED	N,N,N',N'-Tetramethylethylendiamin
TMB	Tetramethylbencidine
UICC	"Union international contre le cancer"
UK	United Kingdom
US	United States of America
VacA	Vacuolating cytotxin A
WHO	World Health Organization

VIII. Publications

Published

Werner S, Chen H, Butt J, Michel A, Knebel P, Holleczek B, Zörnig I, Eichmüller SB, Jäger D, Pawlita M, Waterboer T, Brenner H. Evaluation of the diagnostic value of 64 simultaneously measured autoantibodies for early detection of gastric cancer. *Sci Rep*. 2016 May 3;6:25467. doi: 10.1038/srep25467.

Butt J, Romero-Hernández B, Pérez-Gómez B, Willhauck-Fleckenstein M, Holzinger D, Martin V, Moreno V, Linares C, Dierssen-Sotos T, Barricarte A, Tardón A, Altzibar JM, Moreno-Osset E, Franco F, Requena RO, Huerta JM, Michel A, Waterboer T, Castaño-Vinyals G, Kogevinas M, Pollán M, Boleij A, de Sanjosé S, Del Campo R, Tjalsma H, Aragonés N, Pawlita M. Association of Streptococcus gallolyticus subspecies gallolyticus with colorectal cancer: Serological evidence. *Int J Cancer*. 2016 Apr 1;138(7):1670-9. doi: 10.1002/ijc.29914.

Chen H, Werner S, Butt J, Zörnig I, Knebel P, Michel A, Eichmüller SB, Jäger D, Waterboer T, Pawlita M, Brenner H. Prospective evaluation of 64 serum autoantibodies as biomarkers for early detection of colorectal cancer in a true screening setting. *Oncotarget*. 2016 Mar 29;7(13):16420-32. doi: 10.18632/oncotarget.7500.

Sankaranarayanan R, Prabhu PR, Pawlita M, Gheit T, Bhatla N, Muwonge R, Nene BM, Esmy PO, Joshi S, Poli UR, Jivarajani P, Verma Y, Zomawia E, Siddiqi M, Shastri SS, Jayant K, Malvi SG, Lucas E, Michel A, Butt J, Vijayamma JM, Sankaran S, Kannan TP, Varghese R, Divate U, Thomas S, Joshi G, Willhauck-Fleckenstein M, Waterboer T, Müller M, Sehr P, Hingmire S, Kriplani A, Mishra G, Pimple S, Jadhav R, Sauvaget C, Tommasino M, Pillai MR; Indian HPV Vaccine Study Group. Immunogenicity and HPV infection after one, two, and three doses of quadrivalent HPV vaccine in girls in India: a multicentre prospective cohort study. *Lancet Oncol*. 2016 Jan;17(1):67-77. doi: 10.1016/S1470-2045(15)00414-3.

Murphy G, Michel A, Taylor PR, Albanes D, Weinstein SJ, Virtamo J, Parisi D, Snyder K, Butt J, McGlynn KA, Koshiol J, Pawlita M, Lai GY, Abnet CC, Dawsey SM, Freedman ND. Association of seropositivity to Helicobacter species and biliary tract cancer in the ATBC study. *Hepatology*. 2014 Dec;60(6):1963-71. doi: 10.1002/hep.27193.

In preparation

Butt J, Werner S, Willhauck-Fleckenstein M, Michel A, Waterboer T, Zörnig I, Boleij A, Dramsi S, Brenner H, Pawlita M. Serology of Streptococcus gallolyticus subspecies gallolyticus and its association with colorectal cancer and precursors.