Detection of *Fusobacterium nucleatum* subspecies in saliva of pre-colorectal cancer patients and negative controls, using matrix-assisted laser desorption ionization/tandem time-of-flight mass spectrometry

Hanie Morsi
Department of Microbiology and Immunology,
McGill University, Montreal
July 17, 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Table of contents

List of figures

List of tables

Acknowledgments

First and foremost, I would like to extend my sincerest gratitude to my mentor and previous supervisor, Prof. Dr. Herve Le Moual for accepting me as a graduate student and training me with patience and precision, allowing me to engage in various research projects. His support will never be forgotten. This thesis is dedicated to his memory.

I am forever grateful to my supervisor, Prof. Dr. Simon Tran, for all his advice and supervision throughout my research work in McGill. His door was consistently open to my questions, and he steered me in the right direction whenever he saw necessary. Dr. Tran gave me a concrete example of the welcoming, cheerful, and supportive teacher, scientist, and life coach that I would live up to years to come. I owe it all to him!

I would also like to acknowledge Prof. Dr. Momar Ndao for allowing me the opportunity to use his laboratory facilities to conduct my research, and for his revision of this thesis. A special mention goes to Dr. Makan Golizeh for his tremendous efforts, instructions, and thorough editing of my thesis. I am also grateful to Noah Brosseau for his valuable input and initiatives in the project research and experiments. It was not possible for this work to succeed without their valuable participation and validation.

A special thanks go to Prof. Dr. Elham Emami for her input during the initiation of the study. I am also grateful for Amal Janati, from Dr. Emami lab, for her significant efforts in collecting the saliva samples that made this work doable.

I want to extend my thanks to my advisory committee members, Dr. Samantha Gruenheid and Dr. Corinne Maurice for offering their gracious guidance throughout my project. Dr. Maurice also has allowed me to use her lab and the anaerobic chamber, with the help of William Jogia, which I am genuinely grateful for.

Of particular importance to thank all Dr. Le Moual lab members for their friendship and assistance throughout our times together, especially Anshul Sinha whom I am incredibly appreciative for all of his continuous help and mentorship throughout my presence in McGill and in Montreal, and Natalie Giannakopoulou for her availability to answer my queries and offer her generous support along the way.

I am also indebted for Dr. Tran lab members whom I had the pleasure to work with, especially Ghada Abu Elghanam for her thoughtful insight into this project.

This work would not be possible without the generous funding of Al Ghurair Foundation for Education. Also, special gratitude to Prof. Jong-Ki Kook for providing us the *Fusobacterium nucleatum* strains.

Finally and most importantly, I would like to thank my loving and wonderful family members and my parents, for their unconditional support and for providing me an unending inspiration.

Abbreviations

ABC- ammonium bicarbonate

ACN - acetonitrile

CHCA - alpha-cyano-4-hydroxy-cinnamic acid

CRC- Colorectal cancer

F. nucleatum- *Fusobacterium* nucleatum

Fap 2- Fibroblasts activation protein 2

IAM - iodoacetamide

IBD - inflammatory bowel disease

m/z- mass-to-charge

MAED- microwave-assisted enzymatic digestion

MALDI-TOF MS- matrix-assisted laser desorption ionization-time of flight mass spectrometry

MALDI-TOF/TOF MS- matrix-assisted laser desorption ionization/tandem time-of-flight mass spectrometry

MS- mass spectrometry

MS/MS- tandem mass spectrometry

PCR- polymerase chain reaction

PMF- peptide mass fingerprinting

PSMs- peptide spectral matches

SPE- Solid-phase extraction

TFA- trifluoroacetic acid

TOF/TOF- Tandem time of flight

Contribution of authors

This thesis was written by Hanie Morsi and edited by Dr. Simon Tran, Dr. Momar Ndao, and Dr. Makan Golizeh.

Noah Bresseu has contributed to the methods development with the instructions of Dr. Makan Golzeh. Specifically, Noah has run the AB Sciex 4800™ MALDI TOF/TOF experiment. Noah has also participated in the microwave-assisted digestion experiment as well as the analysis of the patients' and the controls' saliva.

Amal Janati has collected the saliva samples used in this study.

All the experiments were run by Hanie Morsi with the guidance of Dr. Makan Golizeh, unless stated otherwise.

Abstract

Colorectal cancer (CRC) is known to be one of the most aggressive types of cancer. While the gut is a diverse ecological inhabitant of thousands of bacterial species, rising evidence links *Fusobacterium* nucleatum (*F. nucleatum)* with the disease. Whether this bacterium is a causative factor for the disease or a species that merely happens to flourish in the CRC environment, is not fully determined. *F. nucleatum* has been further classified into four genetically different subspecies; *nucleatum*, *polymorphum*, *animalis*, and *vincentii*. These subspecies are thought to vary in their virulence and the various diseases they might cause.

Up to that time, conventional matrix-assisted laser desorption ionization- time-of-flight mass spectrometry (MALDI-TOF MS) has been used as a proteomics tool to detect the *F. nucleatum* four subspecies, in pure cultures. Recently, the more advanced tandem time-of-flight MALDI-MS (MALDI-TOF/TOF MS) technology has been introduced as an efficient and accurate method of detecting bacterial pathogens at microbiology clinical diagnostic laboratories.

Human saliva has been proposed as a potential route of *F. nucleatum* strains to reach and colonize the CRC tissue. Along with the non-invasiveness, time-effectiveness, and ease of acquisition, saliva represents a potential mirror for a myriad of systemic health changes in humans. Hence, human saliva has been extensively used as a promising diagnostic tool in healthcare. For this master's thesis, we explored the capability of MALDI-TOF/TOF MS to identify *F. nucleatum* to the subspecies level in saliva samples. The analyzed saliva specimens

were obtained from patients with precancerous adenoma, the primary precursor for CRC, and from non-diseased controls.

Results: We have optimized a proteomics approach that allows for the successful identification of *F. nucleatum* at the subspecies-level directly from saliva samples. Our findings highlight the presence of the *F. nucleatum* subspecies in both colorectal adenoma patients' saliva and non-diseased controls' saliva. We have also shown a difference in the *F. nucleatum* subtypes inhabiting the diseased and non-diseased subjects. Specifically, *F. nucleatum* subsp. *nucleatum* was absent in the patients' saliva while it was detected in the non-diseased control group. The possible identification of a specific *F. nucleatum* pathogen in the saliva may enable future targeting of this pathogen as a means of screening for, detection, treatment, follow-up or prevention of CRC.

Résumé

Le cancer colorectal (CCR) est connu pour être l'un des types de cancer les plus agressifs. Alors que le côlon est un habitat écologique diversifié de milliers d'espèces bactériennes, de plus en plus de preuves relient *Fusobacterium nucleatum* (*F. nucleatum*) au CCR. La question de savoir si cette bactérie est un facteur causal de la maladie, ou une espèce qui ne fait que prospérer dans l'environnement du CCR, n'est pas totalement résolue. *F. nucleatum* a en effet été classé en quatre sous-espèces génétiquement différentes: *nucleatum*, *polymorphum*, *animalis* et *vincentii*; et on pense que ces sous-espèces varient aussi bien dans leur virulence que dans les maladies qu'elles pourraient causer.

Jusque-là, la spectrométrie de masse conventionnelle par désorption-ionisation laser à temps de vol assistée par matrice (MALDI-TOF MS, du nom en anglais : *Matrix-assisted laser desorption ionization- time-of-flight mass spectrometry*) était utilisée comme outil de protéomique pour détecter les quatre sous-espèces de *F. nucleatum*, dans des cultures pures. Récemment, la technologie plus avancée en temps de vol, MALDI-MS (MALDI-TOF / TOF MS), a été introduite en tant que méthode efficace et précise de détection des bactéries pathogènes dans les laboratoires de diagnostic clinique microbiologique .

La salive a été proposée comme voie potentielle des souches de *F. nucleatum* pour atteindre et coloniser les tumeurs colorectales. En plus de son caractère non invasif, son efficacité en terme de temps et sa facilité d'acquisition, la salive représente un portrait potentiel d'une myriade de changements systémiques dans la santé chez l'homme. Par conséquent, la salive humaine a été largement utilisée comme outil de diagnostic prometteur

dans les soins de santé. Pour cette thèse de maîtrise, nous avons exploré la capacité de MS MALDI-TOF / TOF à identifier *F. nucleatum* au niveau de la sous-espèce dans des échantillons de salive. Les échantillons de salive analysés proviennent de patients diagnostiqués avec des adénomes précancéreux et de témoins non atteints, sachant que les adénomes précancéreux sont des précurseurs de CCRs, .

Résultats: Nous avons optimisé une approche protéomique permettant l'identification réussie de F. *nucleatum* au niveau de la sous-espèce directement à partir d'échantillons de salive. Nos résultats mettent en évidence la présence de la sous-espèce *F. nucleatum* dans la salive des patients souffrant d'un adénome colorectal et dans celle des témoins non atteints. Nous avons également montré une différence entre les sous-types de *F. nucleatum* chez les sujets malades et ceux non malades. Plus précisément, *F*. *nucleatum* subsp. *nucleatum* était absent dans la salive des patients alors qu'il a été détecté dans le groupe témoin non malade. L'identification possible d'un agent pathogène spécifique de *F. nucleatum* dans la salive pourrait permettre un ciblage ultérieur de cet agent pathogène en tant que moyen de dépistage, de détection, de traitement, de suivi ou de prévention du CCR.

1. Literature Review

1.1 *Fusobacterium*

Fusobacterium bacterial genus is a strictly anaerobic non-sporulating, Gram-negative staining bacterial group. It belongs to the family *Bacteriodaceae*. Some of the species belonging to this family are common inhabitants of the oral cavity, specifically: *Fusobacterium nucleatum, Fusobacterium necrophorum,* and *Fusobacterium varium.* While most of the *Fusobacteria* are spindle-shaped, some of them are polymorphous, e.g., filiform-shaped cells, spherical distended or globular-shaped cells(1). Fusobacteria were detected in clinical infections, e.g., pus and gangrene, as well as in children diseases(1–3).

1.2 *Fusobacterium nucleatum* (*F.nucleatum*)

F. nucleatum is one of the most common species that occurs to be found in humans body cavities infections(4,5). Additionally, it is one of the most abundant species in the normal flora of the oral cavity(6). Despite being closely correlated to periodontitis and gingivitis, *F. nucleatum* has been detected in multiple sites of the human body(7), and in a plethora of extraoral infections and conditions, for instance, peritonsillar abscesses(8), bacteremia and liver abscesses(9,10), urinary tract infections(11), intrauterine infections(12), bacterial vaginosis(13), pericarditis and endocarditis(14,15), lung and pleural infections(16,17), septic arthritis and pyomyositis(18), appendicitis(19,20), pancreatic cancer(21), preterm birth and stillbirth(22) and tropical skin ulcers(23).

1.3 *F. nucleatum* subspecies

F. nucleatum is further classified into five subspecies that are normal inhabitants of the oral cavity(24) - *nucleatum*, *polymorphum*, *animalis*, *fusiforme* and *vincentii*. *F. nucleatum* subsp. *fusiforme* and *vincentii* could be classified into one subspecies based on 25 housekeeping genes, including 16S rRNA, *rpoB*, and zinc protease, according to previous genomic analyses(25,26). These subspecies are known to vary in their virulence, pathogenic activity, and different diseases that they cause (3,27–29). As an example, while *F. nucleatum* ssp. *animalis* and *polymorphum* are linked with pregnancy complications and *F. nucleatum* ssp. *nucleatum* is often isolated from periodontal disease, *F. nucleatum* ssp. *vincentii* is usually found in the flora of the oral cavity(28). In a previous study, *F. nucleatum* subsp. *animalis* was found to be the most prevalent *F. nucleatum* subspecies associated with colorectal cancer (CRC) tissue samples(30). Additionally, another study has highlighted the dominance of *F. nucleatum* ssp. *animalis* and F. *nucleatum* ssp. *vincentii* in the fecal samples of CRC patients(31).

1.4 *F. nucleatum* in human saliva, gut, and stool

Although the oral microbiome is known to be the second most diverse bacterial community, coming after the stool microbiome(32,33), it has been less extensively studied than the gut microbiome(34). The oral microbiome constitutes of more than 2000 bacterial taxa(32,35). The gut microbiome could acquire the oral microbes through either the bloodstream or salivary secretion(36). In average, more than 1 L of saliva is secreted by an adult human per day, which generally flows into the gastrointestinal tract. Thus, the gut microbiome composition is affected by its salivary opponent(37). A previous study has elucidated a difference in the microbiota

composition between CRC patients and healthy individuals saliva, stool, and colon tissue(37). According to another study, the level of *F*. *nucleatum* subspecies was found to be more abundant in the stool of CRC patients when compared to healthy controls(38). At the subspecies level, *F. nucleatum* ssp. *animalis* was shown to be the most predominant *F*. *nucleatum* subspecies in CRC tissue samples in another study(39). In another recent research, identical *F*. *nucleatum* strains were found in the saliva and the colon tissues of CRC patients(40).

1.5 Colorectal neoplasms: colorectal adenoma and colorectal cancer (CRC)

Colorectal adenomatous polyps are known to be the most dominant precursors of colorectal cancers. In other words, most of the colorectal carcinomas arise from adenomas(41–44). Adenomas were shown to affect 20 to 53% of people older than 50 years old in the United States(45). CRC is the most prevalent gastrointestinal cancer(46) and a leading cause of death among both men and women around the globe, accounting for more than 690,000 deaths in 2012(47). In 2017, 73 Canadians were diagnosed with CRC every day on average, with almost 10,000 death cases with CRC in the same year(48). Although the exact cause of the disease is not yet elucidated, some risk factors have been highlighted including; smoking, lack of physical activity and obesity, as well as alcohol, red and processed meat consumption. On top of these, both family history and personal history of CRC, colorectal polyps, or inflammatory bowel disease (IBD) are known to be significant risk factors as well(49,50). Recently, accumulating evidence is linking CRC with the intestinal microbiota (36,51–53).

1.5.1 Gastrointestinal microbiome relation to colorectal neoplasms

The human gut is inhabited by more than 100 trillion microorganisms, which serves an essential role in human body health(6). These microorganisms are involved in homeostasis, through regulating a variety of biologic functions in the intestine(54,55). Upon disruption of this intestinal balance, i.e., microbial dysbiosis, a myriad of complications could emerge, e.g., inflammatory bowel disease (IBD) and colorectal tumors(56,57).

1.5.2 *F. nucleatum* in colorectal neoplasms

Several studies have flagged the presence of *Fusobacterium nucleatum* in a significantly higher abundance in colorectal adenomas (58–60) and in CRC-affected tissue compared to the adjacent non-cancerous areas of the same tissue (61–65). These studies are consistent with other reports on the presence of a higher *F. nucleatum* load in CRC tissues even at an early point of clinical lesion development(59,64,66,67). Additionally, other studies have highlighted that *F. nucleatum* promotes the pathogenicity of the disease through synergizing with other Gram-negative bacteria such as *Campylobacter* spp., *Liptotrichia* spp. and *Clostridium difficile* (68–70).

Recent studies have also shown that the presence of *F. nucleatum* might trigger some molecular events that could play a key role of CRC pathogenesis; including BRAF, TP53, CHD7 and CHD8 mutations, microsatellite instability and CpG island methylator phenotype (71,72). Despite a dearth of direct clinical evidence, *F. nucleatum* is now considered a potential contributing factor of CRC susceptibility rather than a mere passenger in the human gut(58,73). In one study performed on Apc min/+ mice, *F. nucleatum* was proven to be involved in

carcinogenesis (59). Another study demonstrated that *F. nucleatum* could cause the CRC cells to grow via β-catenin signaling pathway activation and promotion of oncogenic gene expression through the FadA adhesion molecule (74). FadA is an essential virulence factor of *F. nucleatum* as it aids the attachment of the bacteria to, and invasion of, the endothelial and epithelial host tissue (75,76). *F. nucleatum* invades the cells through binding of FadA to the endothelial cadherin receptor CDH5, upregulating inflammatory genes and promoting the secretion of cytokines, e.g., IL-6, IL-8, IL-10, IL-18, TNF-α and NF-κB resulting in a pro-inflammatory tumor microenvironment that facilitates tumor growth(58,77,78). FadA was also reported to interact with E-cadherin surface molecule, which is a cell adhesion molecule expressed on both CRC and non-CRC host cells and thought to be a strong tumor suppressor(79). This interaction triggers β-catenin signaling, which then leads to the upregulation of Wnt genes such as wnt7a, wnt7b, wnt9a, inflammatory genes such as NF-κB2, oncogenes such as myc and cyclin D1, and transcriptional factors such as lymphoid enhancer factor (LEF-1), eventually leading to tumor proliferation(74). Meanwhile, the bacteria-infected cells increase micro-RNA 21 (miR21) expression through activating TLR-4 signal transduction to MYD88, which leads to the activation of NF-κB (80). Fibroblasts activation protein 2 (Fap 2) is another critical virulence factor of *F. nucleatum*, that leads to the suppression of anti-tumor immunity in various pathways. Fap 2 is believed to bind to TIGIT receptor, which is an inhibitory receptor expressed by natural killer (NK) and T cells, protecting the *F. nucleatum*-infected tumor cells from these cells' attacks (80). Moreover, Fap 2 contributes to the death of human lymphocytes, as well as to the attraction of myeloid-derived suppressor cells (MDSC), which are thought to promote tumor development, through the immunosuppression of the tumor microenvironment(59,81,82). Fap 2 additionally

binds to the Gal-GalNAc host sugar receptor, which is overly expressed in CRC cells, enriching the tumor tissue with *F. nucleatum*(83). The pathways through which *F.nucleatum* is presumed to induce carcinogenicity are shown in **Figure 1**.

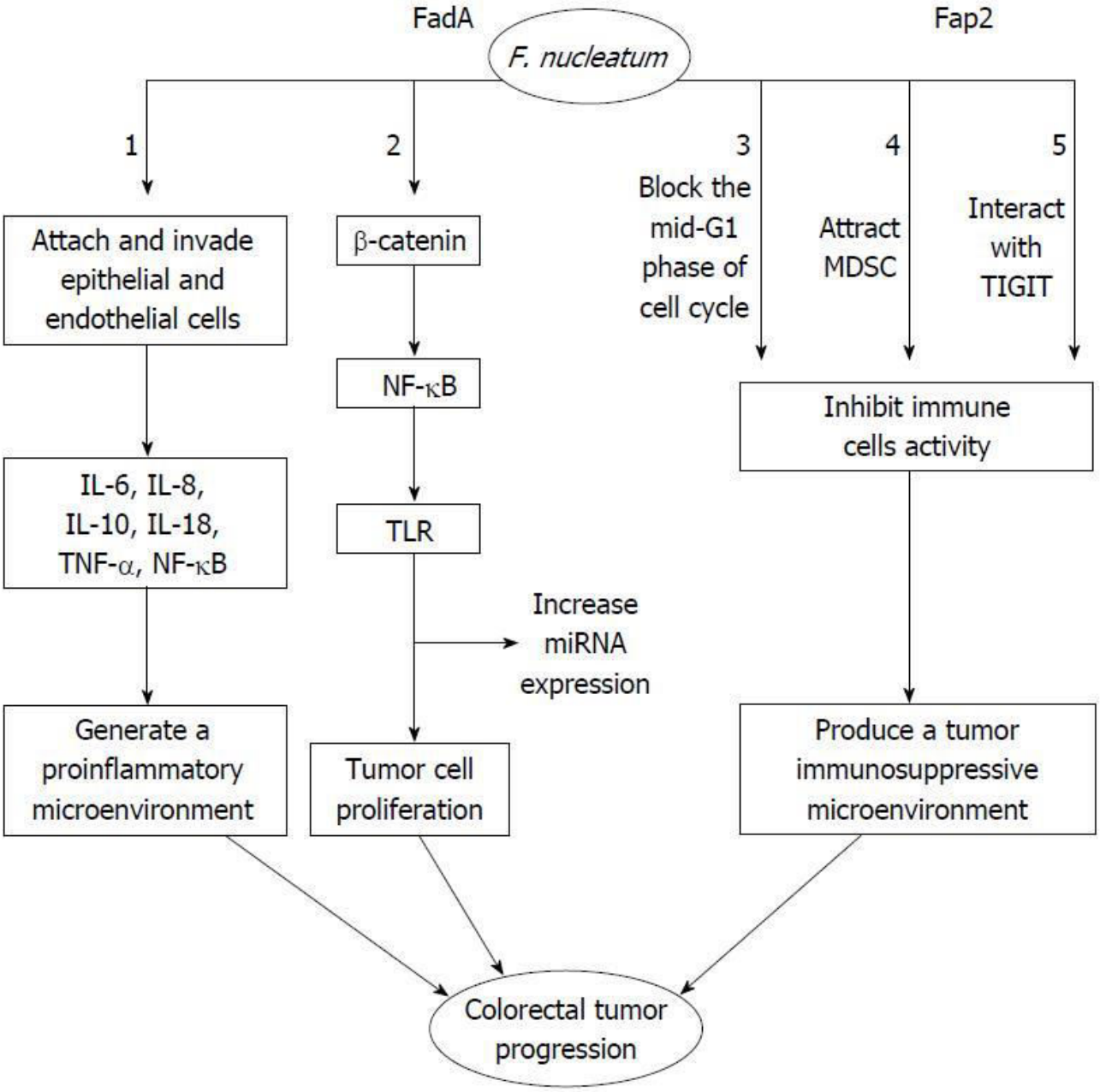

Figure 1. 1 Plausible *F. nucleatum* pathways of inducing CRC pathogenesis mechanisms. FadA in *F. nucleatum* promotes the CRC progression through increasing the inflammatory cytokines level via attachment and invasion of the endothelial and epithelial cells (pathway 1) or through activating E-cadherin β-Catenin signaling which increases the tumor cells proliferation (pathway 2). Fap2 also in *F. nucleatum* aids in CRC progression by producing an immune-suppressed tumor microenvironment (pathways 3,4, and 5). MDSC: Myeloid-derived suppressor cell; TLR: Toll-like receptor**.** Adopted from (84).

1.6 Detection of bacteria in human host tissues

Molecular techniques such as DNA/RNA sequencing, polymerase chain reaction (PCR) and aptamer chemistry are the most widely used and accepted tools for bacterial identification at the subspecies level, also known as "biotyping"(25,85–87). These methods could however have limited applicability and could be both time-consuming and labor-intensive for subspecies biotyping(25,88,89). Here comes the need for a precise and reliable method for bacterial identification at the subspecies levels, as it could be used as the initial step of disease diagnosis and a tool in large-scale epidemiological studies(55). A previous study has used matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (MS) to detect bacterial subspecies in pure cultures from clinical strains where they identified up to more than 80% of the subspecies of clinical isolates in accordance with 16s rRNA sequencing results(56). In another study, it was concluded that MALDI-TOF MS could be implemented as a fast and accurate method to identify the anaerobic bacteria isolated from clinical specimens(57). This type of MS analysis utilizes signature mass spectra generated from bacterial proteins and peptides that are characteristic of each microbe and identify the species by matching these signature spectra against pre-existing spectral libraries(90). Tandem time of flight (TOF/TOF) MS, however, generates peptide-specific spectra through multi-phase selection of proteinogenic peptides using a data-dependent acquisition algorithm(91). In tandem MS (MS/MS), peptides pass through a mass spectrometer a first time to measure their mass-to-charge (m/z) ratios, and then further divided into smaller fragments which their m/z values are also measured and searched against a peptide database for sequence matching and protein identification, a process known as "*de novo* peptide sequencing". Whereas MS/MS provides

more depth and higher efficiency for complex mixtures compared to conventional MS biotyping, it is more costly and time-consuming(59).

Rationale

This study aims to explore the capability of MALDI tandem time-of-flight mass spectrometry, MALDI-TOF/TOF MS, to detect *F. nucleatum* subspecies directly from clinical saliva samples. Further development of this approach could provide a novel non-invasive method for the CRC detection or follow-up screening, and hence comes the need for such a study.

2. Methodology

2.1 MALDI-TOF MS

MS is an analytical technique that is used to determine the exact mass of analytes, resulting in mass spectra, which are schemes of the ions signals represented as abundance versus m/z ratio. These mass spectra are of multiple uses, including, but not limited to, determining the identity or structure of an unknown molecule(92). In a typical mass spectrometer, the sample is ionized by increasing the internal energy of the molecules in an "ionization source" forming ions that then travel through a specific distance inside a "mass analyzer" with their m/z ratio determined by specific physical properties such as their velocity or response to the electric and/or magnetic fields. The ions are then received by a detection device and translated into mass spectra(93).

2.1.1 Ionization source

In MS, molecular entities could be captured and “detected” by a detector only if they are electrically charged, thus the analyte must first be ionized. These ions are called “parent” or “precursor” ions. Ionization sources can be divided into hard and soft(94). While hard ionization techniques break covalent bonds owing to the high energy imparted on the molecules, soft ionization techniques maintain the covalent structure of the analyte(93). A widely used example of hard ionization is electron ionization (EI), in which high energy is transferred to the sample molecules, resulting in variously sized ”fragment” or “product” ions(95). Soft ionization, on the other hand, includes chemical ionization (CI)(96), electrospray ionization (ESI)(97), atmospheric pressure chemical ionization (APCI)(98,99), atmospheric pressure photoionization (APPI) and MALDI(100). Nowadays, MALDI has become one of the most significant ionization methods in clinical laboratories(101). In MALDI, the analyte is mixed in solution with an ultraviolet (UV) irradiation-absorbing chromophore, known as the “matrix”, dried and co-crystallized(90). Laser energy is then applied to the sample and almost entirely absorbed by the matrix. The energy is then transferred to the analyte molecules, inducing excitation and chemical instability resulting in the loss or absorption of one or more protons and/or other charged species forming precursor ions(102). It is worth mentioning that multiple frameworks of MALDI mechanism have been suggested, with the exact one remaining debatable(103–106).

2.1.2 Mass analyzer

The mass analyzer, or mass filter, is responsible for filtering, separating, and occasionally fragmenting the traveling ions into fragment or product ions(101). In clinical laboratories, time-of-flight (TOF), quadrupole and ion traps are the most widely used mass analyzers(107,108).

In a TOF analyzer, ions are accelerated through an electrostatic field and separated based on their velocity, which is dependable on their charge and mass. Given that the voltage (energy) and the tube length (flight distance) are constant, the mass-to-charge ratio is calculated through its proportion with the square of the time-of-flight (109). When combined with MALDI, TOF becomes especially useful in analyzing macromolecules since it has a hypothetically unlimited mass range(110). Additionally, TOF operates in pulses, which makes it suitable to be used with pulsating ionization methods such as MALDI(111).

2.1.3 MALDI-TOF MS usage in microbial identification

MALDI-TOF MS has been used in multiple clinical studies for identifying bacterial or fungal colonies(112). The microbes were initially cultured on agar plates, left to grow colonies, and then the colonies of interest were picked with a sterile tool, e.g., toothpick, pipette tip or a plastic loop, and applied directly to the metal target plate of the MALDI instrument(113). The colonies were then coated and mixed with the matrix solution, e.g., alpha-cyano-4-hydroxy-cinnamic acid (CHCA) in acetonitrile (ACN) and trifluoroacetic acid (TFA) or 2,5-dihydroxybenzoic acid (DHBA) in water, ethanol, and ACN, and left to air dry. In case of analyzing microbes of a specific biosafety concerns, the colonies were pre-treated with one of the matrix solvent components, e.g., ethanol (114), or by the matrix solvent itself(115). When

the spots were dry, the target plate was loaded into the MALDI instrument to run the analysis and generate the mass spectra. The spectra were then searched against a reference spectral database provided by the manufacturer and, in some cases, amended/modified by the laboratory through adding custom strains. A matching score was annotated to the matched spectra by the manufacturer-provided software by comparing the spectra with the reference in the spectral library(101). Despite the evident advantages of this technique in identifying bacteria, such as the rapidity of the analysis that takes almost 5 minutes per sample(116), as well as the capability to identify the bacteria that is close to that of 16S rRNA sequencing, it still has its limitations especially when it comes to database-dependant identification(117–121).

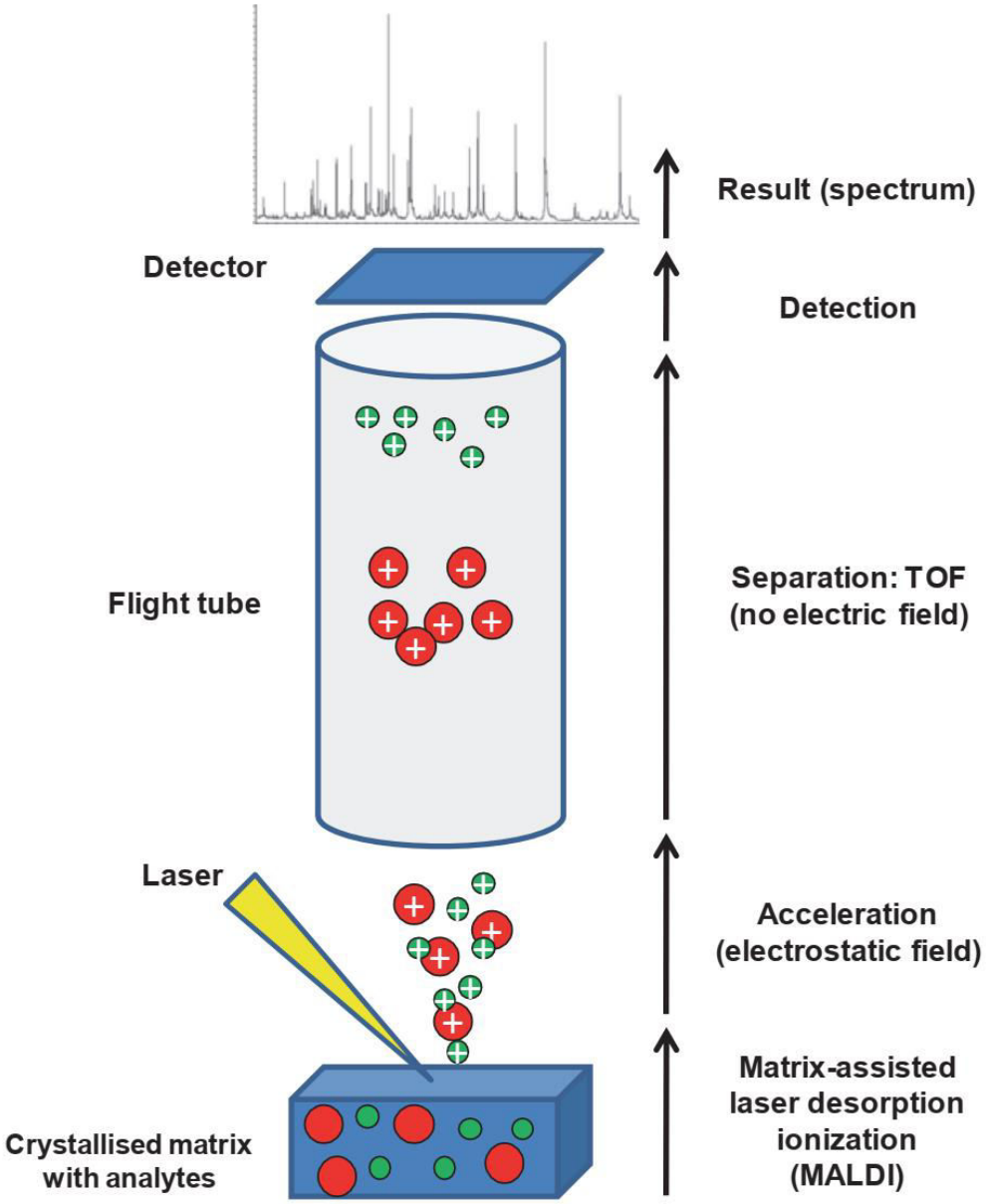

Figure 2. 1 MALDI-TOF MS technical principle of operation. Sample is mixed with a laser-absorbing matrix which facilitates the ionization. Precursor ions are accelerated towards the detector due to a high-voltage electric field, and travel along the flight tube. The ion then hits the detector generating a signal which, along with other travelling ions, is later translated into a mass spectrum. The mass-to-charge (m/z) ratio of an ion is determined from the time it spent to complete the flight tube, known as the time-of-flight. The figure is adopted from(90)**.**

2.1.4 Tandem time-of-flight mass spectrometry (TOF/TOF MS):

The main aim of tandem MS is to further fragment the resulting ions (precursor or parent ions) into smaller fragments or ions (product or daughter ions), which could later be more in-depth analyzed(101,122). Among the ion fragmentation methods available, collision-induced dissociation (CID) and post-source dissociation (PSD) are frequently used in clinical laboratories(123). In CID, parent ions are subjected to high gas pressure and multiple collisions with the molecules of an inert gas, such as nitrogen, argon or helium, in a collision cell that eventually leads to the breakdown of the parent ions bonds and formation of fragment ions. Whereas in PSD, ions receive high excess energies by the ionization source and are consequently destabilized and broken down into fragments immediately post-ionization(124,125). The Autoflex™ Speed instrument (Bruker Daltonics, Germany) that was used in this study implements the PSD fragmentation method. Tandem MS can operate in two configurations: tandem in space and tandem in time. In the former, multiple phases of fragmentation occur within two or more mass analyzers that are distinct and physically separated. Transmission quadrupoles, times-of-flight, or sectors are examples of these elements. On the other hand, in tandem in time, same as what was used in this thesis, a single mass analyzer is used with the data acquisition steps for precursor and product ions conducted in two different but consecutive time intervals, i.e. separated in time rather than in space. Here, the mass spectrometer performs a “precursor ion scan”, selects precursor ions to fragment upon the given criteria during the first time interval, and then performs a “product ion scan” for every precursor ion selected in the previous scan at the subsequent time intervals(126).

Tandem MS is currently considered one of the most efficient methods to identify and characterize proteins in complex mixtures(127). A plethora of studies have reported the implementation of tandem MS for the identification of microorganisms(128–131).

2.2 Preparation of biological samples

Sample preparation, including proteins extraction, clean-up and digestion, is a crucial step to ensure obtaining superior quality MS data(132).

In our experiments, we initially centrifuged human saliva to remove cell debris, insoluble materials, as well as potential solid contaminants. This was followed by protein precipitation to purify and concentrate the protein content of the sample and protein fractionation to reduce the complexity of the proteomics mixture.

2.2.1 Protein fractionation using solid-phase extraction

Solid-phase extraction (SPE) is a frequently applied technique for sample preparation in MS-based proteomics. In this technique, proteins are separated from impurities and collected in multiple fractions based on their different physicochemical attributes and/or their affinity to the sorbent. The sample is dissolved in the liquid (mobile) phase and passed through a column or cartridge filled with the solid (stationary) phase. First, sample constituents with no affinity for the solid phase are washed away then those bound to the sorbent are eluted from the column using an appropriate solvent. The solvent is later evaporated and the collected analyte is reconstituted in the desired solvent(133). In this study, to separate the proteins, we used reverse-phase C4 SPE spin-columns, which can separate analytes according to their affinity to a

non-polar four-carbon (C4) hydrocarbon chain fused to silicon-based microparticles. Short chains (C4 and C8) SPE sorbents are often of choice for protein separation. Columns with larger hydrocarbon chains, such as C18 however, are preferred for smaller molecules, e.g., peptides and amphiphilic organic compounds (134,135). SPE fractionation is an efficient approach to reduce the complexity of biological mixtures, but it increases the amount of time and labor required to conduct and analysis and reduces sample recovery due to inefficient bonding, solubility issues and liquid handling(136).

2.2.2 Protein digestion and desalting of the digest mixture

Protein digestion is a critically important step in samples preparation for MS analysis. The purpose of digestion is to generate molecular fragments, i.e. peptides, with appropriate size that is suitable for the MS analysis(137). Trypsin is the most commonly used enzyme for protein digestion in proteomics owing to its reasonable cost and its high cleavage specificity and reproducibility(138). It is an endopeptidase that acts by cleavage at the C-terminus of preferably arginine and lysine residues(139,140). We have used sequencing grade modified trypsin, which is specifically modified by reductive methylation of its lysine residues for optimal stability and autolysis prevention(141). Unspecific cleavage, and consequently decreased reproducibility of the analysis, could result from the usage of less pure or unmodified trypsin in MS experiments(137).

The next important step in sample preparation is the clean-up. This step aims to further purify and concentrate the digest (peptide) mixture by eliminating salts, buffers, or detergents

used during the previous preparation steps. For this purpose, we used the commercially available ZipTip®C18 pipette tips (Millipore, Ireland), which are 10-µL tips filled with C18 SPE sorbent, devoid of dead volume. The sample is aspirated through the tips and recovered directly into the (MALDI) matrix solution, yielding cleaner and higher quality MS data(138).

2.3 Samples collection

2.3.1 Patients/controls selection

The study participants were recruited as part of an ongoing larger study that aims to investigate the relationship between *Fusobacterium nucleatum* and CRC. All the subjects were within the age range of 40-80 years old with no previous cancer history, familial polyposis, or inflammatory bowel disease and had not been on antibiotic therapy for at least three months before colonoscopy. The individuals were examined by colonoscopy at the Endoscopy Department of *Centre hospitalier de l'Université de Montréal (CHUM)*. Based on the pathologically confirmed screening results, whether it yielded a positive diagnosis of colorectal adenoma, participants were classified into a colorectal-tumor group and a colorectal-tumor-free, i.e., control, group.

2.3.2 Saliva sample collection

Saliva samples of the participants were collected after obtaining their informed consent, given that they had not history of antibiotic therapy in the past three months. In average, the saliva sample production required 2-5 minutes, and for xerostomic (dry mouth) individuals, it required more time, up to 30 minutes. The individual was instructed to leave 30 minutes gap between giving the sample and eating, drinking, smoking, or gum chewing and then accumulate

saliva in the oral cavity for 60 seconds at least. A labeled sterile plastic 50-mL conical collection tube with a flat-top screw cap (Falcon, Fisher Scientific, USA) was opened, and the donor was instructed to spit into the tube. This process could be repeated until the liquid saliva reached the 5-mL line (excluding the foam). The lid was then screwed onto the tube securely, and the fresh samples were kept on ice and then maintained at -80°C pending MALDI-MS analysis.

2.4 Data analysis and proteins identification

Protein digestion using a specific proteolytic enzyme, e.g., trypsin, leads to protein fragments, i.e., peptides, that are specific to each protein (proteogenic peptides), with their molecular masses providing a unique set of data from which their tentative amino acid sequence could be inferred. The mass spectra matching m/z values of theoretical peptide sequences listed in a searched database are called “peptide spectral matches” (PSMs). The computer software then concatenates PSMs from a same protein to identify a potential “protein hit”. This computer-assisted operation is named peptide mass fingerprinting (PMF)(142). The matching degree between the exact mass of a peptide measured by the MS instrument and the theoretical mass of that peptide in the database is indicated by a software-assigned “protein score”. PSMs are then ranked in the order of their scores, i.e., the closeness of their spectra to the database (theoretical) ones. Afterward, according to the user-defined parameters and thresholds, the top scoring peptides are further analyzed to obtain the best “hits.” If no matches were found, the protein, therefore, remains unidentified(143).

In this study, we employed tandem MS to identify proteins, which generates two types of mass spectra. First, the spectra resulting from the initial precursor ion scan which are then

used to choose which ions are to fragment and subject to the product ion scan based on precursor ion signal intensity. This process is known as data-dependant acquisition (DDA)-MS/MS(144,145). A top-ten DDA method was used in this study, i.e., only the 10 most abundant precursor ions were fragmented and characterized by MS/MS, which could potentially mean that only the most abundant proteins were showed in the list of identified proteins. It is worth mentioning that, in theory, the MS/MS analysis could be repeated multiple times on *de novo* product ions (MS^n analysis). However, MS^2 (MS/MS) is usually sufficient for peptide identification(146).

Through bioinformatics analyses, robust and precise identification of the peptides and/or proteins is made possible using a variety of software bundles, search engines, and algorithms(143). Free search engines are available for PMF, e.g., MS-Fit (Protein Prospector, USA), Aldente (ExPASy, Switzerland), and Mascot (Matrix Science, UK). The latter was used in this study. Likewise, there are freely available protein databases for proteomics research, e.g., NCBI, and MSDB. We have chosen UniProt protein database, owing to its comprehensiveness and low redundancy(143).

2.4.1 Mascot score

Mascot score is a probability-based scoring system used to interpret the quality of the Mascot software's search results. It is calculated from -10 Log (**P**), where **P** is the probability that the matching between the MS spectra and the database is a random event(147). For example, a statistical *p*-value of 0.05 as significance threshold would be corresponding to a

score of 13. To further enhance our results, our Mascot score cut-off was set to 20, which is equivalent to a significance *p*-value of 0.01.

3. Materials and Methods

3.1 Bacterial strains growth and culturing

Fusobacterium strains were obtained from a gracious grant from Prof. Joong-Ki Kook, Korean Collection for Oral Microbiology (KCOM) and the Department of Oral Biochemistry, College of Dentistry, Chosun University, China, in the form of freeze-dried bacteria in glass vials. The strains used were as follows: KCOM 2763, *F. nucleatum* subsp. *animalis*; KCOM 1231, subsp. *vincentii*; KCOM 1232, subsp. *polymorphum*; KCOM 1323, subsp. *nucleatum*. The strains were revived by resuspending in 2 mL of sterile water. The suspensions were then sub-cultured on Anaerobic Basal Agar (Oxoid, Thermo Fisher Scientific, USA) in Petri dishes with 5% defibrinated sheep blood at 37°C in an anaerobic chamber (Vinyl Anaerobic Airlock Chamber, Coy, USA), under regulated gas concentration: 87% N_2, 10% CO_2, and 3% H_2, for 2-7 days until evident growth of bacterial colonies.

Liquid cultures of these subspecies were then prepared in Anaerobic Basal Broth (Oxoid) by inoculating 3-5 colonies and incubating in an anaerobic incubator at 37°C for 24-72 h as needed. Diff-Quik™ (Modified Giemsa, Siemens, Germany) stains were used to visualize the bacteria with light microscopy (Primo Star, Zeiss, Germany). The bacterial cultures were then maintained by freezing 500 mL of the liquid culture added to 500 mL of 50% glycerol at -80°C storage prior to MALDI-MS analysis. All the media used were reduced at the chamber for at least 48 hours before usage.

3.3 Bicinchoninic acid (BCA) assay

To obtain an estimate of protein concentration in the saliva samples, 5 mL of saliva was collected from a healthy volunteer lab colleague that did not consume food in 30 minutes and had not consumed antibiotics in three months, into a labelled 50 mL conical collection tube (Falcon) and was maintained at -80°C overnight. The BCA assay was performed using Pierce™ BCA Protein Assay Kit (Thermo Scientific) and following the manufacturer's instruction. First, in a Thermo Scientific™ Pierce™ 96-Well Plate, a series of diluted bovine serum albumin (BSA) standard solutions of known concentrations were prepared in water as reference standards to establish a calibration curve. 20 µL of BSA standard was added in 9 wells as follows; 2000 µg/mL, 1500 µg/mL, 1250 µg/mL, 1000 µg/mL, 750 µg/mL, 500 µg/mL, 250 µg/mL, 125 µg/mL, 62.5 µg/mL, and 31.25 µg/mL. Next, saliva specimen was thawed on ice, then 0.5 mL of thawed saliva was transferred into a 1.5-mL microcentrifuge tube and spun at 1500 g for 15 min (4°C), and the supernatant was transferred into a clean tube. After that, three samples of the supernatant were used as follows; 20 µL per well (undiluted), 4 µL in 16 µL water (5 times diluted), 2 µL in 18 µL water (10 times diluted). These different sample diluents were measured in triplicates. To each well, in addition to a two control wells, 200 µL of BCA working reagent was added. The reagent was prepared using 80 µL of BCA reagent B (copper (II) sulfate) to a 4 mL of BCA reagent A (bicinchoninic acid, 1 : 50 v/v). The microplate was covered, shook on a plate shaker for 30 seconds, and then incubated at 37°C for 30 minutes. Next, it was left to cool at room temperature and then read at 562 nm on a microplate reader. Using the established standard curve, the BCA assay indicated a total protein concentration of 1422 µg/mL saliva.

3.4 MS analysis

3.4.1 Saliva proteome extraction and preparation

To prepare the saliva samples for MALDI-TOF/TOF MS analysis, the protocol used was derived from Dr. Momar Ndao's laboratory protocols, Research Institute of the McGill University Health Centre (RI-MUHC). First, the saliva specimen was thawed on ice, then 0.5 mL of thawed saliva was transferred into a 1.5-mL microcentrifuge tube and spun at 1500 g for 15 min (4°C) to remove the cell debris, solid contaminants or insoluble materials, as described earlier under section 2.2. After that, 0.35 mL of the supernatant (containing approximately 500 µg protein as assessed by the previously described BCA assay) was transferred into another clean microcentrifuge tube. For protein precipitation, 0.7 mL of ice-cold 10% w/v trichloroacetic acid in 90% v/v acetone was then added to the saliva supernatant and kept for at least two hours at -20°C. The samples were then spun at 18,000 g for 5 min (4°C) to obtain a protein pellet, and the supernatant was discarded. The resultant pellet was washed by ice-cold acetone using 0.5 mL by vigorous mixing for 20 seconds, spun at 18000 g for 4 min (4°C). The wash step was repeated, and the pellet was left to air dry at room temperature for 10 min. The pellet was then resuspended in 0.25 mL 0.1 M ammonium bicarbonate (ABC) pH 7.5-8.0 using vigorous mixing for at least 30 seconds. A digital ultrasound homogenizer (Misonix S-4000, USA) was used to fully resuspend the precipitated salivary protein pellet at an amplitude of 3% for a total of 2 min (30 seconds vibration, 30 seconds rest, two cycles) or until the pellet was completely resuspended.

3.4.2 Protein fractionation

Protein fractionation was done using reverse-phase (C4) SPE as follows: four elutions were performed with four different solutions to yield four protein fractionates (1-4) in total per sample: 5%, 20%, 50% and 80% acetonitrile (ACN) in 0.1% TFA to gradually decrease eluent polarity. The procedure was done using four C4 Macro SpinColumns™ (Harvard Apparatus, USA) per sample.The column was gently shook to ensure the sorbent is settled at the bottom, their caps were removed and the columns were transferred into 2-mL microcentrifuge tubes. The columns were rehydrated with 0.5 mL 100% ACN for 15 min and centrifuged for 4 min at 2000 g to remove the solvent. The columns were then equilibrated with 0.5 mL 0.1 M ABC for another 15 min and centrifuged for 4 min at 2000 g. Next, the saliva specimen prepared in the previous step (in 250 µL ABC) was loaded onto the SPE column and spun at 2000 g for 4 min to bind the proteins, and the eluate (unbound fraction) was discarded. After that, to obtain the most polar fractionate, 150 µL of 5% ACN in 0.1% TFA was added to the SPE column, centrifuged at 2000 g for 4 min and the eluate was collected. This step was repeated with the 3 previously mentioned elution solutions into 3 other clean tubes to yield four fractions in total per sample. The fractions were then evaporated under vacuum using a centrifugal vacuum concentrator (Labconco, USA) at 40°C for 2 h or until samples were dry.

3.4.3 Trypsin digestion

To each fraction, 150 µL 0.1 M ABC was added, incubated with 10 µL 0.1 M dithiothreitol in 0.1 M ABC for 30 min at 60-70°C to denature proteins by reducing their disulfide bonds between cysteine residues. To alkylate reduced cysteine residues, fractions

were then incubated in the dark for another 30 min at room temperature with 15 µL 0.25 M iodoacetamide (IAM) in 0.1 M ABC. The pH of the solution was checked and adjusted to 7.5-8.5 if deemed necessary, using 5% ammonium hydroxide to ensure optimal activity of trypsin. To each fraction, 2.5 µg (25 µl 0.1 µg/mL trypsin in 50 mM acetic acid) of sequencing grade modified trypsin (Promega™, USA) was then added and incubated overnight (16–20 h) at 37°C, resulting in a total trypsin : protein ratio of 1 : 50 w/w. The digestion reaction was stopped by adding 10 µL or more of 1% TFA, ensuring the pH was at or slightly below 6.0.

3.4.4 Desalting and spotting on MALDI target plate

Each digest clean-up was done using reverse-phase (C18) SPE utilizing ZipTip® pipette tips (Millipore). First, 60 µL of ACN was added in a 1.5-mL microcentrifuge tube and the SPE material was conditioned by rinsing 5 times with 10 µL ACN. Next, 60 µL of 0.1% TFA was added in a 1.5-mL tube, and the ZipTip® was re-hydrated by rinsing 5 times with 10 µL of this solution. After that, the sample (one saliva fraction) was loaded onto the C18 sorbent by slowly aspirating (up and down pipetting) 10 µL of the sample 10 times. The ZipTip® was then rinsed with 10 µL 0.1% TFA 5 times to remove unbound compounds. MALDI-MS matrix solution was prepared using 10 mg/ml α-cyano-4-hydroxycinnamic acid (CHCA) in 0.1% TFA in 50% ACN, vortexed for 15 minutes and centrifuged at 5000 g for 10 min. 8 µL of the matrix solution was added in a 0.6-mL microcentrifuge tube, and the loaded ZipTip® was eluted into this solution by aspirating 10 times to ensure proper mixing between the sample and the matrix to maximize co-crystallization. Care was taken not to introduce air into the ZipTip® at any point of the

procedure. The ZipTip® was then discarded, and 1.5 µL of the mixed matrix/sample solution was spotted on a clean MTP 384 polished steel target plate (Bruker Daltonics). Each saliva fraction was spotted 4 times (N=4) to correct for reproducibility issues. 1.5 µL of a positive-mode electrospray ionization calibration solution (Agilent Technologies, USA) was also spotted 4 times (N=4) to calibrate the mass spectrometer prior to the analysis. The calibration solution was a commercial mixture of 6 synthetic peptides with signals at m/z 904.4681, 1296.6853, 1570.6774, 2093.0867, 2465.1989, 3657.9294. The spots were left to air dry at room temperature. The remove potential contaminants, the target plate was cleaned using Wenol® metal polish (Reckitt Benckiser, Germany), cleaning solution from MALDI plate cleaning kit (Applied Biosystems, USA) and an electronic duster (Dust-Off®, Falcon, USA) prior to the MALDI-MS analysis.

3.4.5 MALDI-TOF/TOF MS analysis

The analysis was run using Autoflex™ Speed MS instrument (Bruker Daltonics) using BioTools software version 3.2 (Bruker Daltonics). FlexControl software version 3.4 (Bruker Daltonics) was used for data acquisition, FlexAnalysis software version 3.4 (Bruker Daltonics) was used for data processing. Out of the four spots in total per saliva fraction using a positive-mode ionization reflectron TOF acquisition method for m/z 700-4000 and a top-ten DDA TOF/TOF analysis, i.e. 10 MS/MS analysis. In positive-mode ionization molecules absorb one or more protons to form $[M+nH]^{n+}$ ions. The reflectron TOF (reTOF) mode operates the TOF analyzer at the highest resolution and mass accuracy but limits mass range to 6000 Da, which is deemed sufficient for the analysis of peptides. The other operation mode is linear TOF which has unlimited mass range, and is thus an ideal mode for intact (non-digested) proteins but with

considerably lower resolution and mass accuracy. After data acquisition was completed, the spectra were visually inspected in FlexAnalysis software and the spot with the maximum number of peaks was selected out of each quadruplicate for PMF analysis. To do so, the MS spectrum from the selected spot was merged with the corresponding 40 (= 4 x 10) MS/MS spectra acquired from all four spots within the BioTools software. In the case of two equal peaks numbers, the spot with the highest signal-to-noise ratio was chosen. The instrument was calibrated to measure exact mass within 1 ppm accuracy. To identify the proteins, the database search was done in BioTools software, using Mascot search engine for PMF protein search against the *Fusobacterium nucleatum* subset of the UniProt protein database (17.2 MB as of 15-05-2019, containing 84,375 reviewed and unreviewed protein sequences) or *Fusobacterium* (40.3 MB as of 15-05-2019, containing 197,493 reviewed and unreviewed protein sequences). Mass tolerance was set to 200 ppm for MS spectra and 0.4-0.6 Da for the MS/MS ions search.

3.4.6 Bacterial cultures analysis

The previously mentioned *Fusobacterium* strains, KCOM 2763, *F. nucleatum* subsp. *animalis*; KCOM 1231, subsp. *vincentii*; KCOM 1232, subsp. *polymorphum*; KCOM 1323, subsp. *nucleatum* were grown in liquid cultures from previously prepared glycerol stocks. A total of 1 mL mixture of the four strains' liquid cultures was made using 250 µL of each strain. This mixture was prepared for MALDI-MS analysis as described for saliva samples and used as positive control along with the patient and control saliva specimens.

3.5 Experimental design

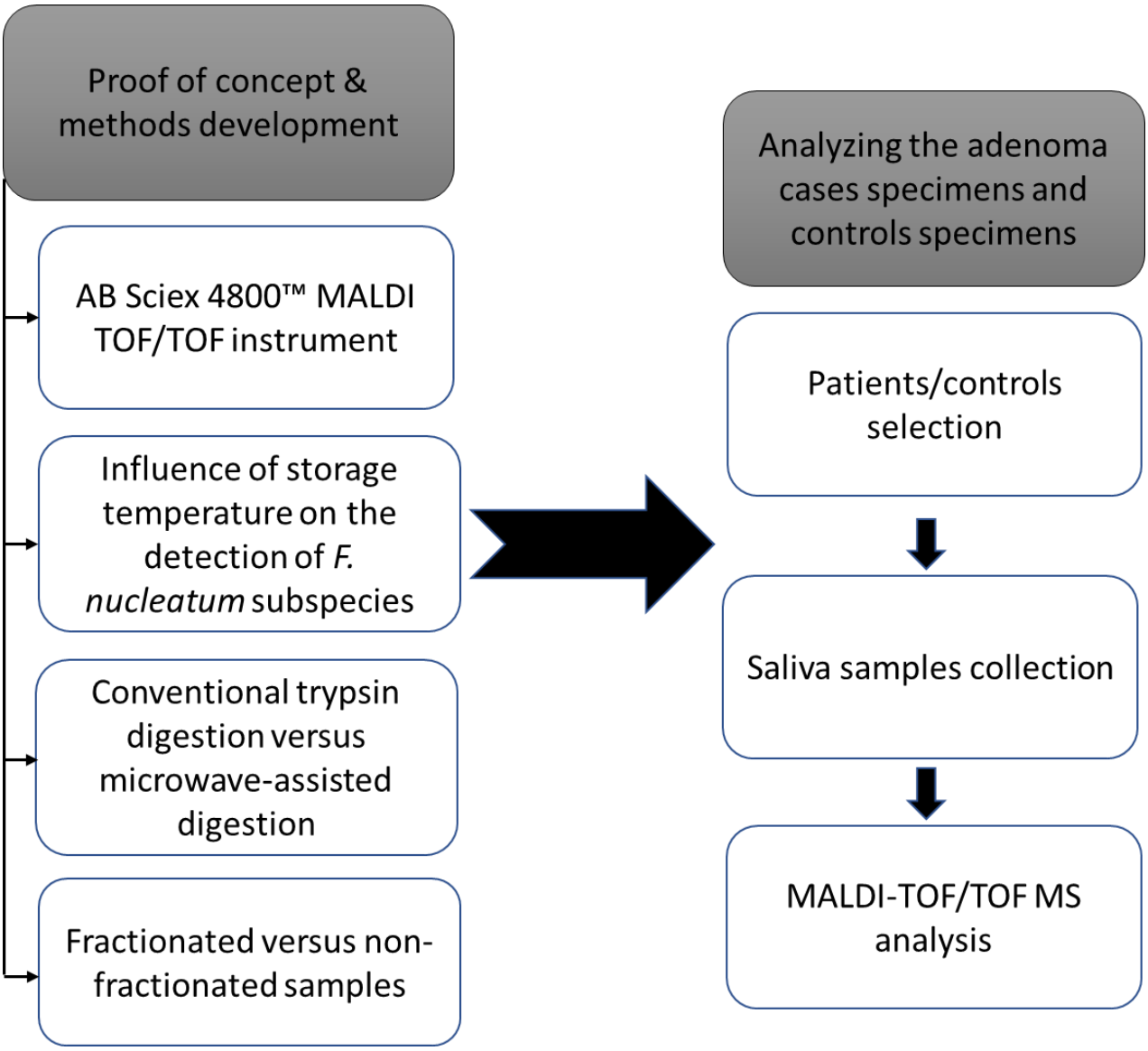

Figure 3. 1 Flowchart showing the experimental design for this study.

3.5.1 Proof of concept and methods development

3.5.1.1 Initial method development with AB Sciex 4800™ MALDI TOF/TOF instrument

In this experiment, we have used 4800 Plus MALDI TOF/TOF™ Analyzer (Applied Biosystems, USA) to detect *F. nucleatum* subspecies in a volunteer saliva sample. A 5 mL of saliva was collected and stored overnight in -80 °C, and the analysis was run using the same discussed experimental approach in section 3.4. The proteins search was done using ProteinPilot™ v2.0 software (SCIEX, USA) against UniProt database.

3.5.1.2 Testing the influence of storage temperature on the detection of F. nucleatum subspecies

A total of 5 mL of saliva was collected twice from a healthy volunteer. The volunteer did not consume food in 30 minute prior to saliva sampling and had no antibiotics treatment history in the past three months. The sampling was done by spitting into two labelled 50 mL conical collection tube (Falcon) - 5 mL per tube. One tube was stored overnight at -80°C and the other was maintained overnight at 4°C. The analysis was performed in triplicate for each of the saliva samples and protein search was done for *Fusobacterium* genus. The two samples were then analyzed by MALDI-MS using AutoFlex Speed™ instrument with the previously described protocol.

3.5.1.3 Conventional trypsin digestion versus microwave-assisted digestion

For proteomics analysis, trypsin digestion is usually performed overnight, i.e. 16–20 h at 37°C. To shorten the analysis time for an increased throughput, a series of experiments were conducted to test trypsin digestion efficiency under microwave (MW)-irradiation conditions. MW-assisted enzymatic digestion (MAED) has been shown to improve the quality of proteomics analysis(148–150) . Four samples of 50 µL 1 mg/mL bovine serum albumin (BSA) in 0.1 M ABC pH 7.5 were prepared. Three were subjected to MAED with different MW power (500-1100 W, 1 min) in a conventional MW cooking oven (Kenmore, USA), and one sample was digested using the conventional digestion conditions. All samples were analyzed in duplicate and spotted for MALDI analysis in quadruplicate (N = 2 x 4 = 8). Per each sample, the spot with the highest MS peak number was selected for PMF and then combined with all the corresponding MS/MS peaks to search against the UniProt database as described in the methods section.

In each microwave run, an additional Eppendorf tube were added with a mercury thermometer sealed to the tube and with its tip immersed in a 50 µL 1 M ABC to measure the temperature after each run. The four samples were then de-salted using ZipTips™ pipette tips and spotted on MALDI target plate (4 spots per sample), loaded into Autoflex™ Speed instrument, and the MADLI-MS was run as described before, searching against UniProt database, using “Other Mammalians” as the class of protein search in BioTools software version 3.2 (Bruker Daltonics) and Mascot search algorithm. FlexAnalysis version 3.4 (Bruker Daltonics) was used to inspect the spectral quality.

3.5.1.4 Comparing fractionated versus non-fractionated samples

A total of 5 mL of saliva was collected twice from a healthy volunteer lab member that did not consume food in 30 minutes and had not consumed antibiotics in three months, into two labelled 50 mL conical collection tube (Falcon) - 5 mL per tube. The two tubes were maintained at -80°C overnight. For MALDI-MS analysis, one sample was prepared using the protocol described before. The other sample was analyzed using the same protocol initially but was not subjected to reverse-phase (C4) solid-phase extraction (SPE) protein fractionation, but rather the saliva supernatant (350 µL) was split into 4 equal parts into four 2 mL Eppendorf tubes and vaporized in the Speed-vac. The rest of the previously described protocol was followed as is.

3.5.2 Analyzing the adenoma cases specimens and controls specimens

A total of 4 patients with colorectal adenomatous polyps, and 8 non-adenoma controls saliva samples were analyzed. One bacterial culture made of the mix of the four strains as described before, was used as a positive control. Each sample had 4 fractions in total, out of which 3 were spotted in quadruplicates. Patients and controls data are shown in **Table A.1**.

The MALDI-MS analysis was done as described before, with the search in Biotools (Bruker Daltonics) software was doe using Mascot in UniProt database, using "*Fusobacterium nucleatum*" query.

4.Results:

4.1 Proof of concept and methods development

4.1.1 Initial method development with AB Sciex 4800™ MALDI TOF/TOF instrument

The experiment was conducted on one 5 mL saliva sample, maintained in -80 °C storage overnight. The proteomics analysis was performed as described in the materials and methods section, and protein identification was made by ProteinPilot™ software version 2.0 (Sciex, Canada) against the *Fusobacterium* subset of the UniProt protein database using the Paragon protein search algorithm(151) for trypsin digestion and iodoacetamide for cysteine alkylation. We were able to detect a total of 15 proteins that belonged to the *Fusobacterium* genus (**Table A.2**). This experiment showed the applicability of the MALDI-MS analysis to distinguish *F. nucleatum* down to the subspecies level. The developed method was later transferred to an Autoflex Speed (Bruker) MALDI-TOF/TOF MS instrument for improved throughput and sensitivity.

4.1.2 Testing the influence of storage temperature on the detection of *F. nucleatum* subspecies

In this experiment, two saliva samples of the same donor were stored overnight at different storage temperatures, 4°C, and -80°C, respectively. MALDI-MS analysis was run as described in the materials and methods section. BioTools software 3.2 (Bruker Daltonics) was used for proteomics analysis utilizing a Mascot search engine (Matrix Science) for protein

identification. The searches were done on UniProt protein database (*Fusobacterium* subset), which included 197,493 reviewed and unreviewed protein sequences. The Mascot score cut off value used was 20, equivalent to a significance *P* value of < 0.01. All the proteins identified scored above 40 on Mascot score, indicating confident hits (**Table 4.1** and **Table 4.2**). *F. nucleatum* subspecies were detected in both samples, noting the lack of significant influence of storage temperature on protein detection using MALDI-TOF/TOF MS analysis. These results also showed the applicability of the MALDI-TOF/TOF MS approach to detect *F. nucleatum* down to the subspecies level.

Protein accession and name	Score	Organism
tr|A0A133NJ24|A0A133NJ24_9FUSO Uncharacterized protein (Fragment)	45	*Fusobacterium equinum*
tr|A0A2G9EFZ2|A0A2G9EFZ2 Polyribonucleotide nucleotidyltransferase	72	*Fusobacterium periodonticum*
tr|A0A2G9EC77|A0A2G9EC77_9FUSO Polyribonucleotide nucleotidyltransferase	72	*Fusobacterium periodonticum*
tr|A0A2D3PLL3|A0A2D3PLL3_9FUSO Polyribonucleotide nucleotidyltransferase	72	*Fusobacterium periodonticum*
tr|A0A2D3M1N1|A0A2D3M1N1_9FUSO Polyribonucleotide nucleotidyltransferase	65	*Fusobacterium periodonticum*
tr|A0A2D3NW42|A0A2D3NW42_9FUSO Polyribonucleotide nucleotidyltransferase	65	*Fusobacterium periodonticum*
tr|A0A2D3PW83|A0A2D3PW83_9FUSO Polyribonucleotide nucleotidyltransferase	58	*Fusobacterium periodonticum*
tr|D4CS73|D4CS73_9FUSO Polyribonucleotide nucleotidyltransferase	53	*Fusobacterium periodonticum* ATCC 33693
tr|K1GLF6|K1GLF6_9FUSO Polyribonucleotide nucleotidyltransferase	52	*Fusobacterium periodonticum* D10
tr|D6LJI6|D6LJI6_9FUSO Mass: Polyribonucleotide nucleotidyltransferase	45	*Fusobacterium periodonticum* 1_1_41FAA
tr|A0A095WG31|A0A095WG31_9FUSO Polyribonucleotide nucleotidyltransferase	59	*Fusobacterium periodonticum* 2_1_31
tr|A0A140NRJ4|A0A140NRJ4_FUSNU Uncharacterized protein	64	*Fusobacterium nucleatum subsp. animalis* 7_1
tr|U7TAG8|U7TAG8_FUSNU Uncharacterized protein	64	*Fusobacterium nucleatum* CTI-1
tr|Q7P8F5|Q7P8F5_FUSNV Uncharacterized protein	64	*Fusobacterium nucleatum subsp. vincentii* ATCC 49256
tr|A0A133NQT1|A0A133NQT1_FUSNU Uncharacterized protein	42	*Fusobacterium nucleatum*
tr|A0A117MW77|A0A117MW77_FUSNC Uncharacterized protein	41	*Fusobacterium nucleatum subsp. nucleatum*

Table 4. 1 List of *Fusobacterium* proteins detected by MALDI-TOF/TOF MS in saliva samples stored at 4°C. Mascot protein score = -10 Log (P), where P represents the probability of a protein hit to be a random event.

Protein accession and name	Score	Organism / strain name
tr|J5TNF0|J5TNF0_9FUSO 2-hydroxyglutaryl-CoA dehydratase, D-component	55	*Fusobacterium necrophorum* subsp. *funduliforme* Fnf 1007
tr|A0A2X3KA92|A0A2X3KA92_9FUSO (R)-2-hydroxyglutaryl-CoA dehydratase subunit alpha	55	*Fusobacterium necrophorum* subsp. *necrophorum*
tr|A0A064A9R0|A0A064A9R0_9FUSO R-phenyllactate dehydratase medium subunit	55	*Fusobacterium necrophorum* BFTR-2
tr|A0A162IIV9|A0A162IIV9_9FUSO 2-hydroxyacyl-CoA dehydratase	55	*Fusobacterium necrophorum* subsp. *funduliforme*
tr|A0A017H320|A0A017H320_9FUSO R-phenyllactate dehydratase medium subunit	55	*Fusobacterium necrophorum* subsp. *funduliforme* B35
tr|H1DBE4|H1DBE4_9FUSO Uncharacterized protein	55	*Fusobacterium necrophorum* subsp. *funduliforme* 1_1_36S
tr|A0A0E2V2D7|A0A0E2V2D7_9FUSO R-phenyllactate dehydratase medium subunit	55	*Fusobacterium necrophorum* DJ-2
tr|X7S779|X7S779_FUSNU Uncharacterized protein	52	*Fusobacterium nucleatum* 13_3C
tr|A0A241Q0P7|A0A241Q0P7_FUSNP Uncharacterized protein	52	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|A0A0S2ZHA3|A0A0S2ZHA3_9FUSO Uncharacterized protein	52	*Fusobacterium hwasookii* ChDC F300
tr|A0A0S2ZQY1|A0A0S2ZQY1_9FUSO Uncharacterized protein	52	*Fusobacterium hwasookii* ChDC F174
tr|X7RXU7|X7RXU7_FUSNU Uncharacterized	53	*Fusobacterium nucleatum* 13_3C
tr|A5TRG2|A5TRG2_Uncharacterized protein	53	*Fusobacterium nucleatum* subsp. *polymorphum* ATCC 10953
tr|A0A1F0DXX1|A0A1F0DXX1_9FUSO Flavodoxin	53	*Fusobacterium* sp. HMSC064B11
tr|A0A0S2ZVK0|A0A0S2ZVK0_FUSNP Flavodoxin	41	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|A0A241Q4N7|A0A241Q4N7_FUSNP Flavodoxin	41	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|X8I3R2|X8I3R2_9FUSO Flavodoxin domain protein	41	*Fusobacterium* sp. OBRC1
tr|A0A246EG78|A0A246EG78_FUSNP Flavodoxin	41	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|A0A1Z3CKB9|A0A1Z3CKB9_FUSNP Flavodoxin	41	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|H1D4D1|H1D4D1_9FUSO Chaperone htpG	62	*Fusobacterium necrophorum* subsp. *funduliforme* 1_1_36S

Table 4. 2 List of *Fusobacterium* proteins detected by MALDI-TOF/TOF MS in saliva samples stored at -80 °C. Mascot protein score = -10 Log (P), where P represents the probability of a protein hit to be a random event.

4.1.3 Conventional trypsin digestion versus microwave-assisted digestion

For each sample, the spot with the highest MS peaks number was picked for PMF and then combined with the resulting MS/MS peaks to be searched against the UniProt database as illustrated in the corresponding methods section.

The highest number of MS peaks, PSMs, and the best Mascot score resulted from Experiment #4, where temperature went up to 65°C using 880 W of MW power. When compared to conventional digestion, all MAED samples yielded a higher protein sequence coverage and hence superior performance over the conventional trypsin digestion (78–85% MAED vs 72% conventional digestion). The data are shown in **Table 4.3.**

Sample #	Maximum temp. °C	MW power (W)	Incubation time	Maximum # MS peaks	# PSMs (Mascot)	Mascot score	Protein sequence coverage (%)
1	37	0	16 h	59	17	70	72.5
2	45	550	1 min	49	20	11	78.0
3	54	1100	1 min	45	21	116	80.1
4	65	880	1 min	55	24	118	83.4

Table 4. 3 Analysis condition and proteomics analysis results for one conventionally digested sample (#1) and three microwave-assisted digested samples (#2-4). Mascot protein score = -10 Log (P), where P represents the probability of a protein hit to be a random event.

4.1.4 Comparing fractionated versus non-fractionated samples

Two saliva samples of the same donor were stored overnight at -80°C. MALDI-TOF/TOF MS analysis was run as described in the materials and methods section, with the modification that one saliva sample was not fractionated. Instead, the saliva supernatant was split into 4 equal portions and evaporated under vaccum. The rest of the protocol was followed as described before. BioTools software (Bruker Daltonics) was used for proteomics analysis utilizing Mascot search engine for protein identification. The protein search was done against UniProt protein database (*Fusobacterium nucleatum* subset), which included 188,063 reviewed and unreviewed protein sequences, with the previously described search parameters.

Proteomics analysis of the non-fractionated saliva sample yielded comparable results to those of the SPE-fractionated sample. Specifically, thirteen *F. nucleatum* proteins were detected in the fractionated sample versus fourteen in the whole lysate (**Table 4.4** and **Table 4.5**).

Protein accession and name	Score	Organism
tr|J8V657|J8V657_9FUSO 2',3'-cyclic nucleotide 2'-phosphodiesterase	41	*Fusobacterium hwasookii* ChDC
tr|F9EP85|F9EP85_FUSNU Cell division topological specificity factor	45	*Fusobacterium nucleatum* subsp. *animalis* ATCC 51191
tr|A0A323TUG0|A0A323TUG0_FUSNU MBL fold metallo-hydrolase	47	Fusobacterium nucleatum
tr|A0A3P1VSX9|A0A3P1VSX9_FUSNU GlcNAc transferas	52	*Fusobacterium nucleatum*
tr|A0A2B7Y5G9|A0A2B7Y5G9_FUSNP Autotransporter domain-containing protein	52	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|A0A1Z3CLX0|A0A1Z3CLX0_FUSNP GlcNAc transferase	38	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|A0A2C6C779|A0A2C6C779_FUSNP GlcNAc transferase	33	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|Q8REY2|Q8REY2_FUSNN Cobalamin biosynthesis protein G	56	*Fusobacterium nucleatum* subsp. *nucleatum*
tr|D5R9Z4|D5R9Z4_FUSN2 CbiG	42	*Fusobacterium nucleatum* subsp. *nucleatum* (strain ATCC 23726 / VPI 4351)
tr|A0A0M4SSI4|A0A0M4SSI4_FUSNC Cobalamin biosynthesis protein G	42	*Fusobacterium nucleatum* subsp. *nucleatum*
tr|Q8REY2|Q8REY2_FUSNN Cobalamin biosynthesis protein G	56	*Fusobacterium nucleatum* subsp. *nucleatum* (strain ATCC 25586 / CIP 101130 / JCM 8532 / LMG 13131)
tr|D5R9Z4|D5R9Z4_FUSN2 Cobalamin biosynthesis protein G	41	*Fusobacterium nucleatum* subsp. *nucleatum* (strain ATCC 23726 / VPI 4351)
tr|A0A0M4SSI4|A0A0M4SSI4_FUSNC Cobalamin biosynthesis protein G	41	*Fusobacterium nucleatum* subsp. *nucleatum*

Table 4.4 The list of *F. nucleatum* proteins detected using SPE fractionation in the saliva samples.

Protein accession and name	Score	Organism
tr|A0A140PV23|A0A140PV23_FUSNU Uncharacterized protein	57	*Fusobacterium nucleatum* subsp. *animalis* 7_1
tr|U7TBG4|U7TBG4_FUSNU Uncharacterized protein	57	*Fusobacterium nucleatum CTI-1*
tr|F7L3H5|F7L3H5_FUSNU Uncharacterized protein	50	Fusobacterium nucleatum subsp. animalis 11_3_
tr|H1HGF2|H1HGF2_FUSNU Uncharacterized protein	50	*Fusobacterium nucleatum* subsp. *animalis* F0419
tr|A0A0M4RME4|A0A0M4RME4_FUSNU Uncharacterized protein	43	*Fusobacterium nucleatum* subsp. *animalis*
tr|A0A1S1MFL2|A0A1S1MFL2_FUSNU Uncharacterized protein	43	*Fusobacterium nucleatum*
tr|A0A377GPB7|A0A377GPB7_FUSNV Subtilase family	43	*Fusobacterium nucleatum* subsp. *vincentii*
tr|R9R973|R9R973_FUSNU Uncharacterized protein	42	*Fusobacterium nucleatum* subsp. *animalis* 4_8
tr|D6BG23|D6BG23_FUSNU Uncharacterized protein	36	*Fusobacterium nucleatum* subsp. *animalis* D11
tr|A0A0M1VX85|A0A0M1VX85_FUSNV Uncharacterized protein	33	*Fusobacterium nucleatum* subsp. *vincentii* 4_1_13
tr|A0A0M4RYS6|A0A0M4RYS6_FUSNU Cupin	60	*Fusobacterium nucleatum* subsp. *animalis*
tr|Q7P2E6|Q7P2E6_FUSNV Hypothetical Cytosolic Protein	39	*Fusobacterium nucleatum* subsp. *vincentii* ATCC 49256
tr|A0A2B7YNZ0|A0A2B7YNZ0_FUSNP Cupin	39	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|A0A246EJ37|A0A246EJ37_FUSNP Cupin	39	*Fusobacterium nucleatum* subsp. *polymorphum*

Table 4.5 The list of *F. nucleatum* proteins detected using the whole lysates (no SPE fractionation) in the saliva samples preparation. The score indicates Mascot score

4.2 Analyzing saliva from adenoma patients and negative controls

As previously described in the materials and methods section, a mixture of 4 liquid cultures representing 4 genetically different *F. nucleatum* subspecies was used as a positive control in the MALDI-MS analysis. Twenty-three proteins from the 4 subspecies were successfully detected in the positive control, signifying the reliability of the protocol, as well as the instrument and parameters used to detect the *F. nucleatum* subspecies **(Table 4.6).**

Protein accession and name	Score	Organism/strain name
tr|X7RV25|X7RV25_FUSNU Uncharacterized protein	42	*Fusobacterium nucleatum* 13_3C
tr|A0A2C6CC76|A0A2C6CC76_FUSNP Uncharacterized protein	30	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|W3XR61|W3XR61_9FUSO Uncharacterized protein	26	*Fusobacterium sp.* CM21
tr|A0A2N6THD0|A0A2N6THD0_FUSNU Uncharacterized protein	43	Fusobacterium nucleatum CTI-6
tr|A0A246EE17|A0A246EE17_FUSNP Uncharacterized protein	43	Fusobacterium nucleatum subsp. polymorphum
tr|D6BFE3|D6BFE3_FUSNU Uncharacterized protein	31	Fusobacterium nucleatum subsp. animalis D11
tr|A0A0M1VSW0|A0A0M1VSW0_FUSNV Uncharacterized protein	29	Fusobacterium nucleatum subsp. vincentii 4_1_13
tr|Q8RF68|Q8RF68_FUSNN Hypothetical cytosolic protein	29	Fusobacterium nucleatum subsp. nucleatum (strain ATCC 25586 / CIP 101130 / JCM 8532 / LMG 13131)
tr|R9R9U6|R9R9U6_FUSNU Uncharacterized protein	29	Fusobacterium nucleatum subsp. animalis 4_8
tr|A0A3P1VVM7|A0A3P1VVM7_FUSNU DUF452 family protein	28	Fusobacterium nucleatum 13_3C
tr|A0A2C6BVR4|A0A2C6BVR4_FUSNP Uncharacterized protein	28	Fusobacterium nucleatum subsp. polymorphum
tr|A0A0S1YT90|A0A0S1YT90_FUSNP Uncharacterized protein	28	Fusobacterium nucleatum subsp. polymorphum
tr|Q7P338|Q7P338_FUSNV Hypothetical Cytosolic Protein	28	Fusobacterium nucleatum subsp. vincentii

tr|A0A241PZR4|A0A241PZR4_FUSNP Uncharacterized protein	28	Fusobacterium nucleatum subsp. polymorphum
tr|A5TWN4|A5TWN4_FUSNP Uncharacterized protein	28	Fusobacterium nucleatum subsp. polymorphum ATCC
tr|F7KY58|F7KY58_FUSNU Uncharacterized protein	28	Fusobacterium nucleatum subsp. animalis 11_3_2
tr|A0A2N6TKV6|A0A2N6TKV6_FUSNU DUF452 domain-containing protein	28	Fusobacterium nucleatum
tr|A0A1S1ME85|A0A1S1ME85_FUSNU Uncharacterized protein	28	Fusobacterium nucleatum
tr|A0A140PQK8|A0A140PQK8_FUSNU Uncharacterized protein	28	Fusobacterium nucleatum subsp. animalis 7_1
tr|A0A377GQL1|A0A377GQL1_FUSNV Uncharacterized protein conserved in bacteria	28	Fusobacterium nucleatum subsp. vincentii
tr|H1HF08|H1HF08_FUSNU Uncharacterized protein	46	Fusobacterium nucleatum subsp. animalis F0419
tr|U7TFI5|U7TFI5_FUSNU Uncharacterized protein	46	Fusobacterium nucleatum CTI-1
tr|U7STJ9|U7STJ9_FUSNU Uncharacterized protein	46	Fusobacterium nucleatum CTI-5

Table 4. 6 Proteins from different *F. nucleatum* strains detected by MALDI-TOF/TOF MS in a liquid culture containing a mixture of four *F. nucleatum* subspecies, namely; *animalis*, *polymorphum*, *vincentii*, and *nucleatum*. Mascot protein score = -10 Log (P), where P represents the probability of a protein hit to be a random event.

Additionally, as discussed in section 3, saliva samples acquired from a total of 4 colorectal adenomatous polyp patients and a total of 8 non-adenoma controls were analyzed using MALDI-TOF/TOF MS. Each saliva fraction was spotted in quadruplicate. The previously described protein search parameters were applied, with "*Fusobacterium nucleatum*" UniProt subset for protein identification in Mascot.

The patients (N = 4) were assigned numbers from 1 to 4, and the controls (N = 8) were assigned numbers from 1 to 8.

The *F. nucleatum* subspecies found in both groups were *F. nucleatum* subsp*. animalis* with the highest # protein IDs detected in both groups, represented as 24 hits in the patients, as opposed to 51 hits in the controls. The next most abundant subspecies were *F. nucleatum* subsp*. vincentii* (22 hits in patients , 24 hits in controls), *F. nucleatum* subsp*. polymorphum* (3 hits in patients , 20 hits in controls). However, the analysis of the patient samples yielded no *F. nucleatum* subsp*. nucleatum* hits, compared to 13 hits of this subspecies that were found in the controls. The numbers of proteins identified from each *F. nucleatum* subspecies are listed in **Table 4.7** and graphically visualized in **Figure 4.1**.

A)

# Patient	1	2	3	4
F. nucleatum* subsp. *animalis	13	0	10	1
F. nucleatum* subsp. *polymorphum	1	1	0	1
F. nucleatum* subsp. *vincentii	5	2	14	1
F. nucleatum* subsp. *nucleatum	0	0	0	0

B)

# Control	1	2	3	4	5	6	7	8
F. nucleatum* subsp. *animalis	13	17	4	4	0	2	4	7
F. nucleatum* subsp. *polymorphum	0	1	0	1	1	12	1	4
F. nucleatum* subsp. *vincentii	2	3	5	4	9	2	1	0
F. nucleatum* subsp. *nucleatum	0	3	0	0	3	3	0	4

Table 4. 7 A) The number of *F. nucleatum* subspecies protein IDs detected per each patient saliva sample (N=4).

B) The number of *F. nucleatum* subspecies protein IDs detected per each control saliva sample (N=8). Data were obtained using AutoFlex™ Speed Mascot search engine in BioTools 3.3 software

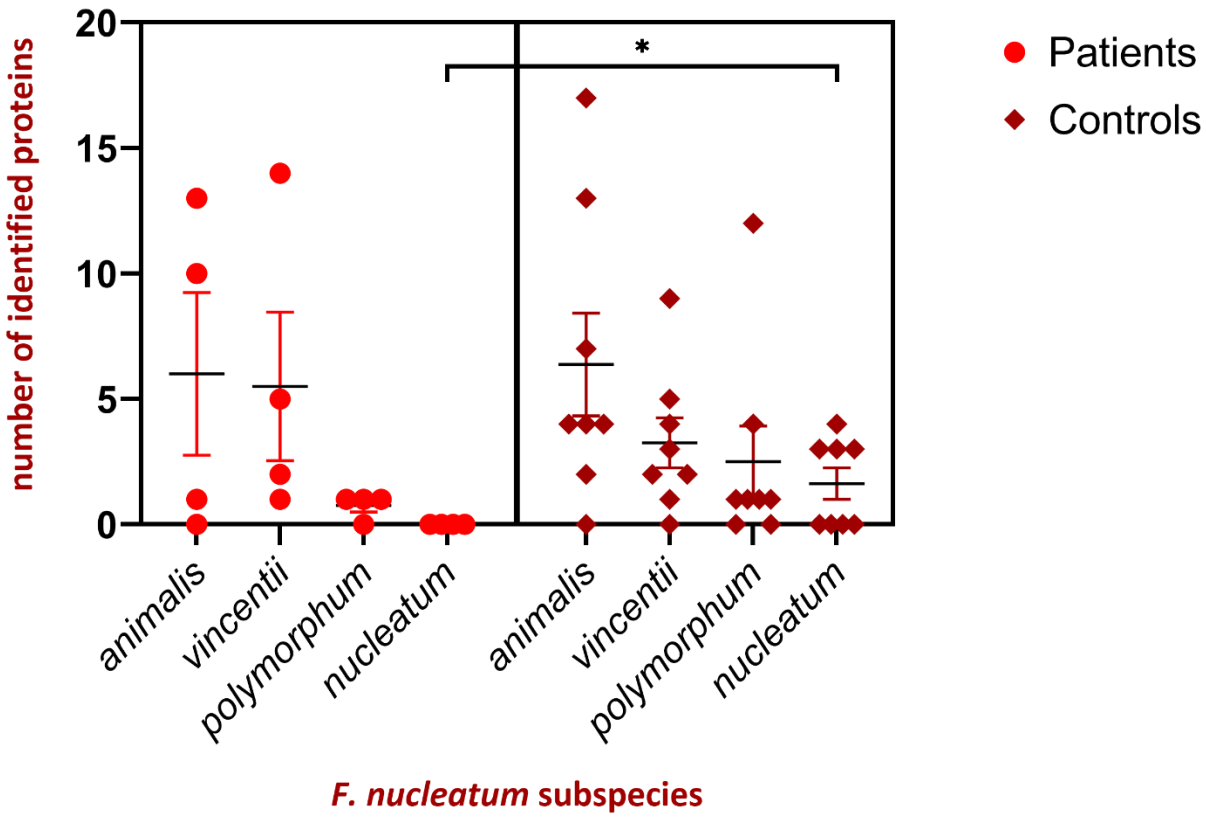

Figure 4. 1 Nested graphs showing the number of *F. nucleatum* subspecies protein IDs detected per each subject in both patients and controls groups. The experiment was done in technical quadruplicate. Values showed are means ± SEM. Significance was assessed using nested t-test * $P < 0.05$

The average of the Mascot score per subspecies per group was calculated, compared and depicted in **Figure 4.2**. In descending order, in the patients, *F. nucleatum* subsp. *polymorphum* had more confident protein hits with an average score of 57.3, followed by *F. nucleatum* subsp. *animalis* (53.3) and *vincentii* (50.3). On the other hand, in the control group,

the average scores were 48.4, 47.8, 46.9, and 43.8 for *F. nucleatum* subsp. *vincentii*, *polymorphum, nucleatum*, and *animalis*, respectively. No significant differences were found between the average Mascot scores for the subspecies, except for *F. nucleatum* subsp. *nucleatum*, which was not detected in the patient group ($P < 0.05$).

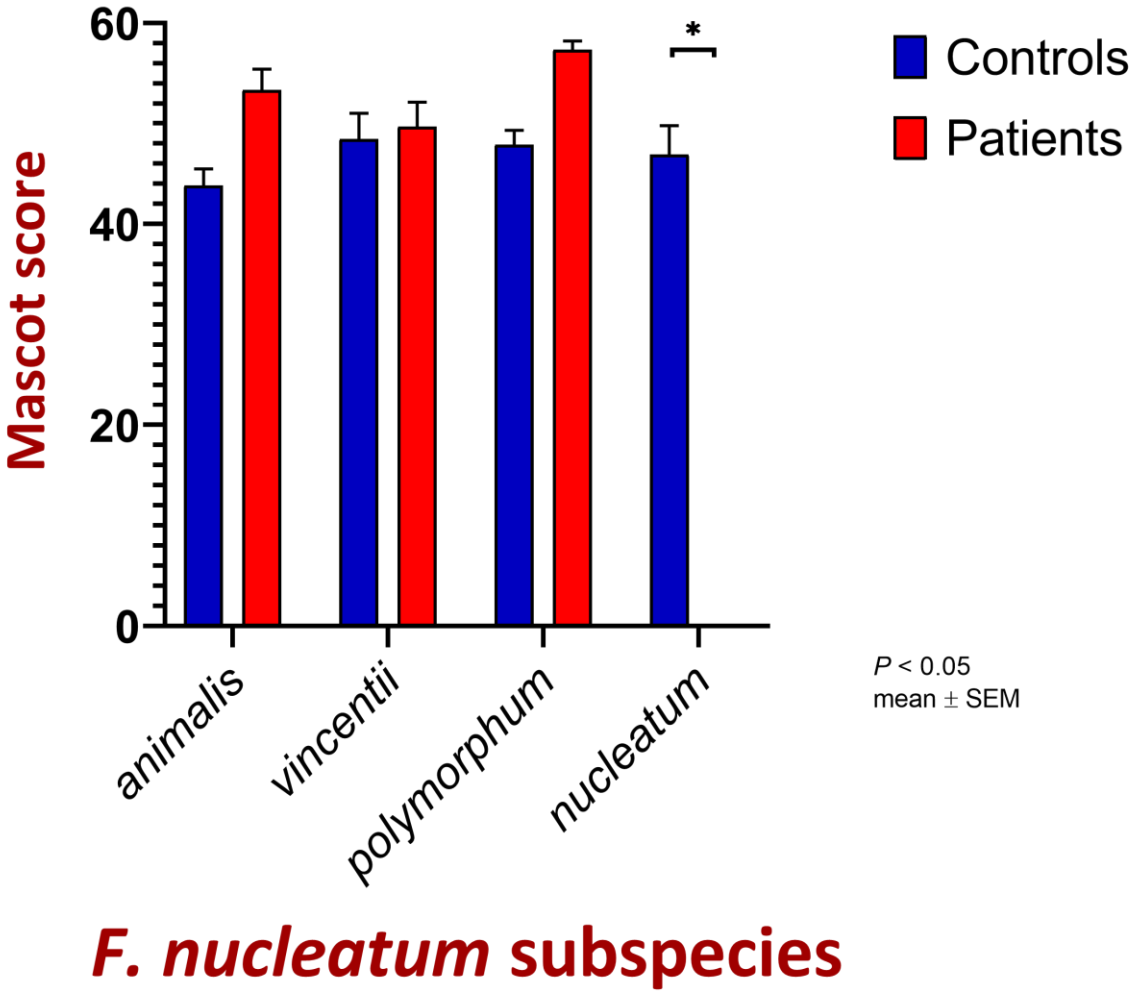

Figure 4. 2 Mean Mascot scores of *F. nucleatum* subspecies proteins detected in each group, patients, and controls. The experiment was done in technical quadruplicate. Values showed raw statistical means ± SEM. * P < 0.05.

A total of 62 proteins from four *F. nucleatum* subspecies were found to be unique to the patients (N = 4), while 133 proteins detected were unique to the controls (N = 8). One protein was detected in both groups, which was an uncharacterized protein from *F. nucleatum* subsp. *vincentii* with accession # A0A2G7H8S4_FUSNV. The results of the overlap analysis are shown in **Figure 4.3**.

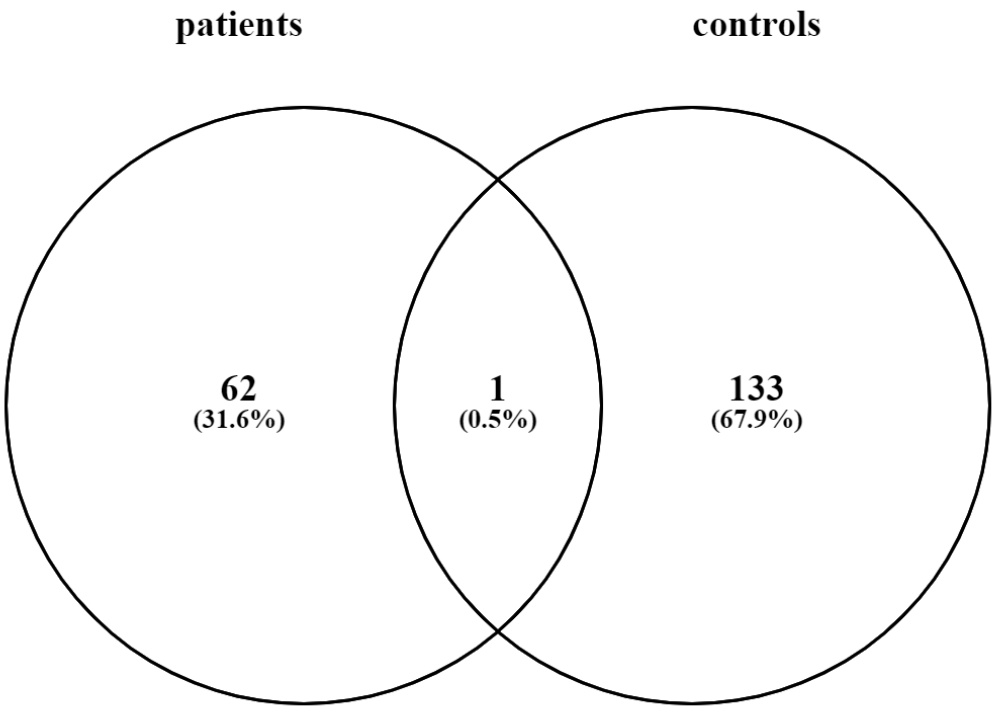

Figure 4. 3 Venn's diagram showing different *F. nucleatum* subspecies proteins detected in the patients and controls groups. A total of 62 unique proteins were detected that belong to *F. nucleatum subspecies* in patients group compared to 133 proteins unique to *F. nucleatum* subspecies in controls. Uncharacterized protein of *F. nucleatum* subsp. *vincentii* (accession A0A2G7H8S4_FUSNV) was the only protein in common between the two groups**.**

4. Discussion

The findings of this study were our ability to detect *F. nucleatum* at the subspecies level with the tandem time-of-flight MS, directly from clinical samples (saliva), without implementing a separate culturing and isolation step. Since employing cultivation-based techniques for microbial identification could add both time and costs burdens(152), using a cultivation-independent technique promotes a more robust and cost-effective bacterial biotyping method.

Standard MALDI-TOF MS was formerly applied for bacterial detection successfully in several investigations(153–156). Moreover, it was used in a previous study to detect the *F. nucleatum* subspecies from clinical isolates(157). Despite the rapidity of the technique, it still has the drawback of being highly dependent on the size and quality of commercially available databases that include signature microbial spectra not to mention the additional cultivation steps involved in their bacterial identification protocol (157,158).

For MALDI-TOF/TOF MS, each experiment was completed in three days, encompassing an overnight protein precipitation step and an overnight trypsin digestion step, but we were able to detect high numbers of proteins that were unique to each of the *F. nucleatum* subspecies albeit from a small number of samples. That goes in hand with previous studies reporting that MS/MS provides more comprehensive and sensitive detection of bacterial taxonomic traits compared to the standard MALDI-TOF MS biotyping(159).

Microwave-assisted digestion has also been reported in the proteomics literature in the trials to make use of the unique characteristics of MW irradiation, with some encouraging results that decreased the digestion time down to 6 minutes in some experiments(149,160–163). We have conducted optimization experiments to compare the two techniques, and have shown that digestion time could be limited to 60 seconds under optimal conditions as shown in section 4.1.3. Of noticeable importance, the time difference between the standard overnight digestion and our 60-second digestion, is promising to boost the throughput of the conducted sample preparation protocol. However, questionable reproducibility of the microwave-assisted digestion when used for multiple samples, would necessitate further development and optimization, and thus was not incorporated in the final method used to analyze the clinical samples.

To further optimize the protocol, we ran the pilot analyses on more than one instrument, namely Autoflex™ Speed (Bruker Daltonics) and 4800™ (AB Sciex). While both instruments were successful in identifying proteins from *F. nucleatum* subspecies in as little as 0.5 mL saliva, the Autoflex™ Speed's performance was superior to 4800™ in terms of throughput, sensitivity and the efficiency of protein identification. This could be owed to a variety of reasons, including the different search algorithms used with each instrument (Mascot for Autoflex™ Speed versus ProteinPilot™ for 4800™). Additionally, the inclusion of a novel technology in the AutoFlex™ system, e.g., LIFT™ fragmentation cell which acts as the product ion source for the MS/MS analysis and significantly enhances sensitivity, resolution and PMF success rate, we opted for Autoflex™ Speed in this study(164).

Before choosing to use MALDI-TOF MS analysis, we aimed at putting in use a quantitative polymerase chain reaction (qPCR)-based method. Motivated by the quantitative nature and the robustness of this technique, we planned to design strain-specific qPCR primers. However, the high genetic similarity between the subspecies makes it challenging to design these highly specific primers. For instance, it was reported that the difference between the 16s rRNA sequences from *F. nucleatum* subspecies were as low as 0.6%–1.9%(25). In such, we got a high rate of overmatches of nucleotide sequences with either *F. nucleatum* subspecies, or other Fusobacteria, when tested primer sequences with a BLAST algorithm against the NCBI database.

It is noteworthy that we also performed an experiment where we did not fractionate saliva proteins, but instead used the intact saliva sample, and the results were comparable to the fractionated sample (**Table A.3** and **Table A.4**). Nonetheless, we chose to adopt the widely accepted samples preparation method, i.e., protein fractionation to reduce sample complexity and increase the detection probability of less abundant proteins and/or minor *F. nucleatum* subspecies(165). A SPE clean up would also be beneficial to the MS instrument as it removes non-protein contaminants such as salts and carbohydrates that are abundant in raw saliva. However, further investigation in this area is recommended as eliminating the fractionation step could significantly improve the throughput of the analysis rendering it more amenable to clinical testing.

Our findings indicate no statistically significant difference between the number and average protein scores of the proteins identified from *F. nucleatum* subspecies in the patients versus those found in the controls, which is in agreement with the findings of another study that reported an absence of a significant difference between the counts of *F. Nucleatum*

subspecies in CRC patients and non-CRC controls(37). Interestingly, however, we were not able to detect any *F. nucleatum* subsp. *nucleatum* proteins in the saliva samples procured from the four adenoma patients recruited for this study. This would raise questions whether this subtype could lose its ecological habitat in the oral cavity in the case of a colon disease, or whether it is an indication of a microbial dysbiosis in the oral flora that mirrors a gut dysbiosis. Although unlikely, the lack of *F. nucleatum* subsp*. nucleatum* proteins could also be explained by the limited number of the saliva samples from adenoma patients included in the analysis.

Out of 195 proteins that were identified in the saliva of the study participants, it was interesting to find out that only a single protein occurred in common between the patients and the controls. This may be referred to the "shotgun" nature of bottom-up proteomics, which is the identification of proteins based on their resulting fragments or peptides from the proteolysis process(166). Shotgun proteomics was named this way by the Yates lab, owing to its similarity to the shotgun genomic sequencing(167). This makes it an arguably poorly reproducible technique for protein profiling applications(168), such as the one used in this study, it could, nonetheless, pave the way to screen for novel protein biomarkers as a diagnostic or curative approach for patients with adenomatous polyps or CRC.

An important setback of this study is the relatively small number of samples included in the MALDI-MS analyses. This is essentially attributable to insufficient number of patients referring to the three gastroenterologists involved in this research project, with their conditions consistent with patient selection criteria reviewed before. However, to further validate our data and explore any possible correlation between the presence and distribution of *F. nucleatum*

subspecies in the saliva of adenoma or CRC patients, it would be beneficial to account for a larger number of cases with a longitudinal follow up analyses.

The nature of bottom-up shotgun proteomics analyses we used, which is based on proteolytic digestion of the proteins and MS/MS analysis of the digest peptides, is yet another limitation of the present study. In shotgun proteomics, analytical reproducibility is a well-known challenge owing to the considerably fragmented nature of the data(131). We tried to reduce this effect by analyzing saliva samples under standardized conditions and in technical quadruplicate. Besides, the absence of a quantitative nature makes MALDI-MS data less reliable as a precise abundance measure. Although the number of PSMs obtained from each subject could be used to surrogate the abundance of corresponding bacteria in their saliva, we should not overlook the possibility that the higher number of PSMs for some bacteria could have been caused by their proteins being better retained during the samples preparation procedure rather than an actual abundance.

Optimizing the same proteomics approach to use MALDI-TOF/TOF MS to detect the *F. nucleatum* subspecies in colon tissues is currently under our investigation. This would be a future direction for this study.

Additionally, we would advocate, under a more flexible timeline and budget conditions, the in-parallel utilization of 16S rRNA sequencing technique to further validate the in-hand results generated by MALDI-TOF/TOF MS.

5. Conclusion

The possible role of *F. nucleatum* subspecies in the occurrence of CRC remains imprecisely determined. Our study presents a novel technique for the detection of *F. nucleatum* subspecies, from saliva specimens that could be later employed for a better understanding of a possible individual role of those subspecies in the CRC development. Subsequently, this will further aid in a possible targeting of a specific subspecies as a modality of screening for, preventing or curing the disease in the future.

Here we have shown that tandem time-of-flight MALDI-MS is a powerful technique to identify the *F. nucleatum* at the subspecies level. Our data advise that some differences in the subspecies existing in the saliva might occur among the colorectal adenoma patients versus the non-adenoma subjects. Understanding this contrast could assist in elucidating the disease pathogenesis further. This opens up certain avenues so that this method could be used under a variety of sampling and analysis protocols, which could improve the output of studies conducted on a larger scale. In addition, this technique may facilitate the identification of other microbial pathogens in saliva, putting in use suitable search parameters.

References

1. Zhou X, Li Y, editors. Chapter 4 - Subgingival Microbes. In: Atlas of Oral Microbiology [Internet]. Oxford: Academic Press; 2015 [cited 2019 Mar 13]. p. 67–93. Available from: http://www.sciencedirect.com/science/article/pii/B9780128022344000045

2. Brook I. Fusobacterial infections in children. J Infect. 1994 Mar 1;28(2):155–65.

3. Bolstad AI, Jensen HB, Bakken V. Taxonomy, biology, and periodontal aspects of Fusobacterium nucleatum. Clin Microbiol Rev. 1996 Jan;9(1):55–71.

4. Boone DR, Castenholz RW, Garrity GM, editors. Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer; 2001. 1 p.

5. Moore WE, Moore LV. The bacteria of periodontal diseases. Periodontol 2000. 1994 Jun;5:66–77.

6. Shang FM, Liu HL. Fusobacterium nucleatum and colorectal cancer: A review. World J Gastrointest Oncol. 2018 Mar 15;10(3):71–81.

7. Bennett KW, Eley A. Fusobacteria: new taxonomy and related diseases. J Med Microbiol. 1993 Oct;39(4):246–54.

8. Jousimies-Somer H, Savolainen S, Mäkitie A, Ylikoski J. Bacteriologic findings in peritonsillar abscesses in young adults. Clin Infect Dis Off Publ Infect Dis Soc Am. 1993 Jun;16 Suppl 4:S292-298.

9. Crippin JS, Wang KK. An unrecognized etiology for pyogenic hepatic abscesses in normal hosts: dental disease. Am J Gastroenterol. 1992 Dec;87(12):1740–3.

10. Scoular A, Corcoran GD, Malin A, Evans BA, Davies A, Miller RF. Fusobacterium nucleatum bacteraemia with multiple liver abscesses in an HIV-I antibody positive man with IgG2 deficiency. J Infect. 1992 May;24(3):321–5.

11. Ribot S, Gal K, Goldblat MV, Eslami HH. The role of anaerobic bacteria in the pathogenesis of urinary tract infections. J Urol. 1981 Dec;126(6):852–3.

12. Chaim W, Mazor M. Intraamniotic infection with fusobacteria. Arch Gynecol Obstet. 1992;251(1):1–7.

13. Hillier SL, Krohn MA, Rabe LK, Klebanoff SJ, Eschenbach DA. The normal vaginal flora, H2O2-producing lactobacilli, and bacterial vaginosis in pregnant women. Clin Infect Dis Off Publ Infect Dis Soc Am. 1993 Jun;16 Suppl 4:S273-281.

14. Infective endocarditis due to Fusobacterium nucleatum: case report and review of the literature. - PubMed - NCBI [Internet]. [cited 2019 Mar 18]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/8416766

15. Truant AL, Menge S, Milliorn K, Lairscey R, Kelly MT. Fusobacterium nucleatum pericarditis. J Clin Microbiol. 1983 Feb;17(2):349–51.

16. Marina M, Strong CA, Civen R, Molitoris E, Finegold SM. Bacteriology of anaerobic pleuropulmonary infections: preliminary report. Clin Infect Dis Off Publ Infect Dis Soc Am. 1993 Jun;16 Suppl 4:S256-262.

17. Bartlett JG. Anaerobic bacterial infections of the lung and pleural space. Clin Infect Dis Off Publ Infect Dis Soc Am. 1993 Jun;16 Suppl 4:S248-255.

18. González-Gay MA, Sánchez-Andrade A, Cereijo MJ, Pulpeiro JR, Armesto V. Pyomyositis and septic arthritis from Fusobacterium nucleatum in a nonimmunocompromised adult. J Rheumatol. 1993 Mar;20(3):518–20.

19. Swidsinski A, Dorffel Y, Loening-Baucke V, Theissig F, Ruckert JC, Ismail M, et al. Acute appendicitis is characterised by local invasion with Fusobacterium nucleatum/necrophorum. Gut. 2011 Jan;60(1):34–40.

20. Han YW. Fusobacterium nucleatum: a commensal-turned pathogen. Curr Opin Microbiol. 2015 Feb;23:141–7.

21. Kei Mitsuhashi KN, Miki Ito HK, Naito YA, Toshiya Shinohara TH, Kimura KH, Yamamoto YS. <Association of Fusobacterium species in pancreatic cancer tissues with molecular features and prognosis..pdf>. Oncotarget. 2015;6(9).

22. Han YW, Redline RW, Li M, Yin L, Hill GB, McCormick TS. Fusobacterium nucleatum Induces Premature and Term Stillbirths in Pregnant Mice: Implication of Oral Bacteria in Preterm Birth. Infect Immun. 2004;72(4):2272–9.

23. Falkler WA, Montgomery J, Nauman RK, Alpers M. Isolation of Fusobacterium nucleatum and electron microscopic observations of spirochetes from tropical skin ulcers in Papua New Guinea. Am J Trop Med Hyg. 1989 Apr;40(4):390–8.

24. Dzink JL, Sheenan MT, Socransky SS. Proposal of three subspecies of Fusobacterium nucleatum Knorr 1922: Fusobacterium nucleatum subsp. nucleatum subsp. nov., comb. nov.; Fusobacterium nucleatum subsp. polymorphum subsp. nov., nom. rev., comb. nov.; and Fusobacterium nucleatum subsp. vincentii subsp. nov., nom. rev., comb. nov. Int J Syst Bacteriol. 1990 Jan;40(1):74–8.

25. Kim HS, Lee DS, Chang YH, Kim MJ, Koh S, Kim J, et al. Application of rpoB and zinc protease gene for use in molecular discrimination of Fusobacterium nucleatum subspecies. J Clin Microbiol. 2010 Feb;48(2):545–53.

26. Kook JK, Park SN, Lim YK, Choi MH, Cho E, Kong SW, et al. Fusobacterium nucleatum subsp. fusiforme Gharbia and Shah 1992 is a later synonym of Fusobacterium nucleatum subsp. vincentii Dzink et al. 1990. Curr Microbiol. 2013 Apr;66(4):414–7.

27. Han YW, Wang X. Mobile microbiome: oral bacteria in extra-oral infections and inflammation. J Dent Res. 2013 Jun;92(6):485–91.

28. Genco RJ, Loos BG. The use of genomic DNA fingerprinting in studies of the epidemiology of bacteria in periodontitis. J Clin Periodontol. 1991 Jul;18(6):396–405.

29. Gharbia SE, Shah HN, Lawson PA, Haapasalo M. Distribution and frequency of Fusobacterium nucleatum subspecies in the human oral cavity. Oral Microbiol Immunol. 1990 Dec;5(6):324–7.

30. Ye X, Wang R, Bhattacharya R, Boulbes DR, Fan F, Xia L, et al. *Fusobacterium Nucleatum* Subspecies *Animalis* Influences Proinflammatory Cytokine Expression and Monocyte Activation in Human Colorectal Tumors. Cancer Prev Res (Phila Pa). 2017 Jul;10(7):398–409.

31. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol. 2014 Nov 28;10:766.

32. Wade WG. The oral microbiome in health and disease. Pharmacol Res. 2013 Mar;69(1):137–43.

33. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012 Jun 13;486(7402):207–14.

34. Kato I, Vasquez AA, Moyerbrailean G, Land S, Sun J, Lin H-S, et al. Oral microbiome and history of smoking and colorectal cancer. J Epidemiol Res. 2016 Oct;2(2):92–101.

35. Warinner C, Rodrigues JFM, Vyas R, Trachsel C, Shved N, Grossmann J, et al. Pathogens and host immunity in the ancient human oral cavity. Nat Genet. 2014 Apr;46(4):336–44.

36. Zackular JP, Rogers MAM, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prev Res Phila Pa. 2014 Nov;7(11):1112–21.

37. Russo E, Bacci G, Chiellini C, Fagorzi C, Niccolai E, Taddei A, et al. Preliminary Comparison of Oral and Intestinal Human Microbiota in Patients with Colorectal Cancer: A Pilot Study. Front Microbiol. 2017;8:2699.

38. Amitay EL, Werner S, Vital M, Pieper DH, Hofler D, Gierse IJ, et al. Fusobacterium and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. Carcinogenesis. 2017 Aug 1;38(8):781–8.

39. Ye X, Wang R, Bhattacharya R, Boulbes DR, Fan F, Xia L, et al. Fusobacterium Nucleatum Subspecies Animalis Influences Proinflammatory Cytokine Expression and Monocyte Activation in Human Colorectal Tumors. Cancer Prev Res Phila. 2017 Jul;10(7):398–409.

40. Yasuhiko Komiya YS. Patients with colorectal cancer have identical strains of Fusobacterium nucleatum in their colorectal cancer and oral cavity. Gut. 2018;

41. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990 Jun 1;61(5):759–67.

42. Hill MJ, Morson BC, Bussey HJ. Aetiology of adenoma--carcinoma sequence in large bowel. Lancet Lond Engl. 1978 Feb 4;1(8058):245–7.

43. Peipins LA, Sandler RS. Epidemiology of colorectal adenomas. Epidemiol Rev. 1994;16(2):273–97.

44. Jass JR. Do all colorectal carcinomas arise in preexisting adenomas? World J Surg. 1989 Feb;13(1):45–51.

45. colorectal-cancer-facts-and-figures-2014-2016.

46. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. CA Cancer J Clin. 2016;66(4):271–89.

47. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015 Mar;65(2):87–108.

48. Canadian Cancer Statistics publication - Canadian Cancer Society [Internet]. www.cancer.ca. [cited 2019 Mar 17]. Available from: https://www.cancer.ca:443/en/cancer-information/cancer-101/canadian-cancer-statistics-publication/?region=on

49. Edwards BK, Ward E, Kohler BA, Eheman C, Zauber AG, Anderson RN, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. Cancer. 2010 Feb 1;116(3):544–73.

50. Danaei G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJ, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. PLoS Med. 2009 Apr 28;6(4):e1000058.

51. Jobin C. Colorectal cancer: looking for answers in the microbiota. Cancer Discov. 2013 Apr;3(4):384–7.

52. Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, et al. Microbial dysbiosis in colorectal cancer (CRC) patients. PLoS One. 2011 Jan 27;6(1):e16393.

53. Mira-Pascual L, Cabrera-Rubio R, Ocon S, Costales P, Parra A, Suarez A, et al. Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. J Gastroenterol. 2015 Feb;50(2):167–79.

54. Holmes E, Li JV, Marchesi JR, Nicholson JK. Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk. Cell Metab. 2012 Nov 7;16(5):559–64.

55. Xu J, Gordon JI. Honor thy symbionts. Proc Natl Acad Sci. 2003 Sep 2;100(18):10452–9.

56. Arthur JC, Jobin C. The complex interplay between inflammation, the microbiota and colorectal cancer. Gut Microbes. 2013 Jun;4(3):253–8.

57. Candela M, Turroni S, Biagi E, Carbonero F, Rampelli S, Fiorentini C, et al. Inflammation and colorectal cancer, when microbiota-host mutualism breaks. World J Gastroenterol. 2014 Jan 28;20(4):908–22.

58. McCoy AN, Araujo-Perez F, Azcarate-Peril A, Yeh JJ, Sandler RS, Keku TO. Fusobacterium is associated with colorectal adenomas. PLoS One. 2013;8(1):e53653.

59. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe. 2013 Aug 14;14(2):207–15.

60. Ito M, Kanno S, Nosho K, Sukawa Y, Mitsuhashi K, Kurihara H, et al. Association of Fusobacterium nucleatum with clinical and molecular features in colorectal serrated pathway. Int J Cancer. 2015 Sep 15;137(6):1258–68.

61. Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan TJ, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science. 2012 Oct 5;338(6103):120–3.

62. Viljoen KS, Dakshinamurthy A, Goldberg P, Blackburn JM. Quantitative profiling of colorectal cancer-associated bacteria reveals associations between fusobacterium spp., enterotoxigenic Bacteroides fragilis (ETBF) and clinicopathological features of colorectal cancer. PLoS One. 2015;10(3):e0119462.

63. Tsai CE, Chiu CT, Rayner CK, Wu KL, Chiu YC, Hu ML, et al. Associated factors in Streptococcus bovis bacteremia and colorectal cancer. Kaohsiung J Med Sci. 2016 Apr;32(4):196–200.

64. Flanagan L, Schmid J, Ebert M, Soucek P, Kunicka T, Liska V, et al. Fusobacterium nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. Eur J Clin Microbiol Infect Dis. 2014 Aug;33(8):1381–90.

65. Gagniere J, Raisch J, Veziant J, Barnich N, Bonnet R, Buc E, et al. Gut microbiota imbalance and colorectal cancer. World J Gastroenterol. 2016 Jan 14;22(2):501–18.

66. Binder Gallimidi A, Fischman S, Revach B, Bulvik R, Maliutina A, Rubinstein AM, et al. Periodontal pathogens Porphyromonas gingivalis and Fusobacterium nucleatum promote tumor progression in an oral-specific chemical carcinogenesis model. Oncotarget. 2015 Sep 8;6(26):22613–23.

67. Mitsuhashi K NK. Association of Fusobacterium species in pancreatic cancer tissues with molecular features and prognosis. 2015;

68. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut. 2017 Jan;66(1):70–8.

69. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, et al. Co-occurrence of anaerobic bacteria in colorectal carcinomas. Microbiome. 2013;1(1):16.

70. Karpathy SE, Qin X, Gioia J, Jiang H, Liu Y, Petrosino JF, et al. Genome Sequence of Fusobacterium nucleatum Subspecies Polymorphum — a Genetically Tractable Fusobacterium. PLOS ONE. 2007 Aug 1;2(8):e659.

71. Mima K, Nishihara R, Qian ZR, Cao Y, Sukawa Y, Nowak JA, et al. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. Gut. 2016 Dec;65(12):1973–80.

72. Tahara T, Yamamoto E, Suzuki H, Maruyama R, Chung W, Garriga J, et al. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. Cancer Res. 2014 Mar 1;74(5):1311–8.

73. Allen-Vercoe E, Strauss J, Chadee K. Fusobacterium nucleatum: an emerging gut pathogen? Gut Microbes. 2011 Sep 1;2(5):294–8.

74. 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 Aug 14;14(2):195–206.

75. Fardini Y WX, M HY. Fusobacterium nucleatum adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity. Mol Microbiol. 2011;82:1468–80.

76. Han YW, Ikegami A, Rajanna C, Kawsar HI, Zhou Y, Li M, et al. Identification and characterization of a novel adhesin unique to oral fusobacteria. J Bacteriol. 2005 Aug;187(15):5330–40.

77. Quah SY, Bergenholtz G, Tan KS. Fusobacterium nucleatum induces cytokine production through Toll-like-receptor-independent mechanism. Int Endod J. 2014 Jun;47(6):550–9.

78. Dharmani P, Strauss J, Ambrose C, Allen-Vercoe E, Chadee K. Fusobacterium nucleatum infection of colonic cells stimulates MUC2 mucin and tumor necrosis factor alpha. Infect Immun. 2011 Jul;79(7):2597–607.

79. Bryant DM, Stow JL. The ins and outs of E-cadherin trafficking. Trends Cell Biol. 2004 Aug;14(8):427–34.

80. Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, et al. Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. Immunity. 2015 Feb 17;42(2):344–55.

81. Kaplan CW, Ma X, Paranjpe A, Jewett A, Lux R, Kinder-Haake S, et al. Fusobacterium nucleatum outer membrane proteins Fap2 and RadD induce cell death in human lymphocytes. Infect Immun. 2010 Nov;78(11):4773–8.

82. Bashir A, Miskeen AY, Hazari YM, Asrafuzzaman S, Fazili KM. Fusobacterium nucleatum, inflammation, and immunity: the fire within human gut. Tumour Biol. 2016 Mar;37(3):2805–10.

83. Abed J, Emgard JE, Zamir G, Faroja M, Almogy G, Grenov A, et al. Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. Cell Host Microbe. 2016 Aug 10;20(2):215–25.

84. Shang F-M, Liu H-L. Fusobacterium nucleatum and colorectal cancer: A review. World J Gastrointest Oncol. 2018 Mar 15;10(3):71–81.

85. Strauss J, White A, Ambrose C, McDonald J, Allen-Vercoe E. Phenotypic and genotypic analyses of clinical Fusobacterium nucleatum and Fusobacterium periodonticum isolates from the human gut. Anaerobe. 2008 Dec 1;14(6):301–9.

86. Adzitey F, Huda N, Ali GRR. Molecular techniques for detecting and typing of bacteria, advantages and application to foodborne pathogens isolated from ducks. 3 Biotech. 2013 Apr;3(2):97–107.

87. Gonzales-Marin C, Spratt DA, Allaker RP. Maternal oral origin of Fusobacterium nucleatum in adverse pregnancy outcomes as determined using the 16S-23S rRNA gene intergenic transcribed spacer region. J Med Microbiol. 2013 Jan;62(Pt 1):133–44.

88. Wassenaar TM, Newell DG. Genotyping of Campylobacter spp. Appl Environ Microbiol. 2000 Jan;66(1):1–9.

89. Urwyler SK, Glaubitz J. Advantage of MALDI-TOF-MS over biochemical-based phenotyping for microbial identification illustrated on industrial applications. Lett Appl Microbiol. 2016 Feb;62(2):130–7.

90. Croxatto A, Prod'hom G, Greub G. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol Rev. 2012 Mar;36(2):380–407.

91. Mann M, Hendrickson RC, Pandey A. Analysis of Proteins and Proteomes by Mass Spectrometry. Annu Rev Biochem. 2001;70(1):437–73.

92. Gross JH. Mass Spectrometry: A Textbook. Springer Science & Business Media; 2006. 533 p.

93. Gross JH. Introduction. In: Gross JH, editor. Mass Spectrometry: A Textbook. Cham: Springer International Publishing; 2017. p. 1–28.

94. McLafferty FW, Senko MW. Mass spectrometry in the development of drugs from traditional medicines. STEM CELLS. 1994;12(1):68–73.

95. Märk TD, Dunn GH. Electron Impact Ionization. Springer Science & Business Media; 2013. 395 p.

96. Mitchum RK, Moler GF, Korfmacher WA. Combined capillary gas chromatography/atmospheric pressure negative chemical ionization/mass spectrometry for the determination of 2,3,7,8-tetrachlorodibenzo-p-dioxin in tissue. Anal Chem. 1980 Dec 1;52(14):2278–82.

97. Whitehouse CM, Dreyer RN, Yamashita Masamichi, Fenn JB. Electrospray interface for liquid chromatographs and mass spectrometers. Anal Chem. 1985 Mar 1;57(3):675–9.

98. Bruins AP. Mass spectrometry with ion sources operating at atmospheric pressure. Mass Spectrom Rev. 1991;10(1):53–77.

99. Huang EC, Wachs T, Conboy JJ, Henion JD. Atmospheric pressure ionization mass spectrometry. Detection for the separation sciences. Anal Chem. 1990 Jul 1;62(13):713A-725A.

100. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom. 1988;2(8):151–3.

101. Chong Y-K, Ho C-C, Leung S-Y, Lau SKP, Woo PCY. Clinical Mass Spectrometry in the Bioinformatics Era: A Hitchhiker's Guide. Comput Struct Biotechnol J. 2018 Jan 1;16:316–34.

102. Karas M, Bachmann D, Bahr U, Hillenkamp F. Matrix-assisted ultraviolet laser desorption of non-volatile compounds. Int J Mass Spectrom Ion Process. 1987 Sep 24;78:53–68.

103. Bae YJ, Kim MS. A Thermal Mechanism of Ion Formation in MALDI. Annu Rev Anal Chem. 2015 Jul 10;8(1):41–60.

104. Lee C, Lu I-C, Hsu HC, Lin H-Y, Liang S-P, Lee Y-T, et al. Formation of Metal-Related Ions in Matrix-Assisted Laser Desorption Ionization. J Am Soc Mass Spectrom. 2016 Sep 1;27(9):1491–8.

105. Knochenmuss R. Ion Yields in the Coupled Chemical and Physical Dynamics Model of Matrix-Assisted Laser Desorption/Ionization. J Am Soc Mass Spectrom. 2015 Oct 1;26(10):1645–8.

106. Kirmess KM, Knochenmuss R, Blanchard GJ, Kinsel GR. MALDI ionization mechanisms investigated by comparison of isomers of dihydroxybenzoic acid. J Mass Spectrom. 2016;51(1):79–85.

107. Jiwan J-LH, Wallemacq P, Hérent M-F. HPLC-high resolution mass spectrometry in clinical laboratory? Clin Biochem. 2011 Jan 1;44(1):136–47.

108. Chace DH. Mass Spectrometry in the Clinical Laboratory. Chem Rev. 2001 Feb 1;101(2):445–78.

109. Wolff MM, Stephens WE. A Pulsed Mass Spectrometer with Time Dispersion. Rev Sci Instrum. 1953 Aug 1;24(8):616–7.

110. Niehuis E, Heller T, Feld H, Benninghoven A. Design and performance of a reflectron based time-of-flight secondary ion mass spectrometer with electrodynamic primary ion mass separation. J Vac Sci Technol A. 1987 Jul 1;5(4):1243–6.

111. El-Aneed A, Cohen A, Banoub J. Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. Appl Spectrosc Rev. 2009 Apr 1;44(3):210–30.

112. Spinali S, van Belkum A, Goering RV, Girard V, Welker M, Van Nuenen M, et al. Microbial Typing by Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry: Do We Need Guidance for Data Interpretation? Doern GV, editor. J Clin Microbiol. 2015 Mar;53(3):760–5.

113. Bizzini A, Greub G. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry, a revolution in clinical microbial identification. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2010 Nov;16(11):1614–9.

114. Lasch P, Nattermann H, Erhard M, Stämmler M, Grunow R, Bannert N, et al. MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. Anal Chem. 2008 Mar 15;80(6):2026–34.

115. Mesureur J, Ranaldi S, Monnin V, Girard V, Arend S, Welker M, et al. A Simple and Safe Protocol for Preparing Brucella Samples for Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry Analysis. J Clin Microbiol. 2016 Feb 1;54(2):449–52.

116. Cherkaoui A, Hibbs J, Emonet S, Tangomo M, Girard M, Francois P, et al. Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. J Clin Microbiol. 2010 Apr;48(4):1169–75.

117. Tam EWT, Chen JHK, Lau ECL, Ngan AHY, Fung KSC, Lee K-C, et al. Misidentification of Aspergillus nomius and Aspergillus tamarii as Aspergillus flavus: characterization by internal transcribed spacer, β-Tubulin, and calmodulin gene sequencing, metabolic fingerprinting, and matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2014 Apr;52(4):1153–60.

118. Tsang C-C, Xiong L, Poon RWS, Chen JHK, Leung K-W, Lam JYW, et al. Gordonia hongkongensis sp. nov., isolated from blood culture and peritoneal dialysis effluent of patients in Hong Kong. Int J Syst Evol Microbiol. 2016 Oct;66(10):3942–50.

119. Tse CW, Curreem SO, Cheung I, Tang BS, Leung K-W, Lau SK, et al. A novel MLST sequence type discovered in the first fatal case of Laribacter hongkongensis bacteremia clusters with the sequence types of other human isolates. Emerg Microbes Infect. 2014 Jun;3(6):e41.

120. Teng JLL, Ho TCC, Yeung RSY, Wong AYP, Wang H, Chen C, et al. Evaluation of 16SpathDB 2.0, an automated 16S rRNA gene sequence database, using 689 complete bacterial genomes. Diagn Microbiol Infect Dis. 2014 Feb;78(2):105–15.

121. Woo PCY, Teng JLL, Yeung JMY, Tse H, Lau SKP, Yuen K-Y. Automated identification of medically important bacteria by 16S rRNA gene sequencing using a novel comprehensive database, 16SpathDB. J Clin Microbiol. 2011 May;49(5):1799–809.

122. Min Fang. Tandem Mass Spectrometry - Based Proteomics, Protein Characterisation and Biomarker Discovery in Microorganisms. In: Haroun N. Shah and Saheer E. Gharbia, editor. Mass Spectrometry for Microbial Proteomics. John Wiley & Sons, Ltd; 2010.

123. Tandem mass spectrometry: dissociation of ions by collisional activation - Shukla - 2000 - Journal of Mass Spectrometry - Wiley Online Library [Internet]. [cited 2019 Jun 4]. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/1096-9888%28200009%2935%3A9%3C1069%3A%3AAID-JMS54%3E3.0.CO%3B2-C

124. Kurnosenko S, Moskovets E. On the high-resolution mass analysis of the product ions in tandem time-of-flight (TOF/TOF) mass spectrometers using a time-dependent re-acceleration technique. Rapid Commun Mass Spectrom RCM. 2010 Jan;24(1):63–74.

125. Spengler B, Kirsch D, Kaufmann R, Cotter RJ. Metastable decay of peptides and proteins in matrix-assisted laser-desorption mass spectrometry. Rapid Commun Mass Spectrom. 1991;5(4):198–202.

126. Mass Spectrometry/Mass Spectrometry. In: Introduction to Mass Spectrometry [Internet]. John Wiley & Sons, Ltd; 2008 [cited 2019 Jun 8]. p. 173–228. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470516898.ch3

127. Vestal ML, Campbell JM. Tandem Time-of-Flight Mass Spectrometry. In: Methods in Enzymology [Internet]. Academic Press; 2005 [cited 2019 Jun 8]. p. 79–108. (Biological Mass Spectrometry; vol. 402). Available from: http://www.sciencedirect.com/science/article/pii/S0076687905020033

128. R YIJ, K EJ. Identification Of Nucleotides, Amino Acids, Or Carbohydrates By Mass Spectrometry [Internet]. US; US 6017693 A, 1996. Available from: https://lens.org/006-506-535-733-566

129. Eng JK, McCormack AL, Yates JR. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom. 1994 Nov;5(11):976–89.

130. Chen W, Laidig KE, Park Y, Park K, Yates JR, Lamont RJ, et al. Searching the Porphyromonas gingivalis genome with peptide fragmentation mass spectra. The Analyst. 2001 Jan;126(1):52–7.

131. Dworzanski JP, Snyder AP, Chen R, Zhang H, Wishart D, Li L. Identification of Bacteria Using Tandem Mass Spectrometry Combined with a Proteome Database and Statistical Scoring. Anal Chem. 2004 Apr 1;76(8):2355–66.

132. Protein Precipitation Method for Salivary Proteins and Rehydration Buffer for Two-Dimensional Electrophoresis [Internet]. Science Alert. [cited 2019 Jun 8]. Available from: https://scialert.net/fulltext/?doi=biotech.2008.686.693

133. Bladergroen MR, van der Burgt YEM. Solid-phase extraction strategies to surmount body fluid sample complexity in high-throughput mass spectrometry-based proteomics. J Anal Methods Chem. 2015;2015:250131.

134. Callesen AK, Madsen JS, Vach W, Kruse TA, Mogensen O, Jensen ON. Serum protein profiling by solid phase extraction and mass spectrometry: A future diagnostics tool? PROTEOMICS. 2009;9(6):1428–41.

135. Jandera P, Henze G. Liquid Chromatography, 1. Fundamentals, History, Instrumentation, Materials. In: Ullmann's Encyclopedia of Industrial Chemistry [Internet]. American Cancer Society; 2011 [cited 2019 Jun 8]. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/14356007.b05_237.pub2

136. Golizeh M. Development of LC-MS/MS methods for the analysis of reactive metabolite protein targets. :290.

137. Encheva V, Parker R. Conventional Approaches for Sample Preparation for Liquid Chromatography and Two-Dimensional Gel Electrophoresis. In: Mass Spectrometry for Microbial Proteomics [Internet]. John Wiley & Sons, Ltd; 2010 [cited 2019 Jun 9]. p. 117–42. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470665497.ch6

138. Thiede B, Höhenwarter W, Krah A, Mattow J, Schmid M, Schmidt F, et al. Peptide mass fingerprinting. Methods. 2005 Mar;35(3):237–47.

139. Brown WE, Wold F. Alkyl isocyanates as active-site-specific reagents for serine protease. Identification of the active-site serine as the site of reaction. Biochemistry. 1973 Feb 27;12(5):835–40.

140. Brown WE, Wold F. Alkyl isocyanates as active-site-specific reagents for serine proteases. Reaction properties. Biochemistry. 1973 Feb 27;12(5):828–34.

141. Rice RH, Means GE, Brown WD. Stabilization of bovine trypsin by reductive methylation. Biochim Biophys Acta. 1977 Jun 24;492(2):316–21.

142. Rapid identification of proteins by peptide-mass fingerprinting. - PubMed - NCBI [Internet]. [cited 2019 Jun 2]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15335725

143. Misra R. Overview of Proteomic Tools and Their Links to Genomics. In: Mass Spectrometry for Microbial Proteomics [Internet]. John Wiley & Sons, Ltd; 2010 [cited 2019 Jun 9]. p. 55–71. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470665497.ch3

144. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol. 1999 Jul;17(7):676–82.

145. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001 Mar;19(3):242–7.

146. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR. Protein sequencing by tandem mass spectrometry. Proc Natl Acad Sci. 1986 Sep 1;83(17):6233–7.

147. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis. 1999 Dec;20(18):3551–67.

148. Pramanik BN, Mirza UA, Ing YH, Liu Y-H, Bartner PL, Weber PC, et al. Microwave-enhanced enzyme reaction for protein mapping by mass spectrometry: A new approach to protein digestion in minutes. Protein Sci. 2002;11(11):2676–87.

149. Reddy PM, Hsu W-Y, Hu J-F, Ho Y-P. Digestion Completeness of Microwave-Assisted and Conventional Trypsin-Catalyzed Reactions. J Am Soc Mass Spectrom. 2010 Mar 1;21(3):421–4.

150. Sun W, Gao S, Wang L, Chen Y, Wu S, Wang X, et al. Microwave-assisted protein preparation and enzymatic digestion in proteomics. Mol Cell Proteomics [Internet]. 2005 Dec 9 [cited 2019 Jun 18]; Available from: https://www.mcponline.org/content/early/2005/12/12/mcp.T500022-MCP200

151. Shilov IV, Seymour SL, Patel AA, Loboda A, Tang WH, Keating SP, et al. The Paragon Algorithm, a Next Generation Search Engine That Uses Sequence Temperature Values and Feature Probabilities to Identify Peptides from Tandem Mass Spectra. Mol Cell Proteomics. 2007 Sep 1;6(9):1638–55.

152. Emerson D, Agulto L, Liu H, Liu L. Identifying and Characterizing Bacteria in an Era of Genomics and Proteomics. BioScience. 2008 Nov 1;58(10):925–36.

153. Mellmann A, Cloud J, Maier T, Keckevoet U, Ramminger I, Iwen P, et al. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. J Clin Microbiol. 2008 Jun;46(6):1946–54.

154. Leli C, Cenci E, Cardaccia A, Moretti A, D'Alò F, Pagliochini R, et al. Rapid identification of bacterial and fungal pathogens from positive blood cultures by MALDI-TOF MS. Int J Med Microbiol IJMM. 2013 May;303(4):205–9.

155. Dupont C, Sivadon-Tardy V, Bille E, Dauphin B, Beretti JL, Alvarez AS, et al. Identification of clinical coagulase-negative staphylococci, isolated in microbiology laboratories, by matrix-assisted laser desorption/ionization-time of flight mass spectrometry and two automated systems. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2010 Jul;16(7):998–1004.

156. Carbonnelle E, Grohs P, Jacquier H, Day N, Tenza S, Dewailly A, et al. Robustness of two MALDI-TOF mass spectrometry systems for bacterial identification. J Microbiol Methods. 2012 May;89(2):133–6.

157. Nie S, Tian B, Wang X, Pincus DH, Welker M, Gilhuley K, et al. Fusobacterium nucleatum subspecies identification by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol. 2015 Apr;53(4):1399–402.

158. Florio W, Tavanti A, Barnini S, Ghelardi E, Lupetti A. Recent Advances and Ongoing Challenges in the Diagnosis of Microbial Infections by MALDI-TOF Mass Spectrometry. Front Microbiol [Internet]. 2018 May 29 [cited 2019 Jun 12];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5986882/

159. Sauer S, Kliem M. Mass spectrometry tools for the classification and identification of bacteria. Nat Rev Microbiol. 2010 Jan;8(1):74–82.

160. Swatkoski S, Gutierrez P, Wynne C, Petrov A, Dinman JD, Edwards N, et al. Evaluation of microwave-accelerated residue-specific acid cleavage for proteomic applications. J Proteome Res. 2008 Feb;7(2):579–86.

161. Hauser NJ, Basile F. Online microwave D-cleavage LC-ESI-MS/MS of intact proteins: site-specific cleavages at aspartic acid residues and disulfide bonds. J Proteome Res. 2008 Mar;7(3):1012–26.

162. Zhong H, Zhang Y, Wen Z, Li L. Protein sequencing by mass analysis of polypeptide ladders after controlled protein hydrolysis. Nat Biotechnol. 2004 Oct;22(10):1291–6.

163. Lill JR, Ingle ES, Liu PS, Pham V, Sandoval WN. Microwave-assisted proteomics. Mass Spectrom Rev. 2007 Oct;26(5):657–71.

164. Suckau D, Resemann A, Schuerenberg M, Hufnagel P, Franzen J, Holle A. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. Anal Bioanal Chem. 2003 Aug 1;376(7):952–65.

165. Golizeh M, Sleno L. Optimized proteomic analysis of rat liver microsomes using dual enzyme digestion with 2D-LC–MS/MS. J Proteomics. 2013 Apr 26;82:166–78.

166. Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates JR. Protein Analysis by Shotgun/Bottom-up Proteomics. Chem Rev. 2013 Apr 10;113(4):2343–94.

167. Yates JR. Mass spectrometry and the age of the proteome. J Mass Spectrom JMS. 1998 Jan;33(1):1–19.

168. Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham A-JL, Bunk DM, et al. Repeatability and Reproducibility in Proteomic Identifications by Liquid Chromatography−Tandem Mass Spectrometry. J Proteome Res. 2010 Feb 5;9(2):761–76.

Appendices

Subject	Age	Gender	Histopathology
Control # 1	69	Female	
Control # 2	42	Male	
Control # 3	76	Male	
Control # 4	78	Male	
Control # 5	66	Female	
Control # 6	48	Female	
Control # 7	57	Male	
Control # 8	53	Female	
Patient # 1	56	Male	Tubulo-villous adenoma
Patient # 2	78	Female	Tubulo-villous adenoma
Patient # 3	70	Male	Tubulo-villous adenoma
Patient # 4	59	Male	Tubulo-villous adenoma

Table A.1 Age and gender of the subjects participating in the study. The controls were colorectal-adenoma free, while the patients had pre-cancerous colorectal adenoma.

Protein accession and name	Score	Organism
tr|J8VK71|J8VK71_9FUSO V-type ATP synthase subunit I	2.00	*Fusobacterium hwasookii*
tr|Q7P754|Q7P754_FUSNV ATPase	1.96	*Fusobacterium nucleatum* subsp. *vincentii* ATCC 49256
tr|A0A2B7YG85|A0A2B7YG85_FUSNP Hemin receptor	1.89	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|J8V310|J8V310_9FUSO Cytoplasmic protein	1.72	*Fusobacterium hwasookii* ChDC F128
tr|A0A323TYZ3|A0A323TYZ3_FUSNU Uncharacterized protein	1.33	*Fusobacterium nucleatum*
tr|A0A2N6TFT4|A0A2N6TFT4_FUSNU Uncharacterized protein	1.07	*Fusobacterium nucleatum*
tr|J8VJX6|J8VJX6_9FUSO Uncharacterized protein	0.86	*Fusobacterium hwasookii* ChDC F128
tr|U7TJP9|U7TJP9_FUSNU Uncharacterized protein (Fragment)	0.75	*Fusobacterium nucleatum* CTI-6
tr|A5TX11|A5TX11_FUSNP Bacteriophage integrase	0.64	*Fusobacterium nucleatum* subsp. *polymorphum* ATCC 10953
tr|A0A2B7YLR6|A0A2B7YLR6_FUSNP	0.48	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|A0A2B7YE51|A0A2B7YE51_FUSNP	0.35	*Fusobacterium nucleatum* subsp. *polymorphum*
tr|F9ERL7|F9ERL7_FUSNU Uncharacterized protein (Fragment)	0.15	*Fusobacterium nucleatum* subsp. *animalis* ATCC 51191
tr|X7RWR5|X7RWR5_FUSNU Aspartate--tRNA ligase	0.08	*Fusobacterium nucleatum* 13_3C
tr|R9RDV4|R9RDV4_FUSNU Uncharacterized protein	0.06	*Fusobacterium nucleatum* subsp. *animalis* 4_8
tr|A5TVJ6|A5TVJ6_FUSNP Uncharacterized protein	0.06	*Fusobacterium nucleatum* subsp. *polymorphum* ATCC 10953

Table A.2 The list of *F. nucleatum* proteins detected by AB Sciex 4800™ MALDI TOF/TOF instrument (Bruker Daltonics). The score indicates ProtScore.