Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork

The Sudabiome: oral and gut microbiome parameters of the Sudanese population including dietary and cultural [Toombak] metagenomics

Thesis presented by

Amel Sami

MSc, NUI BDS Hons

[ORCID no: 0000-0001-8387-4376]

for the degree of

Doctor of Philosophy

University College Cork

School of Microbiology

Head of School/Department: David Clarke

Supervisor[s]: [R Paul Ross, Catherine Stanton, Imad Elimairi, C A Ryan]

December 2022

Table of Content

Dedication

Dedicated to my wonderfully supportive family; my parents, Sami Ahmed and Sarah Shawki, my husband, Mumin Mukhayer and children.

Declaration

"This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism and intellectual property."

Signed: 

Date:15/12/2022

List of Abbreviations

ABC	ATP binding cassette transporters
ANOVA	Analysis of Variance
ATP	Adenosine Triphosphate
ASA	American Society of Anaesthesiologists
ASA-24	Automated self-administered 24-hour recall questionnaire
ASK1	Apoptosis signal regulating kinase 1
ASVs	Amplicon Variant Sequences
B7-DC	Dendritic cell molecule with stimulatory properties for T cells
B7-H1	Anti-apoptotic receptor
BAMH1W	Conserved regions of the Epstein Barr virus genome
BCL-2	B cell leukaemia/lymphoma 2 protein
BIOM	Biological observation matrix
BLAST	Basic local assignment tool
CCA	Canonical correspondence analysis
CCL20	CC chemokine ligand 20
CCR9+	CC chemokine receptor 9
CD4, CD8	Clusters of differentiation 4, 8
Chao1	An estimator of microbiome based on abundance
CpG	5’-C-phosphate-G-3’
CTAB	Cetyl Trimethyl Ammonium Bromide

CTLA-4	Cytotoxic T lymphocyte associated antigen 4
CXCR3	C-X-C chemokine receptor type 3
DADA2	Divisive Amplicon Denoising Algorithm
DeSeq2	Differential expression analysis
DNA	Deoxyribonucleic acid
EDX	Energy Dispersive X-ray Spectroscopy
ELISA	Enzyme linked immunosorbent assay
EPIC - Norfolk FFQ	European Prospective Investigation into Cancer and Nutrition Norfolk Food Frequency Questionnaire
Eta^2	The measure of effect size
ESTOC	European smokeless tobacco council
ExoU	Exoenzyme
FadA	Adhesin of *Fusobacterium nucleatum*
Fap2	Fatty acid binding protein 2
FDA	U.S. Food and Drug Administration
FDR	False Detection Rate
FETA	FFQ Epic tool for analysis
FFPE	Formalin-fixed paraffin embedded
FFQ	Food Frequency questionnaire
FimA	Fimbriae adhesin
FLASH	Fast Length Adjustment of Short Reads
FTND-ST	Fagerstrom Test for Nicotine Dependence-Smokeless Tobacco

G-CSF	Granulocyte colony stimulating factor
GM-CSF	Granulocyte Macrophage-colony stimulating factor
HIV	Human Immunodeficiency Virus
HNC	Head and Neck Cancer
IARC	International Agency for Research on Cancer
IBD	Inflammatory Bowel Disease
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
Irel	Ireland
ITS	Internal transcriber
IUPAC	International union of pure and applied chemistry
JNK	C-Jun-N-terminal-kinase
Jun/Ap-1	Jun protooncogene/ activator protein 1
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LC/MS	Liquid Chromatography/Mass Spectrometry
LDA	Linear discriminant analysis
LEfSe	Linear discriminant analysis Effect Size
LOD	Limit of Detection
MAPK	Mitogen activated kinase pathways
MDPI	Multidimensional poverty index
MMP	Matrix metalloproteinases

NAB	N-nitrosoanabasine
NAT	N-nitrosoanatabine
NF-k β	Nuclear factor of kappa beta
NLRP3	Nucleotide binding domain-like receptor protein 3
NNK	Nicotine derived nitrosamine ketone
NNAL	4 [methylnitrosamine]-1-[3-pyridyl]-1-butanol
NNN	N-nitrosonornicotine
OSCC	Oral squamous cell carcinoma
OTUs	Operational Taxonomic Units
P2X	Purinoreceptor
P53	Tumour suppressor gene
PCA	Principal Component Analysis
PCoA	Principal Co-ordinate Analysis
PCR	Polymerase chain reaction
PD-1	Programmed death 1
PD-L1	Programmed death ligand 1
PI3K	Phosphatidyl inositol-3-kinase
PICRUSt	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
pRb	Retinoblastoma protein
PTEN	Phosphate tensin homologue
PSL13	Periodate resistant antigen

PyNAST	Python nearest alignment space termination
QIIME	Quantitative Insights into Microbial Ecology
RAS	Retrovirus associated DNA sequence
RDA	Redundancy analysis
RNA	Ribonucleic acid
ROS	Reactive oxygen species
TLR	Toll like receptors
TNF	Tumour necrosis factor
TRIS-EDTA	Trisaminomethane Ethylenediamine Tetraacetic Acid
TSNAs	Tobacco specific nitrosamines
TSS	Total sum scaling
SDGs	Sustainable development goals
SEM	Scanning Electron Microscopy
Sud	Sudan
SudIrel	Sudanese living in Ireland
USA	United States of America
WHO	World health organisation
Wnt	Group of signalling transduction pathways

Publications

Published:

Sami A, Elimairi I, Patangia D, Watkins C, Ryan CA, Ross RP, Stanton C. The ultra-structural, metabolomic and metagenomic characterisation of the Sudanese smokeless tobacco 'Toombak'. Toxicol Rep. 2021 Aug 5; 8:1498-1512. doi: 10.1016/j.toxrep.2021.07.008. PMID: 34401360; PMCID: PMC8355839. **[Appendix 1]**

Sami A, Elimairi I, Stanton C, Ross RP, Ryan CA. The Role of the Microbiome in Oral Squamous Cell Carcinoma with Insight into the Microbiome-Treatment Axis. Int J Mol Sci. 2020 Oct 29; 21[21]:8061. doi: 10.3390/ijms21218061. PMID: 33137960; PMCID: PMC7662318. **[Appendix 2]**

Amel Sami, Imad Elimairi, R Paul Ross and Catherine Stanton - The threat of Toombak in Sudan – TResearch, Spring 2022, Volume 17, Number 4 **[Appendix 3]**

Sami A, Elimairi I, Anthony Ryan C, Paul Ross R, Stanton C. Sudanese Toombak smokeless tobacco users harbour significantly altered long-term cortisol body production. Steroids. 2023 Feb 2;193:109189. doi: 10.1016/j.steroids.2023.109189. Epub ahead of print. PMID: 36738817 **[Appendix 4].**

Amel Sami, Imad Elimairi, R Paul Ross, Heyem Salih, Yasmeen Elyass, Marmar Elsiddig and Catherine Stanton. Sudanese Fermented Foods. The Boolean: Snapshots of Doctoral Research at University College Cork DOI: https://doi.org/10.33178/boolean.2022.12.15 **[Appendix 5]**

Submitted:

Sami A, Elimairi I, Ryan CA, Stanton C, Patangia D and Ross RP. Altered oral microbiome in Sudanese smokeless tobacco users carries a newly emerging risk of squamous cell carcinoma development and progression. Scientific reports. Cancer microbiome guest edition.

Sami A, Elimairi I, D Patangia, Ryan CA, Ross RP, Stanton C. The Mycobiome of the Sudanese smokeless tobacco Toombak. Fungal Ecology Elsevier.

List of Conference Abstracts

1. Amel Sami, Imad Elimairi, R Paul Ross, C Anthony Ryan and Catherine Stanton [2022]. First oral microbiome characterisation amongst the Sudanese population and how use of the smokeless tobacco; Toombak can influence the comrade of oral microorganism life. The Dublin University Microbiology Society – Focus meeting 'Microbiomes and Human Health', 13th June 2022 **[Appendix 6]**
2. First time use of the *European Perspective Investigation into Cancer [EP- IC] - Norfolk* food frequency questionnaire on the Sudanese population, retrieving key nutrient intake and dietary consumption patterns Microbiology Research Day, University College Cork 4th May 2022 **[Appendix 7]**
3. Amel Sami, Imad Elimairi, Catherine Stanton, R Paul Ross and CA Ryan, Gut microbiome traditional memory and new reshaping portfolio in relation to migratory movements of the Sudanese population to westernised environments Microbiology Society Annual conference – Belfast, UK 6th April 2022 **[Appendix 8]**
4. Amel Sami, Imad Elimairi, R Paul Ross, C Anthony Ryan and Catherine Stanton. The oral manifestations of the Sudanese smokeless tobacco and its effects on long term cortisol body production [British and Irish Association of Oral Medicine – 18th November 2021, Online **[Appendix 9]**
5. Amel Sami, Catherine Stanton, R Paul Ross. 1st Teagasc Food Research Day, 28th-29th September 2021 - The methodologies and microbiome of Sudanese fermented foods **[Appendix 10]**
6. EU AFRICA SUMMIT 2021 - Case study: EU investment in science research in Sudan and the microbiome role in Africa, Online **[Appendix 11]**
7. Amel Sami, Imad Elimairi, R Paul Ross, C Anthony Ryan and Catherine Stanton. European association of oral medicine 21 - Buccal microbiome: A pilot study on the Sudanese population and the use of the smokeless tobacco, Toombak, on microorganisms of the buccal cavity. European Association of Oral Medicine conference June 2021 **[Appendix 12]**
8. 5th Esther partnership forum – The APC Ireland and National Ribat University Khartoum Sudan partnership 11th November 2020 **[Appendix 13]**

9. Amel Sami, Imad Elimairi, Catherine Stanton. American Academy of Oral Medicine Conference, San Antonio, Texas, USA April 2018 - Toombak… the new king of smokeless tobacco **[Appendix 14]**

Abstract

The Sudabiome is a unique metagenomic project focusing on several critical areas of the Sudanese population. It highlights a pivotal framework in understanding novel aspects of Sudanese health that include oral and gut health, cultural metagenomics, nutritional trends, migratory travel impacts, and the susceptibility or protection against diseases.

Toombak is a fermented smokeless tobacco produced from the *Nicotiana Rustica* tobacco plant, utilised predominantly by Sudanese males. The Toombak is 'dipped' in the oral cavity and replaced several times a day. The microbiome, mycobiome, phylogenetics and metabolomics of 21 pre-prepared and ready to buy Toombak samples purchased from different regions in Khartoum city were assessed, as well as the pH, chemical and microscopical composition of the product. The heavy metals, chromium, cobalt, and copper were high in the pre-prepared form of Toombak, while iron, tobacco-specific nitrosamines [TSNAs], formaldehyde and acetaldehyde were considerably elevated in both types compared to EU regulated smokeless tobacco products.

The phyla, *Firmicutes* and *Actinobacteria* dominated all samples of Toombak. *Virgibacillus* was found to be significantly increased in the pre-prepared form [q=3.6314e-8] compared to the ready to buy forms, while *Corynebacterium casei* [q=1.6417e-5], *Atopococus tabaci* [p=0.05], and *Staphylococcus gallinarum* [p=0.01] were the genera abundant in the ready to buy Toombak in comparison to the pre-prepared form. The mycobiome of the ready to buy Toombak was found to be enriched with *Aspergillus* [85.54%] with the species, *Aspergillus heterocaryoticus* [43.27%] and *austwikii* [39.48%] dominating the samples. Furthermore, utilising PICRUSt, the ready to buy samples were found to harbour an increased activity of metal transport systems [K02006, K02013 and K02015] and an antibiotic transport system [K09687]. The Toombak had a large non-homogenous and irregular particle distribution with increased sodium, while the pH of samples was in the alkaline range [8.73-9.92]. TSNA, formaldehyde and acetaldehyde levels observed in Toombak were

found to be some of the highest compared to other smokeless tobaccos from around the world such as snus from Sweden and moist smokeless tobacco from the USA. These findings highlight that the final composition of Toombak is affected by its preparation method, with Toombak use having the potential to impart many negative consequences on the health of its users.

Dependency on Toombak use was examined through the utilisation of the Fagerstrom test for nicotine dependence smokeless tobacco questionnaire. Nineteen chronic Toombak users had an 85% dependence score. In addition, stress level over a 3-month retrospective period was evaluated through 'scalp-side' hair cortisol analysis. Mean cortisol levels were significantly lower [p=0.023] amongst Toombak users [9.7 pg/ml] compared to non-users [19.4 pg/ml]. While use of Toombak may initially bring anxiolytic effects to its users, blunting of the hypothalamic-pituitary axis chronically ensues in Toombak users, affecting cortisol release. This is likely due to the continued high levels of nicotine exposure and toxic effects that many Toombak compounds cause to body organs, including the adrenal glands.

The Sudan has a rich heritage of food fermentation that involves a vast array of raw starting materials and methodologies of production that have been preserved for centuries. Forty-six Sudanese fermented foods were sourced from six food categories that included, crop [sorghum and millet], plant [Cassia obtusifolia and sesame], fish, animal, and dairy sources to establish these foods metagenomic composition, some of which were explored for the first time. 16S rRNA sequencing was undertaken, while a subset of foods underwent internal transcriber [ITS] sequencing for mycobiome analysis. Beta diversity [p=0.001] was significantly varied between the categories of Sudanese fermented foods. Animal, fish, and plant-based foods exhibited the highest alpha diversity richness [p=8.7229e-07] while fish and plant-based foods had an elevated Shannon index [p=0.001]. Crop-based foods [sorghum and millet] were found to be enriched in the genus *Acetobacter* [96.89%], plant-based foods in *Aeriscardovia* [99.14%], dairy-based foods in *Enhydrobacter* [85.03%], animal-based foods in *Bacteroides* [98.22%] and fish-based foods in *Sporosarcina* [99%].

A loss of *Lactobacillus* abundance was observed between the preparatory [26.91%] and final [4.38%] stages of sorghum and millet fermented foods. The Kawal food,

obtained from the fermentation of the Cassia obtusifolia plant was found to harbour a comprehensive microbiome [*Bacillus, Lactobacillus, Pediococcus, Citrobacter* and *Bifidobacterium*] and mycobiome [*Candida, Aspergillus, Cladosporium* and *Malassezia]* which could potentially harbour unknown but novel probiotic benefits to consumers. Sudanese fermented foods were however also found to be a source for pathogenic bacteria that included *Escherichia-Shigella* found abundantly in the plant-based foods [46.8%] and *Wohlfahrtiimonas* which were found to be abundant in animal-based foods [79.07%]. Intra-food categorical analysis found *Lactobacillus* to be enriched in the dairy food Mish or deep yoghurt [86%] compared to Gergosh or the dairy/legume biscuit [24%] and Garis or fermented camel milk [10%].

Extensive screening of the oral microbiome of users and non-users of Toombak was undertaken through 415 samples detecting salivary [n=72], plaque [n=71], tongue, buccal cheek, floor of the mouth and palatal mucosal [n=272] microbiome variations. In a subset of salivary samples, the mycobiome [ITS] was also assessed. The microbiome of oral squamous cell carcinoma and premalignant samples [n=46] from formalin-fixed paraffin-embedded tissue was further evaluated from another cohort of users and non-users of Toombak. Alpha diversity [richness] was low for tongue microbiome compared to other mucosal locations [p=3.6849e-34] and Beta diversity was found to be significant between the four oral mucosal locations [p<0.001].

Fusobacteria and *Patescibacteria* in saliva and *Actinomyces* in the saliva [p=0.0045], tongue [p=0.013] and palate [q=0.001], were correlated with Toombak use. Skin-associated microbiome were increased amongst the oral cavity of Toombak users and included *Staphylococcaceae* [q=0.037] in the saliva and *Cutibacterium* [q=0.04] in the buccal cheek and floor of the mouth. In Toombak users, utilising LEfSe plotting, the genus *Peptostreptococcus* was most distinct for Toombak use. Non-users of Toombak were found to have abundances in *Prevotella* [buccal, p=0.04] and *Bifidobacterium* [tongue, q=0.0049] while *Scardovia* was discriminant of non-user's tongue microbiome. Virulent strains such as *Prevotella nigrescens* and *Streptococcus equinus [*tongue, p=0.043] were found in the oral microbiome of Toombak users while *Prevotella salivae* and *Streptococcus sobrinus* [tongue, q=0.023] were found in non-users of Toombak. There was a three-fold enrichment of oral *Aspergillus* [78.93%] in the saliva of Toombak users compared to non-users [21.07%] with reduced *Candida*

abundance in Toombak users [4.33%] compared to non-users [95.67%]. Virulent fungal species were also found in smokeless tobacco users [*Candida tropicalis*] compared to non-users [*Candida albicans*].

Premalignant lesion microbiome composition included *Rothia* [p=1.5e-10] and *Peptostreptococcus* [p=6.5e-07] agreeing with the literature. The oral cancer microbiome in Toombak users harboured genera *fav*ouring a poor survival and metastasis that included *Stenotrophomonas,* [p=0.043] and *Schlegelella,* [*p=0.048],* compared to genera found in oral cancer samples from non-users of Toombak such as *Lactobacillus,* [p=0.024]. The genus *Corynebacterium_1* was found to be common to the Toombak product, the oral cavity [saliva, p=0.0054, floor of the mouth, p=0.0028] and oral cancer samples [p=0.044] of Toombak users. These findings highlight that several oral distinctions follow Toombak use, allowing for a microbiome that could potentiate disease and destabilise the normal oral microbial flora while our findings also harbour the potential for microbiome biomarkers to be employed in the future screening of oral carcinoma in Sudan.

In 141 participants, an inter-continental metagenomics study was achieved on residents from Africa [Sudan] and Europe [Ireland] concerning diet frequency, microbiome composition and metabolome activity of stool samples. The European Prospective Investigation into Cancer and Nutrition Norfolk Food Frequency Questionnaire [EPIC - Norfolk FFQ] was modified and employed to compare food groups, nutrient and energy intakes between three cohorts, the Irish living in Ireland [Irel], the Sudanese living in Ireland [SudIrel] and the Sudanese living in Sudan [Sud]. Cereals and cereal products, fats and oil, fish and fish products, fruits and vegetables, meat and meat products, sugar preserves and snacks were consumed at a reduced intake in the Sud cohort with a high strength of association and large effect size [eta^2 = >0.14]. Egg and egg products [eta^2 =0.04], nuts and seeds [eta^2=0.05] and non-alcoholic beverages [eta^2=0.108] were also consumed highly in Sud, albeit with low and medium effect sizes. The average energy intake [measured in kcal/day] between groups was statistically different [p <0.001] and was found to be lowest in Sud [1077 kcal/day], compared to SudIrel [2729 kcal/day] and Irel [3442 kcal/day]. The diet of the Sud cohort was thus found to be low in calcium [p=0.05], carbohydrates, fibre, protein, fat, potassium, sodium, and vitamins A, C, B1, B2, B6, B12 and D [p<0.001]

but high in vitamin B5 [pantothenic acid] and intermediate in Vitamin K levels. These results highlight the variations of dietary consumption between the two populations; Sudan and Ireland and helps set a fundamental previously unmet base for future research to evaluate dietary trends in the Sudanese population, including children.

Microbiome Beta diversity plotting [q=0.001] highlighted a distinct 'sandwiched' presentation to the cohort SudIrel compared to Irel and Sud groups, with R^2 = 63%. Alpha diversity measures however were found to be non-significant between groups [Shannon index or evenness, p=0.74]. Microbiome genera such as *Prevotella*, [q=3.169e-06], *Megasphaera* [q=4.61e-05], *Klebsiella* [q=3.9067-e], *Bifidobacterium* [q=7.023e-03] and *Lactobacillus* [q=2.918e-09] were found to be depleted in the Irel cohort compared to both Sud and SudIrel while in the Irel cohort, the microbiome was instead enriched in *Lachnospira* [q=1.623e-09] and *Bacteroides* [q=3.056e-05]. *Collinsella* was found to be four times more abundant in Sud and SudIrel [44%] compared to Irel [11%].

Principal coordinate analysis of metabolomic compounds from stool samples highlighted 27 metabolites to be significantly varied between Irel and Sud including proline [p=0.003], adipic acid [0.04], and phosphocholine [p=0.02], 21 metabolites to be significantly varied between Sud and SudIrel including sugar alcohol [p=0.04] and N-acetylneuraminic acid [p=0.01] and no significant metabolomic variations between Irel and SudIrel. The short-chain fatty acids, isobutyric [p=0.0] and isovaleric acid [0.002] were significantly lower in the stool samples from those in the Irel cohort. LEfSe plotting of PICRUSt KEGG orthology pathways indicated multi-drug resistance activity pathways in the Irel cohort [K09686], mineral and heavy metal transport activity in Sud [K06199] and in SudIrel, carbohydrate metabolism pathways were prominent [K02794, K02796, K02775, K02773, K02774]. From these results, a migratory impact on the traditional microbiome of the Sudanese population occurs that has not been previously explored which may allow for both the susceptibility and prevention of disease in this population. Those who migrate from the Sudan to Ireland [SudIrel], carry both tradition microbiome preservation [i.e., continued *Prevotella* abundance] and modern microbial carriage [i.e., *Bacteroides* enrichment].

Through these studies, there has been a novel and in depth uncovering of human and environmental microbiome features associated for the first time with the Sudanese population in comparison to Irish controls. Through fermentation techniques for example, cultural [Toombak] and dietary [fermented foods] produces may carry health benefits but also numerous predicaments such as the increased risk of oral cancer development and systemic cortisol imbalance from Toombak use and the contraction of food-borne diseases from contamination during Sudanese fermented food production. The studies on Sudanese human oral and gut health have further provided insight into how the oral and gut microbiome may respond to modern challenges such as migration and economic, dietary and cultural influences.

This project further give light on the merging of next generation sequencing science with economic and population challenges. Continued microbiome research in Sudan should continue to be geared towards sharper preventative disease strategies, management, and treatment outcomes.

1 The role of the microbiome in oral squamous cell carcinoma with insight into the microbiome – treatment axis [literature review]

Amel Sami 1, 2, Imad Elimairi 2, *, Catherine Stanton 1, 3, R. Paul Ross 1 and C. Anthony Ryan 4

1 APC Microbiome Ireland, School of Microbiology, University College Cork, Cork T12 YN60, Ireland.
2 Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, National Ribat University, Nile Street, Khartoum 1111, Sudan
3 Teagasc Food Research Centre, Moorepark, Fermoy, Cork P61 C996, Ireland
4 Department of Paediatrics and Child Health, University College Cork, Cork T12 DFK4, Ireland.

This chapter has been published as **Sami A**, Elimairi I, Patangia D, Watkins C, Ryan CA, Ross RP, Stanton C. The ultra-structural, metabolomic and metagenomic characterisation of the Sudanese smokeless tobacco 'Toombak'. Toxicol Rep. 2021 Aug 5; 8:1498-1512. doi: 10.1016/j.toxrep.2021.07.008. PMID: 34401360; PMCID: PMC8355839. [see Appendix 1]

1.1 Abstract:

Oral squamous cell carcinoma [OSCC] is one of the leading presentations of head and neck cancer [HNC]. The first part of this review will describe the highlights of the oral microbiome in health and normal development while demonstrating how both the oral and gut microbiome can map OSCC development, progression, treatment, and the potential side effects associated with its management. We then scope the dynamics of the various microorganisms of the oral cavity, including bacteria, mycoplasma, fungi, archaea, and viruses, and describe the characteristic roles they may play in OSCC development. We also highlight how the human immunodeficiency viruses [HIV] may impinge on the host microbiome and increase the burden of oral premalignant lesions and OSCC in patients with HIV. Finally, we summarise current insights into the microbiome–treatment axis pertaining to OSCC and show how the microbiome is affected by radiotherapy, chemotherapy, immunotherapy and also how these therapies are affected by the state of the microbiome, potentially determining the success or failure of some of these treatments.

1.2 Introduction

In the Global Cancer report 2018, HNC is thought to be the eighth most commonly occurring cancer in the world [1], with oral squamous cell carcinoma [OSCC] being the most common presentation. According to The International Classification of Diseases for Oncology, World Health Organisation, oral cavity cancer includes the locations of the inner lip [C00.3–C00.9], the tongue, excluding lingual tonsil [C02], the gum [C03], the floor of the mouth [C04], the palate [C05], and other or unspecified parts of the mouth [C06] [2]. Although, in some parts of the world, there is a decreasing trend of OSCC [substituted, in part, by the increasing trend of human papilloma virus-associated oropharyngeal cancer] [3], incidence of the disease is very much still relevant all over the world. In Africa, OSCC statistics are highly underreported, but, in a review by Faggons et al. [2015], the smokeless tobacco products Toombak and Kola nut, used in Sudan and Western African countries, respectively, were the cause of the greatest number of head and neck cancer cases on the continent [4]. Meanwhile, Southeast and Central Asia accounts for almost half of OSCC globally [3], while, in the United Kingdom [up to 2016], there were 52,829 oral

cavity cancer cases reported. In the Republic of Ireland, oral cancer cases are said to be significantly increased in both genders, but particularly amongst women [2,5]. Tobacco [cigarette smoking and smokeless tobacco] and alcohol remain the most established risk factors for the development of OSCC. Betel quid is commonly used in Asia and is the cause for the high oral cancer burden there, while other known aetiological predispositions include low socioeconomic status, poor intake of fruit and vegetables, poor oral health, genetic diseases associated with DNA repair and chronic inflammation. Indeed, on that latter note, chronic inflammation is thought to be the cause of one in four cancers [6]. How this may relate to the aetiopathogenesis of OSCC is an important undertaking. Compelling evidence now exists for the association of chronic periodontal disease and the development of head and neck cancer in general and OSCC specifically. For every millimetre of alveolar bone loss, a four-fold increased risk of developing squamous cell carcinoma of the head and neck cancer unfolds. Furthermore, those with periodontal disease are thought to be at greater risk of poorly differentiated cancers [7,8]. Since periodontal disease is a microbial disease, this sets the tone for the question: what is the role of microorganisms in the aetiology of OSCC? This review sets out to answer this question, delving into how both the oral and gut microbiome may sculpt, in many ways, the outcome of an OSCC from start to finish.

1.3 Review

Role and Function of the Microbiome

The microbiome, including oral, is the genome of all microorganisms, their interactions and their products which include bacteria but also viruses, archaea and fungi [9,10]. The "bacterial" microbiome resides in the lumen of the gut and the oral cavity; however, the skin and vagina also harbour a diverse microbiome. Bacteria in the gut reside either on the epithelium or pass through the lumen, both offering important yet different roles in maintaining homeostasis between the host and the microbiome and have the distinct ability, from an early age, to sculpt the systemic immune response, whereby constant interactions between the gut microbiome and the mucosa regulate inflammation and mediate immune tolerance through bacterial translocation [11,12]. Commensal bacteria help sustain a healthy immune state throughout the whole body, serving to educate the various stages of development of different immune cells that later become necessary in the fight against cancer cells,

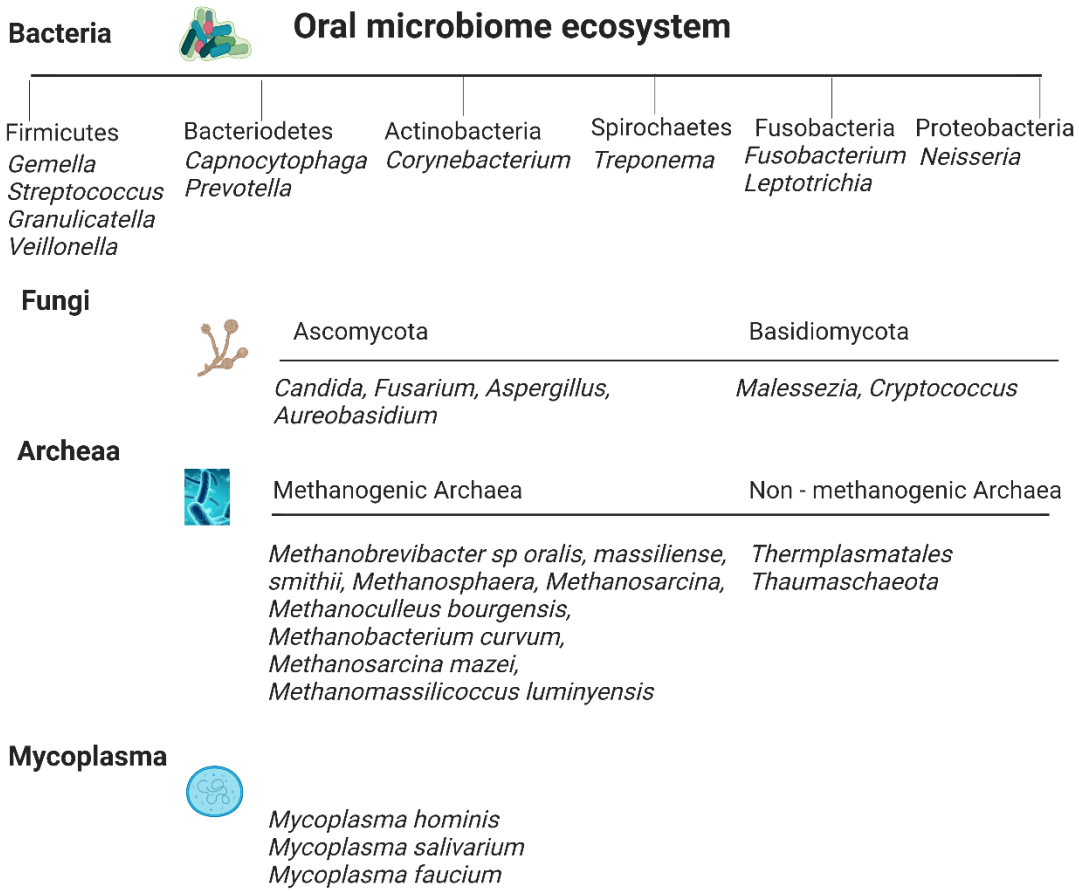
including OSCC [13]. The functions of the gut microbiome include the metabolism of diet components and vitamin production, as well as release of short chain fatty acids, microbial metabolites that further promote anti-inflammatory responses such as the release of interleukin [IL] 10, the inhibition of IL-6, tumour necrosis factor alpha [TNF] [14] and the activation of G protein-coupled receptors, present on epithelial cells as well as immune cells, and have a wide range of immune-inflammatory responses, including tumour cell deactivation [15,16]. In the oral cavity, the normal microbiota colonise space, preventing pathogenic microbes from establishing themselves as well as exerting healthy microbial antagonism and quorum sensing.

Meanwhile, the host immune system is quite successful in maintaining a tolerant state, mainly through the function of mucosal dendritic cells and the secretion of anti-inflammatory cytokines such as IL-10, and prostaglandin E2, which suppress the immune system through the formation of T regulatory cells [17], thus displaying an intricate and well-ordered harmony of stability between oral mucosal dental tissue health and the different microorganisms' requirements for survival. It is only when the environment starts to favour shifts towards a dysbiotic picture that pathogenic bacteria begin to act at the forefront, unravelling their roles in various oral diseases.

Understanding the Microbiome first in health, then in disease

Delwhirst et al. [2010], assembled the Human Oral Microbiome project, identifying 619 bacterial taxa in 13 phyla, utilising 16S rRNA gene sequencing technology [9]. Figure 1 depicts the classification of the oral microbiome ecosystem in health. The most abundant genera of the oral cavity are noted with regard to their respective phyla, while archaea are recognised as either methanogenic or non-methanogenic forms and have been isolated predominantly from subgingival pockets. Fungal genera are also described according to their respective divisions; however, 30% of fungal communities remain largely unculturable. Protozoa such as *Entamoeba gingivalis* have been identified from within the oral cavity [18], while the commonest mycoplasma include *Mycoplasma salivarium* and *orale* [19] and are regarded as parasitic organisms.

Figure 1:



1:The oral microbiome ecosystem classification

A large part of the oral microbiome or "core" microbiome is comparable between individuals; however, every individual also has unique factors [age, behavioural habits, oral health especially saliva, immune status] that can change and modify the microbiome over time. *Bacteroidetes* [mainly the genus *Prevotella*], as well as the *Veillonella* genus, are found to increase in abundance with age, while others such as *Granulicatella* become decreased [22]. Pre-tooth eruption years predominantly display aerobic species such as *Streptococcus salivarius* while post-tooth eruption exhibits a microbiome that is much more diverse with new inhabitants to the hard surfaces of teeth and gingival crevices, particularly anaerobic bacteria. Iatrogenic

environmental changes such as the placement of restorations, orthodontic appliances, implants and dentures can all place a map change in the oral microbiome. *Staphylococcus* species have been shown to have a high affinity for the titanium of implants [23], while those with orthodontic appliances have increased *Streptococcus mutans* and *Lactobacillus* species [24]. In a study by Si et al. [2016], the subspecies *Fusobacterium nucleatum animalis* was found to exclusively colonise the dentures of patients with denture stomatitis. This subspecies has been associated with pathogenic disease, including colorectal cancer, cardiac disease and stillbirth [25]. Neonates born by vaginal delivery, those delivered full term and who undergo breastfeeding are generally in the landscape of a healthy microbiome development, with a much more diverse and stable microbiome profile, while those who are born by Caesarean section, delivered premature, provided with formula milk or exposed to antibiotics early in life generally favour a more dysbiotic course of microbiome development [15,26].

Indeed, any dysbiosis that occurs in the gut microbiome has been associated with many diseases that include IBD, type two diabetes, atherosclerosis, obesity, mental disorders, and cancers of the gastrointestinal tract, as well as lung, breast and haematogenic malignancies [27,28]. Importantly, the oral microbiome is an important factor to consider not just in the pathogenesis of OSCC, but also in non-head and neck-related cancers. Oral bacteria have been implicated in oesophageal squamous cell carcinoma and adenocarcinoma [*Porphyromonas gingivalis* and *Tannerella forsythia*, colorectal cancer [*Treponema denticola, Prevotella intermedia*] and pancreatic cancer [*Actinobacillus Actinomycetemcomitans* and *Porphyromonas gingivalis*] [29]. Bacterial communities of a healthy person and those of an OSCC patient have been shown to differ considerably regarding function, such that those bacterial genes involved in methane metabolism, oxidative phosphorylation and carbon fixation are enriched in the OSCC microbiome, while those involved in amino acid metabolism and folate biosynthesis are not. [30]. Some members of the microbiome are regarded as “key players”, whereby, even in low abundance, their effect is powerful enough to impact on the remaining environment [31]. A richer and diverse microbiome with a higher number of reads and operational taxonomic units is generally well recognised in healthy persons, while edentulousness and disease affiliations have been associated with reduced richness and low diversity in the microbiome [32]. Guerrero-Preston et al. [2016], highlighted this phenomenon in OSCC cases specifically, where the total

number of bacterial operational taxonomic units amongst OSCC samples was 3659, compared to that of healthy individuals [13,849] [33].

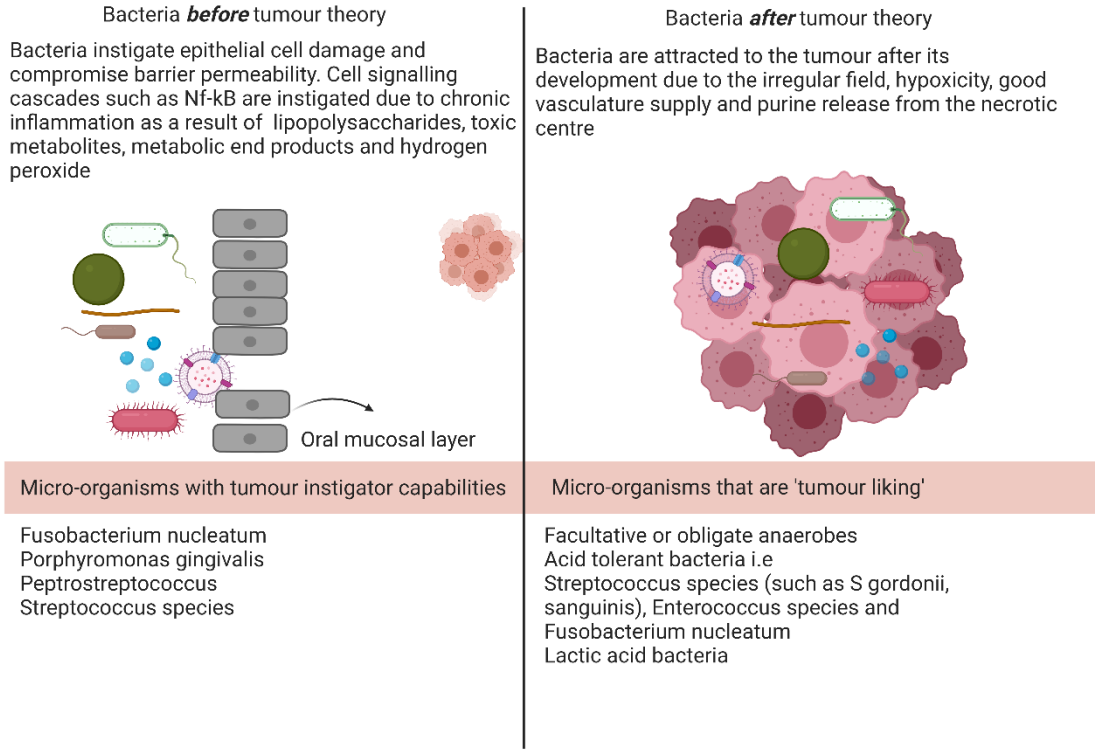
Microbiome and pathogenetic mechanisms underlying OSCC

Ever since Virchow in 1863, the link between "bacteria, inflammation and cancer" progression has been put forward [34]. A statement more applicable to the present day would include how the oral microbiome, the epithelial barrier, the immune system and chronic inflammation exist in an elaborate oncogenic parallelogram [35]. A background of inflammation in relation to cancer development occurs in the context of infections, autoimmune disease and chronic and low-grade stimuli, such as from continuous irritation [36]. The oral microbiome not only drives the chronic inflammation that may precede OSCC but is also involved in the direct influence on the host cell response [37]. Yost et al. [2018], showed that *Fusobacteria* had the highest number of transcripts or active genes in OSCC sites than any other bacteria, followed by *Selenomonas* and *Prevotella,* while *Bacillales, Gemella* and *Neisseria* evidenced more activity in healthy sites. Metatranscriptomic analysis of the OSCC-associated microbiome or "oncobiome" has shown the specific activities that are undertaken by these bacteria such as increased metal ion transport and nitrous oxide reductase, as well as tryptophanase and protease activity [38]. Increased ion transport around cancer sites is associated with catalysing radical ions and the promotion of cancer cell growth [39,40]. Other functional activities of the OSCC-associated microbiome include anaerobic respiration, proteolysis and the response against oxidative stress or radical species damage [41].

To provide lucidity in the relationship between the oral microbiome and OSCC, the "who came first" concept is discussed and illustrated in Figure 2. Here, two broad theories are formulated, in that either a single or group of bacteria may be the direct causative or driving factor in the pathogenesis of OSCC or may have the ability to restructure the microbiome to a more oncobiomic environment [bacteria before tumour] that damages healthy epithelial cells, or that bacterial presence in the OSCC tumour environment is opportunistic and is established after tumour development [bacteria after tumour]. In either concept, nevertheless, it is the metagenomic and transcriptomic events which occur in the oral microbiome, that influence the change in oral cells and in conjunction with the host cell and immune response, these all have

the final say in OSCC development. Furthermore, it is highly likely that a combination of both these theories occurs throughout OSCC progression.

Figure 2:



2: The possible connections between bacteria and development of oral squamous cell carcinoma

In the bacteria after tumour theory, there are many factors that facilitate the attraction of bacteria to a tumour, including the irregular geographical anatomy and increased vasculature of a cancerous lesion, the hypoxic nature of the tumour environment that particularly favours facultative or obligate anaerobic, saccharolytic and acid-tolerant bacterial genera, the reduced host immunity at the tumour site, the presence of purines in the necrotic centre of the tumour, which can serve as nutrients for bacterial growth and cancer cell chemoattraction, as well as adhesion [through cell membrane glycoconjugates] to bacteria [42]. Some of the commonest Phyla associated with the OSCC environment include *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Fusobacteria*, and *Proteobacteria* [43].

These arriving bacteria then reciprocate a synergistic relationship at the cancer site, indirectly promoting the disadvantage of tumour growth and progression, similar to the red complex associated with periodontal disease [44,45]. *Oribacterium* for example, are known to produce acetate as a metabolic by-product, a substrate that is quickly utilised by hypoxic growing OSCC tumours [46].

In the bacteria before tumour theory, the oral mucosa can release numerous inflammatory mediators in response to pathogen-associated molecular patterns of bacteria that include lipopolysaccharides, polysaccharides, peptidoglycans, flagella and microbial DNA and RNA. These activate the host pattern recognition receptor molecules that include toll-like, lectin, nod-like and retinoic acid-inducible gene I-like receptors. Toll-like receptors [TLR] are the most prominent pathogen recognition receptors and can be classified as cell surface or intracellular forms. The cell surface types include TLR4 and TLR5, which recognise bacterial lipopolysaccharides of Gram-negative bacteria and bacterial flagellin, respectively, while intracellular TLRs generally recognise bacterial nucleic acids [47].

The activation of TLRs by bacteria predominantly triggers the MyD88 pathway, which culminates in the activation of the transcription factor; nuclear factor of kappa beta [NF-kβ] and mitogen-activated protein kinase pathways [MAPK], responsible for the many proinflammatory pathways that sustain chronic inflammation [48]. Ultimately, this can eventually progress to haphazard oral cell proliferation and cytoskeletal abnormalities, as well as the inhibition of apoptosis of mutated cells that occurs via the modulation of the Bcl-2 family of proteins and the inactivation of retinoblastoma protein [pRb] as well as the activation of transcription factors such as STAT3 and NFAT [49].

Lipopolysaccharides [endotoxin forms of pathogen-associated molecular patterns in Gram-negative bacteria] also promote the release of IL-1, stimulating several pro-inflammatory cytokines such as IL-6 and TNF alpha, as well as vascular endothelial growth factor. The induction of matrix metalloproteinase [MMP] 9 through IL-1 can degrade the basement membrane and allow for an invasive and aggressive picture of OSCC [50]. The release of IL-6 further induces oxidative stress and can aid the adherence of cancer cells to endothelial cells by the upregulation of intracellular

adhesion molecules. IL-6 also allows for the transcription of many anti-apoptotic gene pathways [37]. Bacteriocins and other proteins such as proteases [which can act like exotoxins] can stimulate host cells to release collagenases, prostaglandins and thromboxane, which further sustain mucosal damage, as well as degrading the immune products of the oral cell, or illicit the release of transforming growth factor β, which can cause abnormal proliferation in cells [51].

TNF alpha primarily exerts its effects through the NF-k β pathway [52], observed to be upregulated in OSCC [53], and is a key pathway in bacterial-associated inflammation. Chronic levels of TNF alpha production have been shown to promote DNA damage by reactive oxygen species [ROS], and support tumour growth through angiogenesis [54]. ROS are involved in all stages of cancer development: initiation, promotion and progression. Thus, the oral microbiome may be an important source for unopposed oxidative stress in epithelial cells that may finally lead to unrepaired DNA mutations, cell division of these abnormally mutated cells and their eventual progression to metastasis [55].

Bacteria are also involved in the release of damaging metabolic end products such as volatile sulphur compounds, organic acids, and aldehydes, as well as the release of nitrosatable compounds. Acetaldehyde is a metabolic end product of many bacteria and has the ability to exert DNA damage and cause excess proliferation of the epithelium. Some of its producers include *Streptococcus* species [*S. gordonii, S. mitis, S. oralis, S. salivarius, S. sanguinis*], *Rothia* species, *Porphyromonas gingivalis* and other alcohol dehydrogenase-possessing organisms, including yeasts such as *Candida albicans* [56]. Through nitrosation pathways, bacteria can catalyse nitrosatable compounds to form N- nitroso compounds, which are associated with the development of cancers. This activity can also occur by fungi [57].

Hydrogen peroxide is another metabolic product that can diffuse passively into the epithelial cell, causing chromosome fragmentation and DNA damage such as double-stranded breaks, base modifications and DNA crosslinking. Hydrogen peroxide can also promote the release of oxidants that lead to post-translational modifications, especially in the P53 gene, tumour suppressor gene [6,35]. Hydrogen peroxide may also have oral cancer preventative pathways, and there is evidence in the existing

literature to support the significantly higher expression of nucleotide binding domain-like receptor protein 3 [NLRP3] inflammasomes in OSCC cells that is associated with increased tumour sizes and lymph node metastasis [58], which oral bacterial hydrogen peroxide can suppress [59].

Volatile sulphur compounds, predominantly hydrogen sulphide and methyl mercaptan, are normally known to cause halitosis, and are produced by bacteria such as *Porphyromonas gingivalis, Prevotella intermedia* and *Fusobacterium nucleatum*, i.e., periodontal pathogens. These compounds have now been implicated in oxidative stress and DNA damage in oral cells. Even low levels of hydrogen sulphide can inhibit the enzyme superoxide dismutase, which is critical in preventing destructive ROS build up in human cells, while methyl mercaptan has been implicated in collagen breakdown, including type 4, and thus may have a role in OSCC invasion across the basement membrane [60]. Host proteins may also be metabolised or fermented into sulphides and nitrosamines by *Firmicutes* and *Bacteroides*, potentiating cell mutations [35].

Premalignant Lesions, the Story before OSCC

There are various similarities and differences that exist in the microbiome environment of potentially malignant lesions such as leukoplakia, submucous fibrosis and lichen planus, as well as OSCC and other cancers. Lee et al. [2017], noted significant variations in the salivary microbiome of those with OSCC and oral premalignant lesions in the genera *Bacillus, Enterococcus, Parvimonas, Peptostreptococcus*, and *Slackia* [61]. Indeed, an interesting species is *Rothia mucilaginosa*, which was found to be abundant in leukoplakic lesions, but less so within OSCC lesions, indicating a possible role in early OSCC establishment [62]. An excellent producer but poor detoxifier of acetaldehyde, this species imparts significant oxidative stress in oral keratinocytes [63]. This species has also been found to be increased more so in those who use smokeless tobacco and may have a significant role in OSCC development, particularly in those with a high alcohol consumption.

Other authors have contrarily noted similarities between the microbiome of oral premalignant lesions and OSCC such as *Cloacibacillus, Gemmiger, Oscillospira*, and *Roseburia*, which were found to be abundant in both oral premalignant lesions and OSCC patients compared to controls [61], while similar microbial patterns have been

observed between leukoplakic lesions and colorectal cancer, which include the abundance of *Fusobacterium, Campylobacter, Leptotrichia*, and *Rothia* species [64].

Taking lichen planus as an example, in a study by Choi et al. [2016], the genera *Leptotrichia* and *Acinetobacter* were increased amongst lichen planus cases, while Streptococcus was significantly decreased compared to controls. Notably, bacteria were found to be significantly more abundant in the basal layer and deeper lamina propria tissues of patients with oral lichen planus compared to controls, suggesting a bacterial causation in the liquefaction process associated with lichen planus histology. The rare finding of the intracellular bacterial infection of T cells was also noted and may explain the unique T cell infiltration that defines oral lichen planus [65] and provide a bacterial cause for why some lesions progress to OSCC.

Pro-tumoral Bacterial Strains

Particular genera from the oral cavity that have been implicated in colorectal cancer development include *Fusobacterium, Peptostreptococcus, Porphyromonas* and *Parvimonas* [66]; thus, it is highly likely that such bacteria also have a role to play in the pathogenesis of OSCC. Examples of OSCC-associated genera include *Actinobacillus, Actinomyces, Aggregatibacter, Bacillus, Campylobacter, Catonella, Citrobacter, Clostridium, Dialister, Eikenella, Enterococcus, Exiguobacterium, Filifactor, Fusobacterium, Gemella, Haemophilus, Klebsiella, Lactobacillus, Micrococcus, Moraxella, Oribacterium, Parvimonas, Peptostreptococcus, Porphyromonas, Propionibacterium, Proteus, Pseudomonas, Rothia, Slackia, Staphylococcus, Streptococcus, Streptomyces* and *Tannerella* [33,61,67–73].

In a study by Sakamoto et al. [1999], the bacterial translocation of members of the oral microbiome, in particular *Peptostreptococcus*, occurred from OSCC metastatic tumours to the cervical lymph nodes. This suggests that the oral microbiome may be directly implicated in the invading front [74] and may even be involved in metastatic events. The species *Peptostreptococcus stomatis* has been isolated from both OSCC and colorectal cancer and is thought to contribute to the invasive tumour environment by the production of several acids such as acetic, butyric, and isovaleric acid [37].

A panel of bacterial species have been isolated from OSCC and include *Aggregatibacter segnis, Capnocytophaga gingivalis, Eubacterium saburreum Exiguobacterium oxidotolerans, Fusobacterium periodonticum, Fusobacterium*

nucleatum, Gemella haemolysans and *morbillorum, Johnsonella ignava, Leptotrichia buccalis, Neisseria flava, Neisseria flavascens, Peptostreptococcus stomatis, Porphyromonas gingivalis, Prevotella oris and melaninogenica, Pseudomonas aeruginosa, Staphylococcus aureus, Streptococcus gordonii, parasanguinis* and *salivarius* [43,49,67,70,75,76]. Rare species to inhabit the oral cavity, such as *Bacteroides fragilis* [77], previously unnamed bacteria such as *Actinomyces [oraltaxon_170]* and *Streptococcus [oral taxon 058]* [70,78] and the environmental species *Dietzia psychralcaliphila* and *Gordonia sputi*, have all also been identified from OSCC samples [68]. Some of these species have been researched in more detail, such as *Fusobacterium nucleatum* and *Porphyromonas gingivalis*, which will be discussed in more detail; however, most have not been pondered in relation to both their singleton and polymicrobial roles in the OSCC-associated microbiome.

Pseudomonas aeruginosa is a rare species to be isolated from OSCC but has been implicated in carcinogenesis due to its ability to cause DNA breakage in epithelial cells [79] as well as having the ability to promote invasion and metastatic change. The bacterium promotes inflammation via its flagella, lipopolysaccharides, and Exoenzyme [ExoU] cytotoxin, which can activate the NF-kB pathway and cause IL-8 secretion [80].

Perhaps the most "award winning" species that have been associated with OSCC development are the Gram-negative, anaerobic and invasive bacteria *Fusobacterium nucleatum* and *Porphyromonas gingivalis*, both of which have been shown to induce OSCC in a mouse model [81]. These two species have also been to date, robustly linked with gastrointestinal cancers [82,83]. Although *Fusobacterium nucleatum* is also related to healthy sites of the oral cavity, it has major roles to play in several oral diseases that include chronic and aggressive periodontal disease and endodontic infections, as well as more recently being implicated in OSCC development [84]. *Fusobacterium nucleatum* has been associated with high-grade dysplasia, lymph node metastasis and poor cancer progression [85,86]. The *Fusobacterium nucleatum* subspecies *polymorphum* was found to be significantly increased in 20 OSCC biopsies in a study by ElHebshi et al. [2017] [79].

Fusobacterium nucleatum can induce the expression of β defensin 2 [an antimicrobial peptide] from oral epithelial cells, which causes the release of potent pro-inflammatory

cytokines such as IL-6 and IL-8 [85,87]. Increased β defensin 2 has been implicated in the pathogenesis of the potentially malignant disorder oral lichen planus, as well as OSCC of the tongue [88]. Through B defensin 2, mast cell degranulation also occurs, causing histamine release that further stimulates β defensin production; thus, a perpetuating cycle of chronic inflammation is allowed to continue and become established with the release of proinflammatory cytokines, the recruitment of immune cells and progressive angiogenesis [88].

An adhesin of *Fusobacterium nucleatum*, FadA, can bind to E-cadherin on epithelial cells, deactivating it to promote mucosal permeability. In turn, free intracellular β catenin allows for increased transcription of Wnt, a transcription factor involved in increased cell proliferation. Fusobacterial proteins such as Fap 2 can block and downgrade natural killer and T cell activity by the binding of Fap 2 with the inhibitory T cell immunoreceptor [28]. Increased activity of MMP9 and MMP13 also occurs, which allows for tumour invasion and metastasis by the degradation of the basement membrane [89].

Fusobacterium nucleatum has been associated with a subtype of colorectal cancer that exhibits molecular gene silencing by CpG island methylation. In OSCC lesions with CpG island methylation, over expression of the immune checkpoint receptor programmed death-ligand 1 [PD-1] has also been shown and thus these tumours may be more sensitive to immune checkpoint blockade [90]. Whether OSCC lesions associated with *Fusobacterium nucleatum* abundance have a better prognosis if treated with drugs such as Pembrolizumab requires further research.

Porphyromonas gingivalis, on the other hand, has been identified as an independent and significant risk factor in cancer-related deaths in the oral cavity and throughout the remaining oral digestive tract, pharynx, oesophagus, stomach, pancreas, liver, and colon and rectum [91]. Higher levels of antibodies against *Porphyromonas gingivalis* infection have been associated with a much higher increased risk of pancreatic cancer [92]. *Porphyromonas gingivalis* has been isolated from non-periodontal disease locations such as the deep crypts of the tongue and the buccal mucosa and is becoming more and more entrenched in its unique and somewhat hidden role in OSCC development [93]. Katz et al. [2011], isolated *Porphyromonas gingivalis* at both higher and more intense levels from OSCC samples as well as from poorly

differentiated OSCC cells; however, further research is required to understand if this bacterium has a specific and common role in the differentiation of OSCC tumours [69]. *Porphyromonas gingivalis* imparts much interference on the host response to oral cancer. The bacterium can activate and complement TLR 2 and 4 and causes the release of pro-inflammatory cytokines such as IL-8. *Porphyromonas gingivalis* works intracellularly to prevent cell apoptosis through the inhibition of cytochrome C, downregulation of caspase 3 activity, secretion of the anti–apoptotic enzyme nucleoside diphosphate kinase that causes the inhibition of the purinoreceptors [P2X] on epithelial cells and the upregulation of microRNA 203, with the subsequent result of apoptosis quelling [94,95] while also upregulating the anti-apoptotic genes Bcl-2 and survivin [28,96,97]. In a study by Geng et al. [2010], *Porphyromonas gingivalis* was found to be associated with the upregulation of colon cancer-associated transcript 1, a long, non-coding RNA that has been found to be upregulated in several cancers and has also been described in oral epithelial cells, as well as in the upregulation of the enzyme production; nicotinamide N methyltransferase, which has been associated with malignancy and cancer stem cells, and was also detected in OSCC patients [93]. Gingipains of *Porphyromonas gingivalis* can cleave and activate MMP9 to its mature form, which can degrade the basement membrane, aiding OSCC metastasis [98].

Furthermore, *Porphyromonas gingivalis* can help OSCC cells to bypass the immune system by activating PD-L1 to bind to its receptor, PD-1, mediating profound T cell inhibition [99], and can induce the expression of the B7-H1 and B7-DC receptors in OSCC cells that leads to the apoptosis of activated T cells [100]. *Porphyromonas gingivalis* has been shown to induce epithelial to mesenchymal transition and can accelerate the cell cycle by affecting the p53, PI3K and cyclin pathways [101], primarily through its FimA adhesin molecule [102]. The bacterium also downgrades the activity of plakophilin, a key molecule in stabilising epithelial cells, and thus can promote metastatic change [93]. The bacterium is also one of the major producers of the carcinogen acetaldehyde and has a role to play in promoting autophagy of cancer cells, promoting chemotherapeutic resistance [103].

Thus, these two species, *Fusobacterium nucleatum* and *Porphyromonas gingivalis*, both have considerable virulence characteristics that allow them to be significantly involved in OSCC progression. They also act in a synergistic fashion, whereby

metabolic end products of *Fusobacterium nucleatum* help promote the growth and the pathogenic potential of *Porphyromonas gingivalis* [104,105].

Furthermore, an important concept in the relationship between the microbiome and OSCC is that of "bacterial trending", portrayed by Mukherjee et al. [2017]. Here, it is suggested that there are multiple different combinations of genera that exist in parallel, but may contribute to the same phenotypical outcome, that is, OSCC development. An example of these are the high *Rothia* and low *Fusobacterium* and high *Eikenella* and high *Fusobacterium* combinations [72]. This indicates that the oncobiomic communities of OSCC are quite dynamic and maybe even interchangeable, as well as being much more individualistic from person to person. Indeed, both intra- and interkingdom connectivity [bacterial–bacterial/ fungal–fungal and bacterial–fungal relationships] may be vital in the better understanding of OSCC pathogenesis. Such examples include the synergistic and mutualistic relationship between *Candida albicans* and *Streptococcus* species, the cross feeding between *Streptococcus gordonii* and *Fusobacterium nucleatum* [106] and the required presence of the latter bacteria in the survival of *Porphyromonas gingivalis*, as has been previously highlighted.

Protective Bacterial Strains against OSCC Development and Progression

Leptotrichia, *Neisseria*, *Streptococcus mitis*, and *Haemophilus parainfluenza* have been shown by several authors to be absent or less abundant in OSCC sites [71,79]. The presence of *Corynebacterium* and *Kingella* in one study was associated with a decreased risk of developing HNC; their presence was suggested to be cancer protective due to their ability to metabolise various toxic compounds by xenobiotic biodegradation. Other members of the oral microbiome that have been associated with a reduced risk of OSCC development include *Neisseria sicca*, *Parvimonas micra* and *Streptococcus anginosus* [78,107] Indeed, *Streptococcus anginosis* has been associated much more with oesophageal cancer rather than oral cancer [108].

Lactic Acid Bacteria, the Controversial Order

Lactic acid bacteria are a controversial issue in both pro- and anti-cancer progression arguments. These bacteria, commonly used in probiotic therapy, have shown success in the management of several disorders that include diarrhoea, allergy to foods, inflammatory diseases and, recently, cancers such as colorectal cancer [109]. Lactic acid bacteria have been shown to harbour anti-cancer growth properties, by promoting

apoptosis, exhibit anti-oxidative protection to cells, increase immune cell numbers and improve tumour suppression gene expressions [109]. Lactic acid bacteria can help eliminate the damaging effects of substances in the body, as well as being capable of the promotion and regulation of both innate and adaptive immune responses in the prevention of tumorigenesis. Such examples include *Lactobacillus rhamnosus* GG, a strain thought to have the ability to lower the chronic inflammation that is linked with cancer development [110].

In the oral cavity, *Lactobacillus* species can inhibit the binding of other bacteria to epithelial cells or even arrest both bacterial and fungal growth [36]. They also have a role to play in improving the efficacy of drugs. The bacterial strain *Lactobacillus fermentum* has been shown to exhibit anti-inflammatory effects against oral cancer and may also work to modify or improve specific treatments against OSCC such as the TLR-9 agonists CpG oligodeoxynucleotides [111]. The addition of *Lactobacillus fermentum*-CQPC08 in a mouse with cancer of the tongue was found to increase serum G-CSF, GM -CSF, IgG and IgM, IL-4, IL-12, TNF-alpha, and IFN-gamma, as well as improving local tissue antioxidants and decreasing malondialdehyde—a damaging lipid peroxidation product—while better macrophage phagocytic ability was also observed [112]. In another study, *Lactobacillus plantarum* was found to induce apoptosis of oral cancer cells by the upregulation of PTEN and MAPK signalling and was thus proposed as a potential probiotic adjuvant in OSCC treatment [113].

On the other hand, lactic acid production by *Lactobacillus, Lactococcus, Leuconostoc, Pediococcus, Streptococcus* and *Peptostreptococcus* has been implicated in OSCC development [51]. Other acids produced by these bacteria include acetic, butyric, isobutyric, isovaleric, and isocaproic acids, which all reduce the environmental pH of an OSCC, contributing to cancer growth and spread. Furthermore, lactic acid itself can enhance gene transcription and DNA repair by the blocking of histone deacetylases [36]. *Lactobacillus*, in one study, was found to be increasingly abundant, relative to the increasing TNM staging of cancer patients [33,114]. Table 1 summarises both the possible protective and harmful roles associated with protumoral and protective bacterial strains in OSCC pathogenesis and progression.

Table 1: Summary of possible associations of harmful and protective bacterial genera and strains and their association with OSCC development and progression

Type of bacteria	Genera/Strains	Possible associated roles in OSCC development and/or progression
Harmful associated [pro-tumoural strains] bacteria	*Actinobacillus*	Can upregulate the production of CCL20 in oral cancer cell lines [139]
	Aggregatibacter	Expression of proinflammatory cytokines [140]
	Capnocytophaga	May stimulate chronic inflammatory processes
	Citrobacter	In murine models, strains of *Citrobacter* induce mucosal hyperplasia and inflammation [141]

Clostridia	Clostridial spores have been shown to have an affinity for hypoxic regions of solid tumours and may be factors in OSCC development Clostridial strains have however been used in cancer therapy [142]
Cutibacterium [*Propionibacterium*]	Induction of pro-inflammatory cytokines [IL-6 and IL-8]. Strains have been associated with prostate cancer [143]
Enterococcus	Extracellular superoxide may produce genomic instability [144] *Enterococcus* strains associated with chronic inflammation and development of other cancers [145]
Eubacterium saburreum	Periodate-resistant antigen [PS L13] [146] may elicit immune reactions
Exiguobacterium oxidotolerans	High catalase activity may protect cancer cells [147]

Fusobacterium	Stimulation of pro-inflammatory cytokines [IL-6, IL-8] FadA adhesin of *Fusobacterium nucleatum* can increase transcription of Wnt, causing increased cell proliferation Fap 2 reduces immune cell activity Increases matrix metalloproteinase activity, leading to degradation of basement membrane
Gemella	Associated with upregulation of interleukins [IL-23A] in cancer [148]
Johnsonella ignava	May stimulate chronic inflammatory processes
Leptotrichia buccalis	Contains a potent phenol-soluble lipopolysaccharide endotoxin that can stimulate inflammation [149]

Oribacterium	Production of acetate as a metabolic by-product, a substrate quickly utilised by hypoxic growing OSCC tumours
Parvimonas	Associated with development of colorectal cancers [150]
Peptostreptococcus stomatis	Has been associated with invading OSCC front
Porphyromonas gingivalis	Activation of complement system Pro-inflammatory cytokine release Prevents cell apoptosis [inhibition of cytochrome C, reduced caspase 3 activity and secretion of nucleoside diphosphate kinase] Further prevents apoptosis by upregulating anti-apoptotic genes BCL2 and surviving Upregulation of colon cancer-associated transcript 1 Increased enzyme production—nicotinamide N methyltransferase Gingipains activate matrix metalloproteinases Bacteria is a potent PD-1/PD- L1 activator, mediating immune

	bypass Induction of epithelial to mesenchymal transition FIMA adhesin molecule of bacteria affects cell cycle and alters tumour suppressor gene expression Acetaldehyde producer
Prevotella	Increased Activation of TLR 2 and promotion of T helper type 17 cells and associated mucosal inflammation Potent activators of pro-inflammatory mediators IL-6 and IL-8 as well as TNF-gamma [151]
Pseudomonas aeruginosa	Chronic inflammatory stimulation DNA cell damage May promote metastasis
Rothia	Potent acetaldehyde producers

		Staphylococcal α-toxin can activate pro-inflammatory cytokines and activate NK-kβ
	Staphylococcus aureus	This bacterium has been found to be in increased abundance in squamous cell carcinoma of the skin [associated with actinic keratosis] [152]
	Streptococcus gordonii, *parasanguinis* and *salivarius*	Streptococcal strains can produce acetaldehyde and show alcohol dehydrogenase enzyme activity [153]
	Tannerella [*forsythia*]	Induces pro-inflammatory cytokine production Cysteine-like proteases may arrest cell cycle in G2 phase [154]
Type of bacteria	**Genera/Strains**	**Role in OSCC prevention and/or cessation**

Protective bacteria	*Streptococcus mitis*	Their presence is very important in the prevention of colonisation of more virulent micro-organisms [155]
	Streptococcus gordonii	May have protective role, such that epithelial cells do not respond to *Porphyromonas gingivalis*-induced epithelial cell proliferation stimulation [156]
	Streptomyces	Some species have been shown to inhibit cell growth [157], including apoptosis of cancer cell lines [158]
	Neisseria	Can break down toxins of tobacco
	Corynebacterium and *Kingella*	Can metabolise various toxic compounds by xenobiotic biodegradation

	Veillonella	Naturally inhabits the dorsum of the tongue and thus may have more protective properties in this region [159]

Lactic acid bacteria have both protumoral and protective. Protumoral factors include lactic acid production that reduced the pH of the cancer environment, potentiating spread. Enhanced gene transcription alters DNA repair by the blocking of histone deacetylases. The protective factors of these bacteria include the binding and degrading of carcinogens, the production of anti-carcinogenic compounds [160], promotion of apoptosis [enhances MAPK activity], providing anti-oxidative [against ROS] protection to cells, increased immune cell numbers [T cells] and the induction of cytokines [INF-gamma, TNF-alpha], and improved tumour suppression gene expression [109].

Microbiome as a Potential Diagnostic Tool

The use of the microbiome as a diagnostic tool may be an important, non-invasive opportunity in the diagnosis of preclinical or early OSCC. Mager et al. [2005], described the species *Capnocytophaga gingivalis, Prevotella melaninogenica* and *Streptococcus mitis* as significantly elevated in the saliva of patients with OSCC with a diagnostic sensitivity and specificity of 80% and 82%, respectively [76]. In a study by Yang et al. [2018], an increased abundance of *Fusobacteria* was found in association with cancer staging, whereby those with Stage 1 and 4 OSCC had 4.35% and 7.92% abundances of this phylum, respectively. Of note in this study was the further decreased abundance of *Fusobacteria* amongst non-OSCC patients [2.98%] [30]. Thus, whether analysis of *Fusobacteria* levels in patients who are at risk of developing OSCC is a useful futuristic and non-invasive diagnostic tool may be of questionable interest. Indeed, the OSCC-associated microbiome may even be accurate enough to differentiate between OSCC and neighbouring cancers. The genera of *Actinomyces, Parvimonas, Selenomonas* and *Prevotella* were found to be in higher abundance in OSCC patients compared to oropharyngeal cancer patients [73].

Effects of Smoking, Smokeless Tobacco and Alcohol on the Oral Microbiome

The oral microbiome is affected by habits such as smoking, smokeless tobacco use and alcohol intake that can modify the oral mucosal ecological niche, allowing for modifications that may predispose a person to OSCC. Smoking is thought to shift the oral microbiome to a more "pathogenic rich" environment with enhanced pro-inflammatory virulence promotion [115]. Furthermore, commensal bacteria such as *Neisseria,* in particular, seem to be lost amongst smokers [114,116], while an abundance of *Fusobacteria, Mogibacterium* and *Tannerella* occurs [117]. Whether *Neisseria* has a protective effect on the oral mucosa requires further investigation; however, the loss of the proteobacteria *Neisseria* and *Haemophilus* is also a pattern amongst gastric cancer patients [118]. Nicotine has been shown to promote bacterial adherence to the oral mucosa [119], while the metabolite of nicotine, cotinine, can significantly increase epithelial cell invasion by *Porphyromonas gingivalis* [120]. Smoking has also been shown to reduce the host response to *Porphyromonas gingivalis* [121] and it is thought that the local immunosuppression caused by smoking can be one cause of a more novel and pathogenic microbiome picture [122].

Furthermore, bacteria play an important role in the increased activation of nitrosamines from different tobacco forms, including smokeless tobacco [123]. Indeed, smokeless tobacco products, in a study by Liu et al. [2016], were shown to promote OSCC-associated genera such as *Eubacterium, Streptococcus* and *Peptostreptococcus,* while tobacco-specific nitrosamines reduced commensal species such as *Veillonella*. Interestingly, in the same study, smokeless tobacco was found to limit the growth of *Fusobacterium nucleatum*, as well as *Eikenella corrodens* [124].

Meanwhile, the effects of alcohol consumption [> 1 unit per week], are also quite significant on the microbiome, with the enrichment of OSCC-associated bacteria such as *Campylobacter* species [64]. In another study, those whose alcohol intake was > 40 g/day, were shown to have significantly increased salivary acetaldehyde production while in combination with smoking, an alcohol consumption<30 g/day [females] and <40 g/day for males, increased salivary acetaldehyde by 100% [125]. Interestingly, while *Neisseria* are generally regarded as commensal organisms and are associated with an anti-OSCC picture, the genus is still involved in the significant production of acetaldehyde from ethanol. *Neisseria* genera were found to be increased amongst alcohol drinkers, suggesting that, among alcohol users, they can turn into commensal pathogens. Furthermore, both oral and gut *Lactobacillus* has been found to be decreased amongst heavy alcohol drinkers [>1–2 drinks/day] and amongst those who both smoke and drink. *Lactobacillus* can break down acetaldehyde and thus alcohol is directly implicated in the disruption of the microbiome, potentiating acetaldehyde-associated carcinogenesis [123,126].

Oral Mycoplasma

In a study by Henrich et al. [2014], It was found that *Mycoplasma salivarium* colonisation was significantly related to the surface of tongue OSCC in a patient with immunosuppressive disease [127]. *Mycoplasma* have been associated with the inhibition of the P53 tumour suppressor gene and the activation of the NF-kβ pathway [128] in rodents. Interestingly, *Mycoplasma salivarium* has been detected by immunohistochemistry from epithelial cells of leukoplakic lesions [129] undergoing hyperplastic changes. Whether oral *mycoplasma* play a role in acting as immunomodulators in targeted persons, particularly those with immunosuppressive states, aiding OSCC progression, is yet to be investigated further.

Oral Archaea

Both methanogenic and non-methanogenic archaic forms have been isolated from the oral cavity. Archaea have been isolated from disease locations such as chronic and aggressive periodontal disease, apical periodontitis, recalcitrant chronic rhinosinusitis and peri-implantitis [130,131]. They can illicit both an IgG humoral immune response [132–135] as well as proinflammatory cytokine reactions that include TNF alpha and IL-1 [20], and thus may be a hidden cause of chronic inflammation in the oral cavity. Importantly, through interspecies hydrogen transfer, a harmonious syntrophic environment develops between archaea and many bacteria, including pathogenic periodontal forms and fermenting bacteria, such that the presence of oral archaea can indirectly support the growth of bacteria that may be involved in the pathogenesis of OSCC [136]. Furthermore, methanogenic archaea are also known to be capable of methylating heavy metals to volatile forms that are toxic to human cells [137], as well as the possibility that the carcinogen formaldehyde can be produced from methane. Increases in methanogenic archaea have been associated with colorectal cancer [138]; however, no literature exists with OSCC; thus, this is an important area for future research.

Oral Mycobiome

The physical [adhesion and exclusion] and chemical [metabolic dependency and quorum-sensing] relationships between the mycobiome and bacteriome in the oral cavity are essential to maintain oral health. One of the limitations so far in microbiome research studies associated with OSCC is the continuous exclusion of fungal dynamics [31].

Similar to the bacterial microbiome, the richness and diversity of fungal species have been found to be significantly reduced amongst OSCC [161]. The absence of certain members of the oral mycobiome has been associated with HNC. *Malassezia* has been found to be increased in healthy individuals compared to cancer patients [162], with its presence found to be negatively correlated with Candidal presence [163], while, in a study by Shay et al. [2020], *Schizophyllum* was enriched in the samples who did not have HNC. This genus is thought to produce schizophylan, a polysaccharide compound with anti-cancer properties; thus, its absence in HNC patients may reflect the loss of its protective role [164]. In another study, the presence of *Emericella* was shown to be decreased in tongue cancer tissue compared to normal tissue. In colon

cancer cells, *Emericella* is thought to increase p53 tumour suppressor expression and thus its absence in OSCC lesions of the tongue may have a role in allowing cancer development [72].

On the other hand, the fungal mycobiome has been implicated in both premalignant disorders and OSCC. *Candida albicans, dubliniensis, tropicalis, pintolopesii,* and *glabrata,* as well as *Saccharomyces cerevisiae,* have been isolated from chronic hyperplastic candidiasis, a premalignant lesion with a higher rate of severe dysplastic transformation [165]. Thus, increased oral Candidal presence has been associated with both oral epithelial dysplasia and its severity [166]. In a study by Perera et al. [2017], the genera *Candida, Hannaella,* and *Gibberella* were increased amongst OSCC lesions and *Candida albicans*, *Candida etchellsii,* and a fungus resembling *Hannaella luteola* was the enriched species in OSCC [161]. *Candida albicans* is also thought to impart a protective role at tumour sites for bacterial species such as *Porphyromonas gingivalis*; the implications of this bacteria in OSCC development are described earlier in this review [167]. Indeed, it is *Candida albicans* that has been the most extensively studied with regard to its role in OSCC development and progression. This species can colonise the epithelium by secreting aspartic proteases as well as stimulating epithelial cell endocytosis, which ultimately leads to keratinocyte cell damage. *Candida* species have been shown to produce N-nitrosobenzylmethylamine and acetaldehyde, both of which are potent carcinogens [165]. Furthermore, Candidal proteinases have been implicated in the degradation of the basement membrane protein laminin-332. Interestingly, this has been found only to occur at acidic pH [168] and thus the acidic nature of a tumour, promoted by certain bacteria, may play a synergistic role with *Candida* species in enhancing this effect.

Viruses

Human Papilloma Virus, Epstein–Barr Virus and Herpes Simplex

While all these viruses are known to cause various different forms of oral disease, they are generally not considered to play a role in OSCC development. The human papilloma virus is not considered to play a role in OSCC pathogenesis, as it is with oropharyngeal cancer, and is found in very low frequency amongst OSCC patients [169,170]; however, the Epstein–Barr virus is controversial. Although present in OSCC cells, this is thought to be due to normal clonal presence and continuous shedding [171,172]; however, a recent meta-analysis concluded that Epstein–Barr

virus infection is in fact statistically associated with an increased risk of OSCC and that DNA regions such as Bam H1W, which is an oncogene, were detected in OSCC tissue [173]. In a large study of 155 OSCC samples from eight different countries and ethnicities, 55% of samples were positive for Epstein–Barr virus, while only 35% were positive for the human papilloma virus, with double infection occurring in 21% of samples [174]. However, there is still no clear evidence for a reproducible aetiopathogenetic link between Epstein–Barr virus and OSCC development.

Herpes simplex

Herpes simplex 1 is another virus associated with oral infection, but with no known association with OSCC development. However, antibody levels to herpes simplex viruses have been found to be higher amongst OSCC patients when compared with controls [175]. Interestingly, it has been shown that lactic acid bacteria and their bacteriocins can act as antiviral agents, particularly to herpes simplex virus. *Lactobacillus* species can prevent viral replication in host cells [176] and can suppress herpes simplex virus type 2 by lactic acid production in the vagina [177] and, as such, this may also be the case orally for herpes simplex type 1. Due to its toxicity to cancer cells, some authors have utilised attenuated forms of the herpes simplex virus in the treatment of OSCC, which was shown in mouse models to inhibit the growth of oral cancer cells [178]. However, there are several limitations to this form of cancer therapy—primarily the elimination of the virus from the cancer cells and the inhibition of the virus by the immune system [179].

The Prodigy of HIV and All Players of the Microbiome

While some authors advocate a similar oral bacterial microbiome between human immunodeficiency virus [HIV] patients and healthy individuals [180,181], others have found considerably reduced microbial diversity amongst HIV-positive persons, with increased abundance of the pathogenic genera *Megasphaera*, *Campylobacter*, *Veillonella* and *Prevotella*, and decreased abundance of commensal forms such as *Lactobacillus* and *Streptococcus* species [46,177]. Indeed, the gut microbiome in HIV infection resembles many other proinflammatory conditions such as IBD of the gut [182].

A pathogenic microbial shift is associated with high plasma HIV RNA levels [*Veillonella* and *Prevotella*], while a commensal shift [*Neisseria* and *Haemophilus*

species] is associated with antiretroviral therapy [46]. However, treatment of HIV still does not restore the microbiome to better health and increased abundance of the pathogenic *Campylobacter, Capnocytophaga, Fusobacterium, Prevotella,* and *Capnocytophaga* genera are still observed [183]. Specific species associated with the OSCC microbiome and its development have been particularly isolated from HIV individuals and include *Capnocytophaga* [180], *Prevotella melaninogenica* and *Rothia mucilaginosa* [46]. The latter species is very much associated with leukoplakic lesions and the carcinogenic release of acetaldehyde [184] and thus it may have a role to play in the high risk associated with developing oral premalignant lesions and OSCC amongst HIV patients [185]. Interestingly, *Porphyromonas gingivalis* may be able to induce HIV reactivation through butyric acid production [186], while also mediating the epithelial cell entry of the virus [187], but whether this bacterium and virus have a synergistic role to play in OSCC development in HIV positive patients is a source for further investigation.

The mycobiome of those with HIV varied considerably compared to controls in a study by Mukherjee et al. [2014] [180]. Interestingly, in this study, the fungus *Pichia* was found only in non-HIV persons and was found to have considerable inhibitory properties in the growth of *Candida, Aspergillus* and *Fusarium* [180]. This would suggest that the oral candidiasis that often presents itself in HIV-infected persons may be, in part, due to the absence of the antagonistic role that *Pichia* plays in healthy persons.

1.4 The Microbiome–Treatment Axis in OSCC

During OSCC treatment, both the oral and gut microbiome are affected by many factors that include host diet, side effects of surgery and radiation, antibiotic administration, as well as local implications such as the development of oral mucositis and dry mouth [188]. A new concept in cancer treatment is pharmacomicrobiomics, that is, how the gut microbiome interacts with drugs, affecting the success or failure of a treatment. Through the use of probiotics, prebiotics and antibiotics, subsequent interventions may be carried out to gear up the microbiome for a successful interaction with a wide range of cancer therapies. Early stage OSCC is predominantly treated via a surgery and radiation-based approach, while, where used, chemotherapy is often in

the form of the platinum-based drug cisplatin, which is used to radiosensitise oral cancer cells [189] or in more advanced case settings.

The oral microbiome has been shown to undergo significant changes during and after radiotherapy for OSCC [190]; however, the use of intensely modulated radiotherapy can help protect and stabilise the oral microbiome compared to conventional forms of radiation [191]. The side effects of radiation for OSCC include the development of mucositis, xerostomia, taste modification, dysphagia, osteoradionecrosis, and trismus [192]. In radiation therapy, higher abundances of the cariogenic bacteria *Streptococcus mutans* and *Lactobacillus*, as well as *Staphylococcus*, *Enterococcus*, and the fungus *Candida albicans*, parallel to a reduced presence of the commensal microorganism *Neisseria* [193,194], have been described. Indeed, *Lactobacillus* species may be increased up to 6 months post-radiation, as xerostomia is thought to create a favourable growth environment [195].

However, HNC patients who undergo radiation have been shown to have the distinct trend of increased *Candida albicans* counts and *Enterobacteriaceae* as part of their oral microbiome [196]. In comparison, the xerostomic microbiome of those with Sjogren's syndrome was associated with only increased numbers of *Streptococcus mutans* [197]. Increased abundance of *Enterococcus* species in radiated patients may be associated with a high percentage of antibiotic resistance and may further contribute to infections of the oral cavity during OSCC treatment [198]. *Enterococcus* species can also enhance microbial proteolytic activity on fibronectin, further promoting mucosal damage and inflammation from radiation [104]. Radiation also has an inhibitory role in antimicrobial proteins that would normally keep *Candida* growth in check [162].

Soil genera such as *Derxia* and *Luteococcus*, the latter isolated from blood, have also been sequenced from patients with HNC undergoing radiation [199]. Interestingly, a significant decrease in abundance occurred post-radiation therapy for the OSCC-associated species *Prevotella melaninogenica* in those with HNC [190]. With relation to osteoradionecrosis, the most predominant species from affected mandibles were *Campylobacter gracilis*, *Fusobacterium nucleatum*, *Streptococcus intermedius*, *Peptostreptococcus*, and *Prevotella* species [200]. *Actinomyces israelii* has been

significantly associated with the necrotic bone of this condition and is implicated in the chronicity and purulent characteristics [201].

Cancer chemotherapy has a great impact on both the oral and gut microbiome. With the former, this is thought to shift to more acid-tolerant bacteria, but still leading to a relatively commensal [*Lactobacillus* and *Streptococcus*] picture. [202]. Napenas et al. [2012], found a predominance of *Gemella haemolysans* and *Streptococcus mitis* species amongst patients undergoing chemotherapy; however, the most common genera identified from the oral microbiome during cancer chemotherapy include *Streptococcus, Staphylococcus, Klebsiella,* and *Enterobacter* [203].

The gut microbiome has a vital impact on the response to cancer chemotherapeutics and is involved by improving or compromising the efficacy of cancer drugs, as well as mediating their toxic effects. The microbiome has also recently been thought to play a role in chemotherapy-induced pain [204]. Locally, intra-tumour bacteria may also modulate chemotherapeutic effects [205].

The combination of *Lactobacillus* species [*Lactobacillus acidophilus*] and cisplatin was found to reduce tumour size and improve survival in a lung cancer mouse model [206]. The microbiome may also help upregulate proapoptotic pathways that improve the efficacy of cisplatin chemotherapy [207]. Other chemotherapeutic drugs used in advanced OSCC may also be affected by the microbiome. Gemcitabine is a synthetic pyrimidine antimetabolite that has efficacy in some cases of late stage OSCC [208]. *Gammaproteobacteria* have been shown to infer gemcitabine resistance, and thus co administration with ciprofloxacin can improve efficacy of this drug [205].

Oral mucositis has been associated with a marked loss of commensal bacteria such as *Actinomyces, Streptococcus, Granulicatella,* and *Veillonella* and an increased abundance of pathogenic bacteria such as *Enterobacteriaceae, Pseudomonas* and *Staphylococcus* species as well as *Escherichia coli, Fusobacterium nucleatum,* [in particular, the presence of *Fusobacterium nucleatum* subspecies *vincentii* in more severe mucositis], *Clostridiales, Treponema denticola* and *maltophilum, Porphyromonas gingivalis, Parvimonas micra* and *Prevotella oris* [207,209,210].

Porphyromonas gingivalis, *Candida glabrata,* and *Candida kefyr*, in a study by Laheij et al. [2012], were significantly associated with ulcerative oral mucositis [209]. Indeed, species diversity is found to differ between those who develop mild and severe

forms of mucositis [211]. Furthermore, shifts in the oral microbiome during chemotherapy or radiotherapy may allow for the dominance of mucolytic bacteria, particularly *Streptococcus* species that can degrade the protective mucous layer, further predisposing patients to oral mucositis development. The selective inhibition of Gram-negative bacteria, in a placebo-controlled double-blind randomised study by Wijers et al. [2001], was found not to improve radiation-induced mucositis grade [212,213].

Patients who develop the clinical disease of oral candidiasis during radiotherapy or chemotherapy have been shown to harbour increased Candidal load, comprising of *Candida albicans* and *Candida dubliniensis* predominantly, in combination with acid-tolerant bacteria that include *Lactobacillus salivarius*, *oris* and *crispatus* and *Streptococcus* species such as *parasanguinis* [162]. A synergistic relationship between *Candida albicans* and *several Streptococcus* species has been suggested [214]. *Streptococcus gordonii* interacts with *Candida albicans*, for example, through its adhesin receptor, improving hyphal development [215] and thus promoting *candida* tissue invasion. *Candida albicans* remains the most pathogenic species associated with radiotherapy [216].

Finally, in HNC, tumours are often regarded as "cold" or with reduced T cell infiltration, in part due to immune checkpoints, providing cancer protection from the immune system. Pembrolizumab and nivolumab are the commonest anti-PD-1 checkpoint inhibitors and are approved by the FDA for the treatment of patients with recurrent or metastatic head and neck squamous cell carcinoma or those with cisplatin resistance. [217,218]. However, there have been patients who respond well to anti PD-1/L-1 therapy, while others seem to be resistant. Recent evidence now exists to suggest that bacteria may have a potent effect on the successful ability of immune checkpoint inhibitors, which also include the CTLA-4 immunotherapy-type drugs. The species *Akkermansia muciniphila* is thought to promote better success with anti-PD-1 therapy by causing the infiltration of CCR9+ CXCR3+ CD4+ T lymphocytes to tumour sites as well as improving the immunosurveillance and preventing the immunosuppression of the host [219,220]. *Ruminococcaceae* were found to be abundant in patients with melanoma responsive to PD-1 therapy, and increased CD4+ and CD8+ T cells, while better cytokine profiles against the melanoma were associated with higher abundances of the *Clostridiales* order, the *Ruminococcaceae* family and the *Faecalibacterium*

genus in those patients [221]. In another study, a combination of *Bifidobacterium* and PD-L1 therapy was shown to have an almost complete abolishing effect on solid tumours due to enhanced dendritic cell activity and CD8+ T cell presence in the tumour field [222]. So important is the microbiome on the effect of anti-PD-1 therapy, that a recent clinical trial involved implementing faecal transplantation from responsive donors in anti-PD-1 therapy to recalcitrant cases and has shown promising results in increasing CD8 T cell infiltrations [223] in tumours in patients that were non-responsive to anti-PD1 therapy.

The effect of the CTLA-4 blockade is now shown to depend on distinct *Bacteroides* species, while *Firmicutes*, in particular *Faecalibacterium*, were associated with an improved clinical response to ipilimumab, by increasing CD8+ cell infiltration to tumours; however, at the same time, this concurred with the increased presentation of a more toxic side effect profile [224]. High levels of the short chain fatty acids butyrate and propionate have also been associated with resistance to the CTLA-4 blockade [225].

Conclusions

This review highlights the current science linking both the gut and oral microbiome to OSCC aetiopathogenesis and management, but also serves to shed a light on the limitations that exist in this field. Our understanding of the role of the gut and oral microbiome on OSCC development and behaviour is still evolving but has grown considerably in the last decade. However, it is still difficult to fully impart the true mechanisms by which the oral microbiome may directly lead to OSCC development [226]. Strong reproducible data have now emerged that fully link specific families, genera and species to OSCC aetiopathogenesis; however, there is still room for studies that are longitudinal and that employ increased participant numbers with similar methodologies and less variables. It is imperative that researchers begin to undertake such projects, examples of which have already been employed in other areas of cancer research such as The Cancer Genome Atlas project and the International Genomic Consortium, which provide robust current insights into molecular cancer associations [227]. Longitudinal studies are critical in assessing dynamic and more accurate shifts in the oral microbiome before, during and after OSCC development and can provide key insights into the association of the microbiome with OSCC.

1.5 References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **2018**, 68, 394–424.

2. Conway, D.I.; Purkayastha, M.; Chestnutt, I.G. The changing epidemiology of oral cancer: Definitions, trends, and risk factors. Br. Dent. J. **2018**, 225, 867–873.

3. Chaturvedi, A.K.; Anderson, W.F.; Lortet-Tieulent, J.; Curado, M.P.; Ferlay, J.; Franceschi, S.; Rosenberg, P.S.; Bray, F.; Gillison, M.L. Worldwide trends in incidence rates for oral cavity and oropharyngeal cancers. J. Clin. Oncol. **2013**, 31, 4550–4559.

4. Faggons, C.E.; Mabedi, C.; Shores, C.G.; Gopal, S. Review: Head and neck squamous cell carcinoma in sub-Saharan Africa. Malawi Med. J. **2015**, 27, 79–87.

5. Ali, H.; Sinnott, S.-J.; Corcoran, P.; Deady, S.; Sharp, L.; Kabir, Z. Oral cancer incidence and survival rates in the Republic of Ireland, 1994–2009. BMC Cancer **2016**, 16, 950.

6. Hussain, S.P.; Hofseth, L.J.; Harris, C.C. Radical causes of cancer. Nat. Rev. Cancer **2003**, 3, 276–285.

7. Tezal, M.; Sullivan, M.A.; Hyland, A.; Marshall, J.R.; Stoler, D.; Reid, M.E.; Loree, T.R.; Rigual, N.R.; Merzianu, M.; Hauck, L.; et al. Chronic periodontitis and the incidence of head and neck squamous cell carcinoma. Cancer Epidemiol. Biomark. Prev. **2009**, 18, 2406–2412.

8. Javed, F.; Warnakulasuriya, S. Is there a relationship between periodontal disease and oral cancer? A systematic review of currently available evidence. Crit. Rev. Oncol. Hematol **2016**, 97, 197–205.

9. Dewhirst, F.E.; Chen, T.; Izard, J.; Paster, B.J.; Tanner, A.C.R.; Yu, W.-H.; Lakshmanan, A.; Wade, W.G. The human oral microbiome. J. Bacteriol. **2010**, 192, 5002–5017.

10. Orlandi, E.; Iacovelli, N.A.; Tombolini, V.; Rancati, T.; Polimeni, A.; De Cecco, L.; Valdagni, R.; De Felice, F. Potential role of microbiome in oncogenesis, outcome prediction and therapeutic targeting for head and neck cancer. Oral Oncol. **2019**, 99, 104453.

11. Liang, D.; Leung, R.K.-K.; Guan, W.; Au, W.W. Involvement of gut microbiome in human health and disease: Brief overview, knowledge gaps and research opportunities. Gut Pathog. **2018**, 10, 3.

12. MacFie, J. Current status of bacterial translocation as a cause of surgical sepsis. Br. Med. Bull. **2004**, 71, 1–11.

13. Ivanov, I.I.; Honda, K. Intestinal commensal microbes as immune modulators. Cell Host Microbe **2012**, 12, 496–508.

14. Manzo, V.E.; Bhatt, A.S. The human microbiome in hematopoiesis and hematologic disorders. Blood **2015**, 126, 311–318.

15. Mohajeri, M.H.; Brummer, R.J.M.; Rastall, R.A.; Weersma, R.K.; Harmsen, H.J.M.; Faas, M.; Eggersdorfer, M. The role of the microbiome for human health: From basic science to clinical applications. Eur. J. Nutr. **2018**, 57, 1–14.

16. Vivarelli, S.; Salemi, R.; Candido, S.; Falzone, L.; Santagati, M.; Stefani, S.; Torino, F.; Banna, G.L.; Tonini, G.; Libra, M. Gut Microbiota and Cancer: From Pathogenesis to Therapy. Cancers **2019**, 11, 38. Int. J. Mol. Sci. **2020**, 21, 8061 19 of 29

17. Sultan, A.S.; Kong, E.F.; Rizk, A.M.; Jabra-Rizk, M.A. The oral microbiome: A Lesson in coexistence. PLoS Pathog. **2018**, 14, e1006719.

18. Bonner, M.; Fresno, M.; Gironès, N.; Guillén, N.; Santi-Rocca, J. Reassessing the Role of Entamoeba gingivalis in Periodontitis. Front. Cell. Infect. Microbiol. **2018**, 8.

19. Patil, S.; Rao, R.S.; Raj, A.T. Role of Mycoplasma in the Initiation and Progression of Oral Cancer. J. Int. Oral Health **2015**, 7, i–ii.

20. Nkamga, V.D.; Henrissat, B.; Drancourt, M. Archaea: Essential inhabitants of the human digestive microbiota.Hum. Microbiome J. **2017**, 3, 1–8.

21. Dame-Teixeira, N.; De Cena, J.A.; Côrtes, D.A.; Belmok, A.; Borges, L.G.D.A.; Marconatto, L.; Giongo, A.; Kyaw, C.M. Presence of Archaea in dental caries biofilms. Arch. Oral Biol. **2019**, 110, 104606.

22. Sampaio-Maia, B.; Monteiro-Silva, F. Acquisition and maturation of oral microbiome throughout childhood: An update. Dent. Res. J. **2014**, 11, 291–301.

23. Pokrowiecki, R.; Mielczarek, A.; Zar˛eba, T.; Tyski, S. Oral microbiome and peri-implant diseases: Where are we now? Clin. Risk Manag. **2017**, 13, 1529–1542.

24. Lucchese, A.; Bondemark, L.; Marcolina, M.; Manuelli, M. Changes in oral microbiota due to orthodontic appliances: A systematic review. J. Oral Microbiol. **2018**, 10, 1476645.

25. Shi, B.; Wu, T.; McLean, J.; Edlund, A.; Young, Y.; He, X.; Lv, H.; Zhou, X.; Shi, W.; Li, H.; et al. The Denture-Associated Oral Microbiome in Health and Stomatitis. mSphere **2016**, 1, e00215-16.

26. Hill, C.J.; Lynch, D.B.; Murphy, K.; Ulaszewska, M.; Je_ery, I.B.; O'Shea, C.A.; Watkins, C.; Dempsey, E.; Mattivi, F.; Tuohy, K.; et al. Evolution of gut microbiota composition from birth to 24 weeks in the Infantmet Cohort. Microbiome **2017**, 5, 4.

27. Zhuang, H.; Cheng, L.; Wang, Y.; Zhang, Y.-K.; Zhao, M.-F.; Liang, G.-D.; Zhang, M.-C.; Li, Y.-G.; Zhao, J.-B.; Gao, Y.-N.; et al. Dysbiosis of the Gut Microbiome in Lung Cancer. Front. Cell. Infect. Microbiol. **2019**, 9, 112.

28. Raskov, H.; Burcharth, J.; Pommergaard, H.-C. Linking Gut Microbiota to Colorectal Cancer. J. Cancer **2017**, 8, 3378–3395.

29. Mascitti, M.; Togni, L.; Troiano, G.; Caponio, V.C.A.; Gissi, D.B.; Montebugnoli, L.; Procaccini, M.; Muzio, L.L.; Santarelli, A. Beyond Head and Neck Cancer: The Relationship Between Oral Microbiota and Tumour Development in Distant Organs. Front. Cell. Infect. Microbiol. **2019**, 9.

30. Yang, C.-Y.; Yeh, Y.-M.; Yu, H.-Y.; Chin, C.-Y.; Hsu, C.-W.; Liu, H.; Huang, P.-J.; Hu, S.-N.; Liao, C.-T.; Chang, K.-P.; et al. Oral Microbiota Community Dynamics Associated With Oral Squamous Cell Carcinoma Staging. Front. Microbiol. **2018**, 9, 862.

31. Krom, B.; Kidwai, S.; Cate, J.T. Candida and Other Fungal Species. J. Dent. Res. **2014**, 93, 445–451.

32. Gazdeck, R.K.; Fruscione, S.R.; Adami, G.R.; Zhou, Y.; Cooper, L.F.; Schwartz, J.L.; Gazdeck, K.R. Diversity of the oral microbiome between dentate and edentulous individuals. Oral Dis. **2019**, 25, 911–918.

33. Guerrero-Preston, R.; Godoy-Vitorino, F.; Jedlicka, A.; Rodríguez-Hilario, A.; González, H.; Bondy, J.; Lawson, F.; Folawiyo, O.; Michailidi, C.; Dziedzic, A.; et al. 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. Oncotarget **2016**, 7, 51320–51334.

34. Virchow, R. An Address on the Value of Pathological Experiments. BMJ **1881**, 2, 198–203.

35. Pang, X.; Tang, Y.-J.; Ren, X.-H.; Chen, Q.-M.; Tang, Y.-L.; Liang, X.-H. Microbiota, Epithelium, Inflammation, and TGF-_ Signaling: An Intricate Interaction in Oncogenesis. Front. Microbiol. **2018**, 9, 1353.

36. Guerrero-Preston, R.; White, J.R.; Godoy-Vitorino, F.; Rodríguez-Hilario, A.; Navarro, K.; González, H.; Michailidi, C.; Jedlicka, A.; Canapp, S.; Bondy, J.; et al. High-resolution microbiome profiling uncovers Fusobacterium nucleatum, Lactobacillus gasseri/johnsonii, and Lactobacillus vaginalis associated to oral and oropharyngeal cancer in saliva from HPV positive and HPV negative patients treated with surgery and chemo-radiation. Oncotarget **2017**, 8, 110931–110948. Int. J. Mol. Sci. **2020**, 21, 8061 20 of 29

37. Karpinski, T.M. Role of Oral Microbiota in Cancer Development. Microorganisms **2019**, 7, 20.

38. Yost, S.; Stashenko, P.; Choi, Y.; Kukuruzinska, M.; Genco, C.A.; Salama, A.;Weinberg, E.O.; Kramer, C.D.; Frias-Lopez, J. Increased virulence of the oral microbiome in oral squamous cell carcinoma revealed by metatranscriptome analyses. Int. J. Oral Sci. **2018**, 10, 32.

39. Börnigen, D.; Ren, B.; Pickard, R.; Li, J.; Ozer, E.; Hartmann, E.M.; Xiao,W.; Tickle, T.; Rider, J.; Gevers, D.; et al. Alterations in oral bacterial communities are associated with risk factors for oral and oropharyngeal cancer. Sci. Rep. **2017**, 7, 17686.

40. Weinberg, E.D. The role of iron in cancer. Eur. J. Cancer Prev. **1996**, 5, 19–36.

41. Stashenko, P.; Yost, S.; Choi, Y.; Danciu, T.; Chen, T.; Yoganathan, S.; Kressirer, C.; Ruiz-Tourrella, M.; Das, B.; Kokaras, A.; et al. The Oral Mouse Microbiome Promotes Tumorigenesis in Oral Squamous Cell Carcinoma. mSystems **2019**, 4, e00323-19.

42. Sonalika,W.G.; Tayaar, S.A.; Bhat, K.G.; Patil, B.R.; Muddapur, M.V. Oral microbial carriage in oral squamous cell carcinoma patients at the time of diagnosis and during radiotherapy—A comparative study. Oral Oncol. **2012**, 48, 881–886.

43. Zhao, H.; Chu, M.; Huang, Z.; Yang, X.; Ran, S.; Hu, B.; Zhang, C.; Liang, J. Variations in oral microbiota associated with oral cancer. Sci. Rep. **2017**, 7, 11773.

44. Suzuki, N.; Yoneda, M.; Hirofuji, T. Mixed red-complex bacterial infection in periodontitis. Int. J. Dent. **2013**, 2013, 587279.

45. Cummins, J.; Tangney, M. Bacteria and tumours: Causative agents or opportunistic inhabitants? Infect. Agent. Cancer **2013**, 8, 11.

46. Lewy, T.; Hong, B.Y.; Weiser, B.; Burger, H.; Tremain, A.; Weinstock, G.; Anastos, K.; George, M.D. Oral Microbiome in HIV-Infected Women: Shifts in the Abundance of Pathogenic and Beneficial Bacteria Are Associated with Aging, HIV Load, CD4 Count, and Antiretroviral Therapy. Aids Res. Hum. Retrovir. **2019**, 35, 276–286.

47. Kawasaki, T.; Kawai, T. Toll-like receptor signaling pathways. Front. Immunol. **2014**, 5, 461.

48. Klimesova, K.; Zakostelska, Z.J.; Tlaskalova-Hogenova, H. Oral Bacterial and Fungal Microbiome Impacts Colorectal Carcinogenesis. Front. Microbiol. **2018**, 9.
49. Chocolatewala, N.; Chaturvedi, P.; Desale, R. The role of bacteria in oral cancer. Indian J. Med. Paediatr. Oncol. **2010**, 31, 126–131.

50. Voronov, E.; Shouval, D.S.; Krelin, Y.; Cagnano, E.; Benharroch, D.; Iwakura, Y.; Dinarello, C.A.; Apte, R.N. IL-1 is required for tumor invasiveness and angiogenesis. Proc. Natl. Acad. Sci. USA **2003**, 100, 2645–2650.

51. La Rosa, G.R.M.; Gattuso, G.; Pedullà, E.; Rapisarda, E.; Nicolosi, D.; Salmeri, M. Association of oral dysbiosis with oral cancer development. Oncol. Lett. **2020**, 19, 3045–3058.

52. Wu, Y.; Zhou, B.P. TNF-alpha/NF-kappaB/Snail pathway in cancer cell migration and invasion. Br. J. Cancer **2010**, 102, 639–644.

53. Tang, D.; Tao, D.; Fang, Y.; Deng, C.; Xu, Q.; Zhou, J. TNF-Alpha Promotes Invasion and Metastasis via NF-Kappa B Pathway in Oral Squamous Cell Carcinoma. Med. Sci. Monit. Basic Res. **2017**, 23, 141–149.

54. Szlosarek, P.; Charles, K.A.; Balkwill, F.R. Tumour necrosis factor-alpha as a tumour promoter. Eur. J. Cancer **2006**, 42, 745–750.

55. Choudhari, S.K.; Chaudhary, M.; Gadbail, A.R.; Sharma, A.; Tekade, S. Oxidative and antioxidative mechanisms in oral cancer and precancer: A review. Oral Oncol. **2014**, 50, 10–18.

56. Meurman, J.H.; Uittamo, J. Oral micro-organisms in the etiology of cancer. Acta Odontol. Scand. **2008**, 66, 321–326.

57. Blaser, M.J. Understanding microbe-induced cancers. Cancer Prev. Res. **2008**, 1, 15–20.

58. Wang, H.; Luo, Q.; Feng, X.; Zhang, R.; Li, J.; Chen, F. NLRP3 promotes tumor growth and metastasis in human oral squamous cell carcinoma. BMC Cancer **2018**, 18, 500.

59. Erttmann, S.F.; Gekara, N.O. Hydrogen peroxide release by bacteria suppresses inflammasome-dependent innate immunity. Nat. Commun. **2019**, 10, 3493.

60. Yaegaki, K. Oral malodorous compounds are periodontally pathogenic and carcinogenic. Jpn. Dent. Sci. Rev. **2008**, 44, 100–108. Int. J. Mol. Sci. **2020**, 21, 8061 21 of 29

61. Lee,W.H.; Chen, H.M.; Yang, S.F.; Liang, C.; Peng, C.Y.; Lin, F.M.; Tsai, L.L.;Wu, B.C.; Hsin, C.H.; Chuang, C.Y.; et al. Bacterial alterations in salivary microbiota and their association in oral cancer. Sci. Rep. **2017**, 7, 16540.

62. Halboub, E.; Al-Ak'hali, M.S.; Alamir, A.H.; Homeida, H.E.; Baraniya, D.; Chen, T.; Al-Hebshi, N.N. Tongue microbiome of smokeless tobacco users. BMC Microbiol. **2020**, 20, 201.

63. Amer, A.; Whelan, A.; Al-Hebshi, N.N.; Healy, C.M.; Moran, G.P. Acetaldehyde production by Rothia mucilaginosa isolates from patients with oral leukoplakia. J. Oral Microbiol. **2020**, 12, 1743066.

64. Amer, A.; Galvin, S.; Healy, C.M.; Moran, G.P. The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for Fusobacterium, Leptotrichia, Campylobacter, and Rothia Species. Front. Microbiol. **2017**, 8, 2391.

65. Choi, Y.S.; Kim, Y.; Yoon, H.J.; Baek, K.J.; Alam, J.; Park, H.K.; Choi, Y. The presence of bacteria within tissue provides insights into the pathogenesis of oral lichen planus. Sci. Rep. **2016**, 6, 29186.

66. Flemer, B.;Warren, R.D.; Barrett, M.P.; Cisek, K.; Das, A.; Je_ery, I.B.; Hurley, E.; O'Riordain, M.; Shanahan, F.; O'Toole, P.W. The oral microbiota in colorectal cancer is distinctive and predictive. Gut **2017**, 67, 1454–1463.

67. Nagy, K.N.; Sonkodi, I.; Szöke, I.; Nagy, E.; Newman, H.N. The microflora associated with human oral carcinomas. Oral Oncol. **1998**, 34, 304–308.

68. Hooper, S.J.; Crean, S.J.; Lewis, M.A.O.; Spratt, D.A.; Wade, W.G.; Wilson, M.J. Viable bacteria present within oral squamous cell carcinoma tissue. J. Clin. Microbiol. **2006**, 44, 1719–1725.

69. Katz, J.; Onate, M.D.; Pauley, K.M.; Bhattacharyya, I.; Cha, S. Presence of Porphyromonas gingivalis in gingival squamous cell carcinoma. Int. J. Oral Sci. **2011**, 3, 209–215.

70. Pushalkar, S.; Ji, X.; Li, Y.; Estilo, C.; Yegnanarayana, R.; Singh, B.; Li, X.; Saxena, D. Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. BMC Microbiol.**2012**, 12, 144.

71. Schmidt, B.L.; Kuczynski, J.; Bhattacharya, A.; Huey, B.; Corby, P.M.; Queiroz, E.L.S.; Nightingale, K.; Kerr, A.R.; DeLacure, M.D.; Veeramachaneni, R.; et al.

Changes in abundance of oral microbiota associated with oral cancer. PLoS ONE **2014**, 9, e98741.

72. Mukherjee, P.K.;Wang, H.; Retuerto, M.; Zhang, H.; Burkey, B.; Ghannoum, M.A.; Eng, C. Bacteriome and mycobiome associations in oral tongue cancer. Oncotarget **2017**, 8, 97273–97289.

73. Lim, Y.; Fukuma, N.; Totsika, M.; Kenny, L.; Morrison, M.; Punyadeera, C. The Performance of an Oral Microbiome Biomarker Panel in Predicting Oral Cavity and Oropharyngeal Cancers. Front. Cell. Infect. Microbiol. **2018**, 8, 267.

74. Sakamoto, H.; Naito, H.; Ohta, Y.; Tanakna, R.; Maeda, N.; Sasaki, J.; Nord, C.E. Isolation of bacteria from cervical lymph nodes in patients with oral cancer. Arch. Oral Biol. **1999**, 44, 789–793.

75. Byakodi, R.; Krishnappa, R.; Keluskar, V.; Bagewadi, A.; Shetti, A. The microbial flora associated with oral carcinomas. Quintessence Int. **2011**, 42, e118–e123.

76. Mager, D.L.; Ha_ajee, A.D.; Devlin, P.M.; Norris, C.M.; Posner, M.R.; Goodson, J.M. The salivary microbiota as a diagnostic indicator of oral cancer: A descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. J. Transl. Med. **2005**, 3, 27.

77. Al-Hebshi, N.N.; Nasher, A.T.; Idris, A.M.; Chen, T. Robust species taxonomy assignment algorithm for 16S rRNA NGS reads: Application to oral carcinoma samples. J. Oral Microbiol. **2015**, 7, 28934.

78. Hayes, R.B.; Ahn, J.; Fan, X.; Peters, B.A.; Ma, Y.; Yang, L.; Agalliu, I.; Burk, R.D.; Ganly, I.; Purdue, M.P.; et al. Association of Oral MicrobiomeWith Risk for Incident Head and Neck Squamous Cell Cancer. Jama Oncol.**2018**, 4, 358–365.

79. Al-Hebshi, N.N.; Nasher, A.T.; Maryoud, M.Y.; Homeida, H.E.; Chen, T.; Idris, A.M.; Johnson, N.W. Inflammatory bacteriome featuring Fusobacterium nucleatum

and Pseudomonas aeruginosa identified in association with oral squamous cell carcinoma. Sci. Rep. **2017**, 7, 1834.

80. Lima, C.; Calegari-Silva, T.; Pereira, R.M.; Santos, S.; Lopes, U.; Plotkowski, M.-C.; Saliba, A. ExoU Activates NF-_B and Increases IL-8/KC Secretion during Pseudomonas aeruginosa Infection. PLoS ONE **2012**, 7, e41772. Int. J. Mol. Sci. **2020**, 21, 8061 22 of 29

81. Gallimidi, A.B.; Fischman, S.; Revach, B.; Bulvik, R.; Maliutina, A.; Rubinstein, A.M.; Nussbaum, G.; Elkin, M. Periodontal pathogens Porphyromonas gingivalis and Fusobacteriumnucleatum promote tumor progression in an oral-specific chemical carcinogenesis model. Oncotarget **2015**, 6, 22613–22623.

82. Liu, X.-b.; Gao, Z.-y.; Sun, C.-t.; Wen, H.; Gao, B.; Li, S.-b.; Tong, Q. The potential role of P.gingivalis in gastrointestinal cancer: A mini review. Infect. Agents Cancer **2019**, 14, 23.

83. Castellarin, M.;Warren, R.L.; Freeman, J.D.; Dreolini, L.; Krzywinski, M.; Strauss, J.; Barnes, R.;Watson, P.; Allen-Vercoe, E.; Moore, R.A.; et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. **2012**, 22, 299–306.

84. Han, Y.W. Fusobacterium nucleatum: A commensal-turned pathogen. Curr. Opin. Microbiol. **2015**, 23, 141–147.

85. Flanagan, L.; Schmid, J.; Ebert, M.; Soucek, P.; Kunicka, T.; Liska, V.; Bruha, J.; Neary, P.; Dezeeuw, N.; Tommasino, M.; et al. Fusobacterium nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. Eur. J. Clin. Microbiol. Infect. Dis. **2014**, 33, 1381–1390.

86. Mitsuhashi, K.; Nosho, K.; Sukawa, Y.; Matsunaga, Y.; Ito, M.; Kurihara, H.; Kanno, S.; Igarashi, H.; Naito, T.; Adachi, Y.; et al. Association of Fusobacterium species in pancreatic cancer tissues with molecular features prognosis. Oncotarget **2015**, 6, 7209–7220.

87. Kolenbrander, P.E.; Andersen, R.N.; Blehert, D.S.; Egland, P.G.; Foster, J.S.; Palmer, R.J., Jr. Communication among oral bacteria. Microbiol. Mol. Biol. Rev. **2002**, 66, 486–505.

88. Salem, A.; Almahmoudi, R.; Hagström, J.; Stark, H.; Nordström, D.; Salo, T.; Eklund, K.K. Human _-Defensin 2 Expression in Oral Epithelium: Potential Therapeutic Targets in Oral Lichen Planus. Int. J. Mol. Sci. **2019**, 20, 1780.

89. Rubinstein, M.R.;Wang, X.; Liu,W.; Hao, Y.; Cai, G.; Han, Y.W. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/_-catenin signaling via its FadA adhesin. Cell Host Microbe **2013**, 14, 195–206.

90. Brennan, K.; Koenig, J.L.; Gentles, A.J.; Sunwoo, J.B.; Gevaert, O. Identification of an atypical etiological head and neck squamous carcinoma subtype featuring the CpG island methylator phenotype. EBioMedicine **2017**,17, 223–236.

91. Ahn, J.; Segers, S.; Hayes, R.B. Periodontal disease, Porphyromonas gingivalis serum antibody levels andorodigestive cancer mortality. Carcinogenesis **2012**, 33, 1055–1058.

92. Michaud, D.S.; Izard, J.; Wilhelm-Benartzi, C.S.; You, D.H.; Grote, V.A.; Tjønneland, A.; Dahm, C.C.; Overvad, K.; Jenab, M.; Fedirko, V.; et al. Plasma antibodies to oral bacteria and risk of pancreatic cancer in a large European prospective cohort study. Gut **2013**, 62, 1764–1770.

93. Geng, F.; Liu, J.; Guo, Y.; Li, C.; Wang, H.; Wang, H.; Zhao, H.; Pan, Y. Persistent Exposure to Porphyromonas gingivalis Promotes Proliferative and Invasion Capabilities, and Tumorigenic Properties of Human Immortalized Oral Epithelial Cells. Front. Cell. Infect. Microbiol. **2017**, 7, 57.

94. Moffatt, C.E.; Lamont, R.J. Porphyromonas gingivalis induction of microRNA-203 expression controls suppressor of cytokine signaling 3 in gingival epithelial cells. Infect. Immun. **2011**, 79, 2632–2637.

95. Yilmaz, O.; Yao, L.; Maeda, K.; Rose, T.M.; Lewis, E.L.; Duman, M.; Lamont, R.J.; Ojcius, D.M. ATP scavenging by the intracellular pathogen Porphyromonas gingivalis inhibits P2X7-mediated host-cell apoptosis. Cell. Microbiol. **2008**, 10, 863–875.

96. Tribble, G.D.; Kerr, J.E.; Wang, B.Y. Genetic diversity in the oral pathogen Porphyromonas gingivalis: Molecular mechanisms and biological consequences. Future Microbiol. **2013**, 8, 607–620.

97. Michaud, D.S. Role of bacterial infections in pancreatic cancer. Carcinogenesis **2013**, 34, 2193–2197.

98. Inaba, H.; Sugita, H.; Kuboniwa, M.; Iwai, S.; Hamada, M.; Noda, T.; Morisaki, I.; Lamont, R.J.; Amano, A. Porphyromonas gingivalis promotes invasion of oral squamous cell carcinoma through induction of proMMP9 and its activation. Cell. Microbiol. **2014**, 16, 131–145.

99. Robayo, D.A.G.; Hernandez, R.F.; Erira, A.; Kandaurova, L.; Juarez, C.L.; Juárez, V.; Cid-Arregui, A. Oral Microbiota Associated with Oral and Gastroenteric Cancer. Open Microbiol. J. **2020**, 14, 1–17.

100. Groeger, S.; Domann, E.; Gonzales, J.R.; Chakraborty, T.; Meyle, J. B7-H1 and B7-DC receptors of oral squamous carcinoma cells are upregulated by Porphyromonas gingivalis. Immunobiology **2011**, 216, 1302–1310. Int. J. Mol. Sci. **2020**, 21, 8061 23 of 29

101. Kuboniwa, M.; Hasegawa, Y.; Mao, S.; Shizukuishi, S.; Amano, A.; Lamont, R.J.; Yilmaz, O.P. gingivalis accelerates gingival epithelial cell progression through the cell cycle. Microbes Infect. **2008**, 10, 122–128.

102. Whitmore, S.E.; Lamont, R.J. Oral bacteria and cancer. PLoS Pathog. **2014**, 10, e1003933.

103. Patil, S.; Rao, R.S.; Raj, A.T. Dual Role of Autophagy in Oral Cancer. J. Int. Oral Health **2015**, 7, i–ii.

104. Chattopadhyay, I.; Verma, M.; Panda, M. Role of Oral Microbiome Signatures in Diagnosis and Prognosis of Oral Cancer. Technol. Cancer Res. Treat. **2019**, 18.

105. Polak, D.; Wilensky, A.; Shapira, L.; Halabi, A.; Goldstein, D.; Weiss, E.I.; Houri-Haddad, Y. Mouse model of experimental periodontitis induced by Porphyromonas gingivalis/Fusobacterium nucleatum infection: Bone loss and host response. J. Clin. PeriOdontol. **2009**, 36, 406–410.

106. Brennan, C.A.; Garrett, W.S. Fusobacterium nucleatum-symbiont, opportunist and oncobacterium. Nat. Rev. Microbiol. **2019**, 17, 156–166. [CrossRef] [PubMed]
107. Perera, M.; Al-Hebshi, N.N.; Speicher, D.J.; Perera, I.; Johnson, N.W. Emerging role of bacteria in oral carcinogenesis: A review with special reference to perio-pathogenic bacteria. J. Oral Microbiol. **2016**, 8, 32762.

108. Morita, E.; Narikiyo, M.; Yano, A.; Nishimura, E.; Igaki, H.; Sasaki, H.; Terada, M.; Hanada, N.; Kawabe, R. Different frequencies of Streptococcus anginosus infection in oral cancer and esophageal cancer. Cancer Sci. **2003**, 94, 492–496.

109. Zhong, L.; Zhang, X.; Covasa, M. Emerging roles of lactic acid bacteria in protection against colorectal cancer. World J. Gastroenterol. Wjg **2014**, 20, 7878–7886.

110. Banna, G.L.; Torino, F.; Marletta, F.; Santagati, M.; Salemi, R.; Cannarozzo, E.; Falzone, L.; Ferraù, F.; Libra, M. Lactobacillus rhamnosus GG: An Overview to Explore the Rationale of Its Use in Cancer. Front. Pharmacol. **2017**, 8.

111. Iida, N.; Dzutsev, A.; Stewart, C.A.; Smith, L.; Bouladoux, N.;Weingarten, R.A.; Molina, D.A.; Salcedo, R.; Back, T.; Cramer, S.; et al. Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. Science **2013**, 342, 967–970.

112. Liu, B.; Zhang, J.; Yi, R.; Zhou, X.; Long, X.; Pan, Y.; Zhao, X. Preventive E_ect of Lactobacillus fermentum CQPC08 on 4-Nitroquineline-1-Oxide Induced Tongue Cancer in C57BL/6 Mice. Foods **2019**, 8, 93.

113. Asoudeh-Fard, A.; Barzegari, A.; Dehnad, A.; Bastani, S.; Golchin, A.; Omidi, Y. Lactobacillus plantarum induces apoptosis in oral cancer KB cells through upregulation of PTEN and downregulation of MAPK signalling pathways. Bioimpacts **2017**, 7, 193–198.

114. Wu, J.; Peters, B.A.; Dominianni, C.; Zhang, Y.; Pei, Z.; Yang, L.; Ma, Y.; Purdue, M.P.; Jacobs, E.J.; Gapstur, S.M.; et al. Cigarette smoking and the oral microbiome in a large study of American adults. ISME J. **2016**, 10, 2435–2446.

115. Shah, S.A.; Ganesan, S.M.; Varadharaj, S.; Dabdoub, S.M.;Walters, J.D.; Kumar, P.S. The making of a miscreant: Tobacco smoke and the creation of pathogen-rich biofilms. Npj. Biofilms Microbiomes **2017**, 3, 26.

116. Al-Zyoud,W.; Hajjo, R.; Abu-Siniyeh, A.; Hajjaj, S. Salivary Microbiome and Cigarette Smoking: A First of Its Kind Investigation in Jordan. Int. J. Environ. Res. Public Health **2019**, 17, 256.

117. Pimentel, S.P.; Fontes, M.; Ribeiro, F.V.; Corrêa, M.G.; Nishii, D.; Cirano, F.R.; Casati, M.Z.; Casarin, R.C.V. Smoking habit modulates peri-implant microbiome: A case-control study. J. Periodontal Res. **2018**, 53, 983–991.

118. Hu, J.; Han, S.; Chen, Y.; Ji, Z. Variations of Tongue Coating Microbiota in Patients with Gastric Cancer. Biomed. Res. Int. **2015**, 2015, 173729.

119. Piatti, G.; Gazzola, T.; Allegra, L. BACTERIAL ADHERENCE IN SMOKERS AND NON-SMOKERS. Pharmacol. Res. **1997**, 36, 481–484.

120. Cogo, K.; Calvi, B.M.; Mariano, F.S.; Franco, G.C.; Gonçalves, R.B.; Groppo, F.C. The e_ects of nicotine and cotinine on Porphyromonas gingivalis colonisation of epithelial cells. Arch. Oral Biol. **2009**, 54, 1061–1067.

121. Bagaitkar, J.; Demuth, D.R.; Daep, C.A.; Renaud, D.E.; Pierce, D.L.; Scott, D.A. Tobacco upregulates P. gingivalis fimbrial proteins which induce TLR2 hyposensitivity. PLoS ONE **2010**, 5, e9323. Int. J. Mol. Sci. **2020**, 21, 8061 24 of 29

122. Huang, C.; Shi, G. Smoking and microbiome in oral, airway, gut and some systemic diseases. J. Transl. Med. **2019**, 17, 225.

123. Ahn, J.; Chen, C.; Hayes, R. Oral microbiome and oral and gastrointestinal cancer risk. Cancer Causes Control. **2012**, 23, 399–404.

124. Liu, M.; Jin, J.; Pan, H.; Feng, J.; Cerniglia, C.E.; Yang, M.; Chen, H. E_ect of smokeless tobacco products on human oral bacteria growth and viability. Anaerobe **2016**, 42, 152–161.

125. Homann, N.; Tillonen, J.; Meurman, J.; Rintamäki, H.; Lindqvist, C.; Rautio, M.; Jousimies-Somer, H.; Salaspuro, M. Increased salivary acetaldehyde levels in heavy drinkers and smokers: A microbiological approach to oral cavity cancer. Carcinogenesis **2000**, 21, 663–668.

126. Fan, X.; Peters, B.A.; Jacobs, E.J.; Gapstur, S.M.; Purdue, M.P.; Freedman, N.D.; Alekseyenko, A.V.;Wu, J.; Yang, L.; Pei, Z.; et al. Drinking alcohol is associated with variation in the human oral microbiome in a large study of American adults. Microbiome **2018**, 6, 59.

127. Henrich, B.; Rumming, M.; Sczyrba, A.; Velleuer, E.; Dietrich, R.; Gerlach,W.; Gombert, M.; Rahn, S.; Stoye, J.; Borkhardt, A.; et al. Mycoplasma salivarium as a dominant coloniser of Fanconi anaemia associated oral carcinoma. PLoS ONE **2014**, 9, e92297.

128. Logunov, D.Y.; Scheblyakov, D.V.; Zubkova, O.V.; Shmarov, M.M.; Rakovskaya, I.V.; Gurova, K.V.;Tararova, N.D.; Burdelya, L.G.; Naroditsky, B.S.; Ginzburg, A.L.; et al. Mycoplasma infection suppresses p53, activates NF-kappaB and cooperates with oncogenic Ras in rodent fibroblast transformation. Oncogene **2008**, 27, 4521–4531.

129. Mizuki, H.; Kawamura, T.; Nagasawa, D. In situ immunohistochemical detection of intracellular Mycoplasma salivarium in the epithelial cells of oral leukoplakia. J. Oral Pathol. Med. **2015**, 44, 134–144.

130. Sogodogo, E.; Fellag, M.; Loukil, A.; Nkamga, V.D.; Michel, J.; Dessi, P.; Fournier, P.-E.; Drancourt, M. Nine Cases of Methanogenic Archaea in Refractory Sinusitis, an Emerging Clinical Entity. Front. Public Health **2019**, 7, 38.

131. Grine, G.; Terrer, E.; Boualam, M.A.; Aboudharam, G.; Chaudet, H.; Ruimy, R.; Drancourt, M. Tobaccosmoking- related prevalence of methanogens in the oral fluid microbiota. Sci. Rep. **2018**, 8, 9197.

132. Yamabe, K.; Maeda, H.; Kokeguchi, S.; Tanimoto, I.; Sonoi, N.; Asakawa, S.; Takashiba, S. Distribution of Archaea in Japanese patients with periodontitis and humoral immune response to the components. Fems Microbiol. Lett. **2008**, 287, 69–75.

133. Sogodogo, E.; Doumbo, O.; Aboudharam, G.; Kouriba, B.; Diawara, O.; Koita, H.; Togora, S.; Drancourt, M. First characterization of methanogens in oral cavity in Malian patients with oral cavity pathologies. BMC Oral Health **2019**, 19, 232.

134. Li, C.L.; Jiang, Y.T.; Liu, D.L.; Qian, J.; Liang, J.P.; Shu, R. Prevalence and quantification of the uncommon Archaea phylotype Thermoplasmata in chronic periodontitis. Arch. Oral Biol. **2014**, 59, 822–828.

135. Ferrari, A.; Brusa, T.; Rutili, A.; Canzi, E.; Biavati, B. Isolation and characterization of Methanobrevibacter oralis sp. nov. Curr. Microbiol. **1994**, 29, 7–12.

136. Macario, E.C.d.; Macario, A.J. Methanogenic archaea in health and disease: A novel paradigm of microbial pathogenesis. Int. J. Med. Microbiol. **2009**, 299, 99–108.

137. Horz, H.-P.; Conrads, G. Methanogenic Archaea and oral infections - ways to unravel the black box. J. Oral Microbiol. **2011**, 3.

138. Guindo, C.O.; Drancourt, M.; Grine, G. Digestive tract methanodrome: Physiological roles of human microbiota-associated methanogens. Microb. Pathog. **2020**, 149, 104425.

139. Kusama, K.; Inoue, H.; Miyazaki, Y.; Kikuchi, K.; Sakashita, H.; Ochiai, K. Microorganisms and cancer of the oral cavity. Integr. Cancer Sci. Ther. **2016**, 3, 510–515.

140. Benso, B. Virulence factors associated with Aggregatibacter actinomycetemcomitans and their role in promoting periodontal diseases. Virulence **2017**, 8, 111–114.

141. Umar, S. Citrobacter Infection and Wnt signaling. Curr. Colorectal Cancer Rep. **2012**, 8.

142. Minton, N.P. Clostridia in cancer therapy. Nat. Rev. Microbiol. **2003**, 1, 237–242.

143. Ugge, H.; Carlsson, J.; Söderquist, B.; Fall, K.; Andén, O.; Davidsson, S. The influence of prostatic Cutibacterium acnes infection on serum levels of IL6 and CXCL8 in prostate cancer patients. Infect. Agents Cancer **2018**, 13, 34. Int. J. Mol. Sci. **2020**, 21, 8061 25 of 29

144. Amarnani, R.; Rapose, A. Colon cancer and enterococcus bacteremia co-infection: A dangerous alliance. J. Infect. Public Health **2017**, 10, 681–684.

145. Maekawa, T.; Fukaya, R.; Takamatsu, S.; Itoyama, S.; Fukuoka, T.; Yamada, M.; Hata, T.; Nagaoka, S.; Kawamoto, K.; Eguchi, H.; et al. Possible involvement of Enterococcus infection in the pathogenesis of chronic pancreatitis and cancer. Biochem. Biophys. Res. Commun. **2018**, 506, 962–969.

146. Skaug, N.; Hofstad, T. Eubacterium saburreum L13 antigen containing a ketohexose as a main sugar constituent. Infect. Immun. **1983**, 39, 179–183.

147. Yumoto, I.; Hishinuma-Narisawa, M.; Hirota, K.; Shingyo, T.; Takebe, F.; Nodasaka, Y.; Matsuyama, H.; Hara, I. Exiguobacterium oxidotolerans sp. nov., a novel alkaliphile exhibiting high catalase activity. Int. J. Syst. Evol. Microbiol. **2004**, 54, 2013–2017.

148. Ge, W.; Hu, H.; Cai, W.; Xu, J.; Hu, W.; Weng, X.; Qin, X.; Huang, Y.; Han, W.; Hu, Y.; et al. High-risk Stage III colon cancer patients identified by a novel five-gene mutational signature are characterized by upregulation of IL-23A and gut bacterial translocation of the tumor microenvironment. Int. J. Cancer **2020**, 146, 2027–2035.

149. Weinberger, M.; Wu, T.; Rubin, M.; Gill, V.J.; Pizzo, P.A. Leptotrichia buccalis Bacteremia in Patients with Cancer: Report of Four Cases and Review. Rev. Infect. Dis. **1991**, 13, 201–206.

150. Yu, J.; Zhao, L.; Zhao, R.; Long, X.; Coker, O.O.; Sung, J.J.Y. The role of Parvimonas micra in intestinal tumorigenesis in germ-free and conventional APCmin/+ mice. J. Clin. Oncol. **2019**, 37, 531.

151. Larsen, J.M. The immune response to Prevotella bacteria in chronic inflammatory disease. Immunology **2017**, 151, 363–374.

152. Kullander, J.; Forslund, O.; Dillner, J. Staphylococcus aureus and squamous cell carcinoma of the skin. Cancer Epidemiol. Biomark. Prev. **2009**, 18, 472–478.

153. Kurkivuori, J.; Salaspuro,V.; Kaihovaara, P.; Kari, K.; Rautemaa, R.; Grönroos, L.; Meurman, J.H.; Salaspuro, M. Acetaldehyde production from ethanol by oral streptococci. Oral Oncol. **2007**, 43, 181–186.

154. Malinowski, B.; W˛esierska, A.; Zalewska, K.; Sokołowska, M.M.; Bursiewicz, W.; Socha, M.; Ozorowski, M.; Pawlak-Osi´ nska, K.; Wici ´ nski, M. The role of Tannerella forsythia and Porphyromonas gingivalis in pathogenesis of esophageal cancer. Infect. Agents Cancer **2019**, 14, 3.

155. Mitchell, J. Streptococcus mitis: Walking the line between commensalism and pathogenesis. Mol. Oral Microbiol. **2011**, 26, 89–98.

156. Ohshima, J.;Wang, Q.; Fitzsimonds, Z.R.; Miller, D.P.; Sztukowska, M.N.; Jung, Y.-J.; Hayashi, M.; Whiteley, M.; Lamont, R.J. Streptococcus gordonii programs epithelial cells to resist ZEB2 induction by Porphyromonas gingivalis. Proc. Natl. Acad. Sci. USA **2019**, 116, 8544–8553.

157. Maragheh, B.F.A.; Fatourachi, P.; Mohammadi, S.M.; Valipour, B.; Behtari, M.; Dehnad, A.;Nozad Charoudeh, H. Streptomyces Levis ABRIINW111 Inhibits SW480 Cells Growth by Apoptosis Induction. Adv. Pharm. Bull. **2018**, 8, 675–682.

158. Khan, N.; Yılmaz, S.; Aksoy, S.; Uzel, A.; Tosun, Ç.; Kirmizibayrak, P.B.; Bedir, E. Polyethers isolated from the marine actinobacterium Streptomyces cacaoi inhibit autophagy and induce apoptosis in cancer cells. Chem Biol. Interact. **2019**, 307, 167–178.

159. Liljemark, W.F.; Gibbons, R.J. Ability of Veillonella and Neisseria species to attach to oral surfaces and their proportions present indigenously. Infect. Immun. **1971**, 4, 264–268.

160. Rafter, J. Lactic acid bacteria and cancer: Mechanistic perspective. Br. J. Nutr. **2002**, 88, S89–S94.

161. Perera, M.; Al-Hebshi, N.N.; Perera, I.; Ipe, D.; Ulett, G.C.; Speicher, D.J.; Chen, T.; Johnson, N.W. A dysbiotic mycobiome dominated by Candida albicans is identified within oral squamous-cell carcinomas. J. Oral Microbiol. **2017**, 9, 1385369.

162. Diaz, P.I.; Hong, B.-Y.; Dupuy, A.K.; Choquette, L.; Thompson, A.; Salner, A.L.; Schauer, P.K.; Hegde, U.; Burleson, J.A.; Strausbaugh, L.D.; et al. Integrated Analysis of Clinical and Microbiome Risk Factors Associated with the Development of Oral Candidiasis during Cancer Chemotherapy. J. Fungi **2019**, 5, 49.

163. Robinson, S.; Peterson, C.; Sahasrabhojane, P.; Ajami, N.J.; Shelburne, S.; Kontoyiannis, D.; Galloway-Peña, J. Observational Cohort Study of Oral Mycobiome and InterKingdom Interactions over the Course of Induction Therapy for Leukemia. mSphere **2020**, 5. Int. J. Mol. Sci. **2020**, 21, 8061 26 of 29

164. Shay, E.; Sangwan, N.; Padmanabhan, R.; Lundy, S.; Burkey, B.; Eng, C. Bacteriome and mycobiome and bacteriome-mycobiome interactions in head and neck squamous cell carcinoma. Oncotarget **2020**, 11, 2375–2386.

165. Bakri, M.; Mohd Hussaini, H.; Rachel Holmes, A.; David Cannon, R.; Mary Rich, A. Revisiting the association between candidal infection and carcinoma, particularly oral squamous cell carcinoma. J. Oral Microbiol.**2010**, 2.

166. McCullough, M.; Jaber, M.; Barrett, A.W.; Bain, L.; Speight, P.M.; Porter, S.R. Oral yeast carriage correlates with presence of oral epithelial dysplasia. Oral Oncol. **2002**, 38, 391–393.

167. Bartnicka, D.; Gonzalez-Gonzalez, M.; Sykut, J.; Koziel, J.; Ciaston, I.; Adamowicz, K.; Bras, G.; Zawrotniak, M.; Karkowska-Kuleta, J.; Satala, D.; et al. Candida albicans Shields the Periodontal Killer Porphyromonas gingivalis from Recognition by the Host Immune System and Supports the Bacterial Infection of Gingival Tissue. Int. J. Mol. Sci. **2020**, 21, 1984.

168. Pärnänen, P.; Kari, K.; Virtanen, I.; Sorsa, T.; Meurman, J.H. Human laminin-332 degradation by Candida proteinases. J. Oral Pathol. Med. **2008**, 37, 329–335.

169. de Abreu, P.M.; Có, A.C.G.; Azevedo, P.L.; do Valle, I.B.; de Oliveira, K.G.; Gouvea, S.A.; Cordeiro-Silva, M.F.; Louro, I.D.; de Podestá, J.R.V.; Lenzi, J.; et al. Frequency of HPV in oral cavity squamous cell carcinoma. BMC Cancer **2018**, 18, 324.

170. Machado, J.; Reis, P.P.; Zhang, T.; Simpson, C.; Xu, W.; Perez-Ordonez, B.; Goldstein, D.P.; Brown, D.H.; Gilbert, R.W.; Gullane, P.J.; et al. Low prevalence of human papillomavirus in oral cavity carcinomas. Head Neck Oncol. **2010**, 2, 6.

171. Cruz, I.; Van den Brule, A.J.; Steenbergen, R.D.; Snijders, P.J.; Meijer, C.J.; Walboomers, J.M.; Snow, G.B.; Van derWaal, I. Prevalence of Epstein-Barr virus in oral squamous cell carcinomas, premalignant lesions and normal mucosa–a study using the polymerase chain reaction. Oral Oncol. **1997**, 33, 182–188.

172. Wilms, T.; Khan, G.; Coates, P.J.; Sgaramella, N.; Fåhraeus, R.; Hassani, A.; Philip, P.S.; Spaak, L.N.; Califano, L.; Colella, G.; et al. No evidence for the presence of Epstein-Barr virus in squamous cell carcinoma of the mobile tongue. PLoS ONE **2017**, 12, e0184201.

173. She, Y.; Nong, X.; Zhang, M.;Wang, M. Epstein-Barr virus infection and oral squamous cell carcinoma risk: A meta-analysis. PLoS ONE **2017**, 12, e0186860.

174. Jalouli, J.; Jalouli, M.M.; Sapkota, D.; Ibrahim, S.O.; Larsson, P.A.; Sand, L. Human papilloma virus, herpes simplex virus and epstein barr virus in oral squamous cell carcinoma from eight different countries. Anticancer. Res. **2012**, 32, 571–580.

175. Meurman, J.H. Infectious and dietary risk factors of oral cancer. Oral Oncol. **2010**, 46, 411–413.

176. al Kassaa, I.; Hober, D.; Hamze, M.; Chihib, N.E.; Drider, D. Antiviral Potential of Lactic Acid Bacteria and Their Bacteriocins. Probiotics Antimicrob. Proteins **2014**, 6.

177. Li, N.; Ma, W.-T.; Pang, M.; Fan, Q.-L.; Hua, J.-L. The Commensal Microbiota and Viral Infection: A Comprehensive Review. Front. Immunol. **2019**, 10.

178. Lou, E.; Kellman, R.M.; Shillitoe, E.J. Effect of herpes simplex virus type-1 on growth of oral cancer in an immunocompetent, orthotopic mouse model. Oral Oncol. **2002**, 38, 349–356.

179. Shillitoe, E.J.; Pellenz, C. Factors that limit the effectiveness of herpes simplex virus type 1 for treatment of oral cancer in mice. Clin. Cancer Res. **2005**, 11, 3109–3116.

180. Mukherjee, P.K.; Chandra, J.; Retuerto, M.; Sikaroodi, M.; Brown, R.E.; Jurevic, R.; Salata, R.A.; Lederman, M.M.; Gillevet, P.M.; Ghannoum, M.A. Oral mycobiome analysis of HIV-infected patients: Identification of Pichia as an antagonist of opportunistic fungi. PLoS Pathog. **2014**, 10, e1003996.

181. Kistler, J.; Arirachakaran, P.; Poovorawan, Y.; Dahlén, G.;Wade,W. The Oral Microbiome in HIV-positive Individuals. J. Med. Microbiol. **2015**, 64.

182. Herrera, S.; Martínez-Sanz, J.; Serrano-Villar, S. HIV, Cancer, and the Microbiota: Common Pathways Influencing Different Diseases. Front. Immunol. **2019**, 10.

183. Li, Y.; Saxena, D.; Chen, Z.; Liu, G.; Abrams, W.R.; Phelan, J.A.; Norman, R.G.; Fisch, G.S.; Corby, P.M.; Dewhirst, F.; et al. HIV infection and microbial diversity in saliva. J. Clin. Microbiol. **2014**, 52, 1400–1411.

184. Fulcher, J.A. Is the Oral Microbiome Important in HIV-Associated Inflammation? mSphere **2020**, 5, e00034-20. Int. J. Mol. Sci. **2020**, 21, 8061 27 of 29

185. Chikumaru, H.; Okubo, M.; Ueda, A.; Shirai, A.; Ishikawa, Y.; Mitsudo, K.; Ohta, Y.; Kaneko, A.; Tohnai, I.; Ishigatsubo, Y. HIV-Infected Persons Are at High Risk of Developing Precancerous Lesions of the Oral Cavity. Open Forum Infect. Dis. **2016**, 3.

186. Imai, K.; Yamada, K.; Tamura, M.; Ochiai, K.; Okamoto, T. Reactivation of latent HIV-1 by a wide variety of butyric acid-producing bacteria. Cell Mol. Life Sci. **2012**, 69, 2583–2592.

187. Mantri, C.K.; Chen, C.; Dong, X.; Goodwin, J.S.; Xie, H. Porphyromonas gingivalis-mediated Epithelial Cell Entry of HIV-1. J. Dent. Res. **2014**, 93, 794–800.

188. Pouncey, A.L.; Scott, A.J.; Alexander, J.L.; Marchesi, J.; Kinross, J. Gut microbiota, chemotherapy and the host: The influence of the gut microbiota on cancer treatment. Ecancermedicalscience **2018**, 12, 868.

189. Le, X.; Hanna, E.Y. Optimal regimen of cisplatin in squamous cell carcinoma of head and neck yet to be determined. Ann. Transl. Med. **2018**, 6, 229.

190. Mougeot, J.-L.C.; Stevens, C.B.; Almon, K.G.; Paster, B.J.; Lalla, R.V.; Brennan, M.T.; Mougeot, F.B. Caries-associated oral microbiome in head and neck cancer radiation patients: A longitudinal study. J. Oral Microbiol. **2019**, 11, 1586421.

191. Shao, Z.Y.; Tang, Z.S.; Yan, C.; Jiang, Y.T.; Ma, R.; Liu, Z.; Huang, Z.W. Effects of intensity-modulated radiotherapy on human oral microflora. J. Radiat Res. **2011**, 52, 834–839.

192. Gaetti-Jardim, E., Jr.; Jardim, E.C.G.; Schweitzer, C.M.; da Silva, J.C.L.; Oliveira, M.M.; Masocatto, D.C.;Dos Santos, C.M. Supragingival and subgingival microbiota from patients with poor oral hygiene submitted to radiotherapy for head and neck cancer treatment. Arch. Oral Biol. **2018**, 90, 45–52.

193. De Ryck, T.; Duprez, F.; Bracke, M.; Van de Wiele, T.; Vanhoecke, B. Dynamic Shifts in the Oral Microbiome during Radiotherapy. Clin. Res. Infect. Dis. **2015**, 2, 1013.

194. Almståhl, A.; Wikström, M.; Fagerberg-Mohlin, B. Microflora in oral ecosystems in subjects with radiation-induced hyposalivation. Oral Dis. **2008**, 14, 541–549.

195. Guobis, Ž.; Kareivien ˙ e, V.; Baseviˇcien˙ e, N.; Paipalien˙ e, P.; Niedzelskien˙ e, I.; Sabalys, G.; Kubilius, R.; Gervickas, A. Microflora of the Oral Cavity in Patients with Xerostomia. Medicina [Kaunas] **2011**, 47, 646–651.

196. Arrifin, A.; Heidari, E.; Burke, M.; Fenlon, M.R.; Banerjee, A. The Effect of Radiotherapy for Treatment of Head and Neck Cancer on Oral Flora and Saliva. Oral Health Prev. Dent. **2018**, 16, 425–429.

197. Almståh, I.A.;Wikström, M.; Stenberg, I.; Jakobsson, A.; Fagerberg-Mohlin, B. Oral microbiota associated with hyposalivation of different origins. Oral Microbiol. Immunol. **2003**, 18, 1–8.

198. Komiyama, E.Y.; Lepesqueur, L.S.S.; Yassuda, C.G.; Samaranayake, L.P.; Parahitiyawa, N.B.; Balducci, I.; Koga-Ito, C.Y. Enterococcus Species in the Oral Cavity: Prevalence, Virulence Factors and Antimicrobial Susceptibility. PLoS ONE **2016**, 11, e0163001.

199. Hu, Y.-J.; Wang, Q.; Jiang, Y.-T.; Ma, R.; Xia, W.-W.; Tang, Z.-S.; Liu, Z.; Liang, J.-P.; Huang, Z.-W. Characterization of oral bacterial diversity of irradiated patients by high-throughput sequencing. Int. J.Oral Sci. **2013**, 5, 21–25.

200. Aas, J.A.; Reime, L.; Pedersen, K.; Eribe, E.R.K.; Abesha-Belay, E.; Støre, G.; Olsen, I. Osteoradionecrosis contains a wide variety of cultivable and non-cultivable bacteria. J. Oral Microbiol. **2010**, 2.

201. Hansen, T.; Kunkel, M.; Kirkpatrick, C.J.; Weber, A. Actinomyces in infected osteoradionecrosis–underestimated? Hum. Pathol **2006**, 37, 61–67.

202. Jensen, S.B.; Mouridsen, H.T.; Bergmann, O.J.; Reibel, J.; Brünner, N.; Nauntofte, B. Oral mucosal lesions, microbial changes, and taste disturbances induced by adjuvant chemotherapy in breast cancer patients. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod. **2008**, 106, 217–226.

203. Napeñas, J.J.; Brennan, M.T.; Coleman, S.; Kent, M.L.; Noll, J.; Frenette, G.; Nussbaum, M.L.; Mougeot, J.-L.;Paster, B.J.; Lockhart, P.B.; et al. Molecular methodology to assess the impact of cancer chemotherapy on the oral bacterial flora: A pilot study. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod. **2010**, 109, 554–560.

204. Ramakrishna, C.; Corleto, J.; Ruegger, P.M.; Logan, G.D.; Peacock, B.B.; Mendonca, S.; Yamaki, S.; Adamson, T.; Ermel, R.; McKemy, D.; et al. Dominant

Role of the Gut Microbiota in Chemotherapy Induced Neuropathic Pain. Sci. Rep. **2019**, 9, 1–16.

205. Panebianco, C.; Andriulli, A.; Pazienza, V. Pharmacomicrobiomics: Exploiting the drug-microbiota interactions in anticancer therapies. Microbiome **2018**, 6, 92. Int. J. Mol. Sci. **2020**, 21, 8061 28 of 29

206. Gui, Q.F.; Lu, H.F.; Zhang, C.X.; Xu, Z.R.; Yang, Y.H. Well-balanced commensal microbiota contributes to anti-cancer response in a lung cancer mouse model. Genet. Mol. Res. **2015**, 14, 5642–5651.

207. Hong, B.-Y.; Sobue, T.; Choquette, L.; Dupuy, A.K.; Thompson, A.; Burleson, J.A.; Salner, A.L.; Schauer, P.K.;Joshi, P.; Fox, E.; et al. Chemotherapy-induced oral mucositis is associated with detrimental bacterial dysbiosis. Microbiome **2019**, 7, 66.

208. Raguse, J.D.; Gath, H.J.; Bier, J.; Riess, H.; Oettle, H. Gemcitabine in the treatment of advanced head and neck cancer. Clin. Oncol. [R Coll Radiol.] **2005**, 17, 425–429.

209. Laheij, A.M.; de Soet, J.J.; von dem Borne, P.A.; Kuijper, E.J.; Kraneveld, E.A.; van Loveren, C.;Raber-Durlacher, J.E. Oral bacteria and yeasts in relationship to oral ulcerations in hematopoietic stem cell transplant recipients. Support. Care Cancer **2012**, 20, 3231–3240.

210. Napeñas, J.J.; Brennan, M.T.; Bahrani-Mougeot, F.K.; Fox, P.C.; Lockhart, P.B. Relationship between mucositis and changes in oral microflora during cancer chemotherapy. Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod. **2007**, 103, 48–59.

211. Sonis, S.T. The Chicken or the Egg? Changes in Oral Microbiota as Cause or Consequence of Mucositis During Radiation Therapy. EBioMedicine **2017**, 18, 7–8.

212. Wijers, O.B.; Levendag, P.C.; Harms, E.R.; Gan-Teng, A.M.; Schmitz, P.I.; Hendriks, W.D.; Wilims, E.B.; van der Est, H.; Visch, L.L. Mucositis reduction by

selective elimination of oral flora in irradiated cancers of the head and neck: A placebo-controlled double-blind randomized study. Int. J. Radiat. Oncol. Biol. Phys. **2001**, 50, 343–352.

213. Stokman, M.A.; Spijkervet, F.K.; Burlage, F.R.; Dijkstra, P.U.; Manson,W.L.; de Vries, E.G.; Roodenburg, J.L. Oral mucositis and selective elimination of oral flora in head and neck cancer patients receiving radiotherapy: A double-blind randomised clinical trial. Br. J. Cancer **2003**, 88, 1012–1016.

214. Diaz, P.I.; Xie, Z.; Sobue, T.; Thompson, A.; Biyikoglu, B.; Ricker, A.; Ikonomou, L.; Dongari-Bagtzoglou, A. Synergistic interaction between Candida albicans and commensal oral streptococci in a novel in vitro mucosal model. Infect. Immun. **2012**, 80, 620–632.

215. Bamford, C.V.; d'Mello, A.; Nobbs, A.H.; Dutton, L.C.; Vickerman, M.M.; Jenkinson, H.F. Streptococcus gordonii modulates Candida albicans biofilm formation through intergeneric communication. Infect. Immun. **2009**, 77, 3696–3704.

216. Vanhoecke, B.; De Ryck, T.; Stringer, A.; Van de Wiele, T.; Keefe, D. Microbiota and their role in the pathogenesis of oral mucositis. Oral Dis. **2015**, 21, 17–30.

217. Mohan, S.P.; Bhaskaran, M.K.; George, A.L.; Thirutheri, A.; Somasundaran, M.; Pavithran, A. Immunotherapy in Oral Cancer. J. Pharm. Bioallied Sci. **2019**, 11, S107–S111.

218. Mei, Z.; Huang, J.; Qiao, B.; Lam, A.K.-y. Immune checkpoInt pathways in immunotherapy for head and neck squamous cell carcinoma. Int. J. Oral Sci. **2020**, 12, 16.

219. Routy, B.; Le Chatelier, E.; Derosa, L.; Duong, C.P.M.; Alou, M.T.; Daillère, R.; Fluckiger, A.; Messaoudene, M.; Rauber, C.; Roberti, M.P.; et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science **2018**, 359, 91–97.

220. Zheng, Y.; Wang, T.; Tu, X.; Huang, Y.; Zhang, H.; Tan, D.; Jiang, W.; Cai, S.; Zhao, P.; Song, R.; et al. Gut microbiome affects the response to anti-PD-1 immunotherapy in patients with hepatocellular carcinoma. J. Immunother. Cancer **2019**, 7, 193.

221. Gopalakrishnan, V.; Spencer, C.N.; Nezi, L.; Reuben, A.; Andrews, M.C.; Karpinets, T.V.; Prieto, P.A.; Vicente, D.; Hoffman, K.; Wei, S.C.; et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science **2018**, 359, 97–103.

222. Sivan, A.; Corrales, L.; Hubert, N.;Williams, J.B.; Aquino-Michaels, K.; Earley, Z.M.; Benyamin, F.W.; Lei, Y.M.; Jabri, B.; Alegre, M.L.; et al. Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science **2015**, 350, 1084–1089.

223. Baruch, E.N.; Youngster, I.; Ortenberg, R.; Ben-Betzalel, G.; Katz, L.H.; Lahat, A.; Barshack, I.; Dick-Necula, D.; Mamtani, R.; Bloch, N.; et al. Abstract CT042: Fecal microbiota transplantation [FMT] and re-induction of anti-PD-1 therapy in refractory metastatic melanoma patients - preliminary results from a phase I clinical trial [NCT03353402]. Clin. Trials **2019**, 79, CT042. Int. J. Mol. Sci. **2020**, 21, 8061 29 of 29

224. Chaput, N.; Lepage, P.; Coutzac, C.; Soularue, E.; Le Roux, K.; Monot, C.; Boselli, L.; Routier, E.; Cassard, L.; Collins, M.; et al. Baseline gut microbiota predicts clinical response and colitis in metastatic melanoma patients treated with ipilimumab. Ann. Oncol. **2017**, 28, 1368–1379.

225. Coutzac, C.; Jouniaux, J.-M.; Paci, A.; Schmidt, J.; Mallardo, D.; Seck, A.; Asvatourian, V.; Cassard, L.; Saulnier, P.; Lacroix, L.; et al. Systemic short chain fatty acids limit antitumor effect of CTLA-4 blockade in hosts with cancer. Nat. Commun. **2020**, 11, 2168.

226. Lax, A.J.; Thomas, W. How bacteria could cause cancer: One step at a time. Trends Microbiol. **2002**, 10, 293–299.

227. Amin, M.B.; Greene, F.L.; Edge, S.B.; Compton, C.C.; Gershenwald, J.E.; Brookland, R.K.; Meyer, L.; Gress, D.M.; Byrd, D.R.;Winchester, D.P. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. CA Cancer J. Clin. **2017**, 67, 93–99.

2- [A] The ultra-structural, metabolomic and metagenomic characterisation of the Sudanese smokeless tobacco 'Toombak'

Amel Sami 1,2, Imad Elimairi 2, *, Dhrati Patangia 1, Claire Watkins 3, Catherine Stanton 1,3, R. Paul Ross 1, and C. Anthony Ryan 4

1 APC Microbiome Ireland, School of Microbiology, University College Cork, Cork T12 YN60, Ireland.
2 Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, National Ribat University, Nile Street, Khartoum 1111, Sudan
3 Teagasc Food Research Centre, Moorepark, Fermoy, Cork P61 C996, Ireland
4 Department of Paediatrics and Child Health, University College Cork, Cork T12 DFK4, Ireland.

This chapter has been published as **Sami A**, Elimairi I, Patangia D, Watkins C, Ryan CA, Ross RP, Stanton C. The ultra-structural, metabolomic and metagenomic characterisation of the Sudanese smokeless tobacco 'Toombak'. Toxicol Rep. 2021 Aug 5; 8:1498-1512. doi: 10.1016/j.toxrep.2021.07.008. PMID: 34401360; PMCID: PMC8355839. [see Appendix 2]

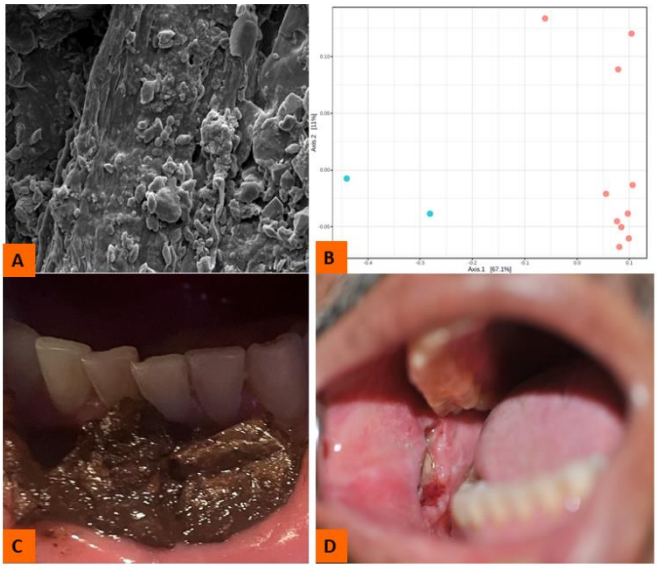
2.1 Abstract

Toombak is a smokeless tobacco produced from the *Nicotiana Rustica* tobacco plant from Sudan. Pre-prepared and ready to buy Toombak samples were analysed using mass spectrometry [heavy metals], gas and liquid chromatography [metabolomics], 16S rRNA metagenomic sequencing [microbiome] and Phylogenetic Investigation of Communities by Reconstruction of Unobserved States [PICRUSt], scanning electron microscopy with energy dispersive X-ray spectroscopy [SEM-EDX] and pH analysis.

Chromium, cobalt, and copper were high in the pre-prepared form of Toombak while iron, tobacco specific nitrosamines [TSNAs], formaldehyde and acetaldehyde were high in both types. *Firmicutes* and *Actinobacteria* dominated Toombak. Samples of ready to buy Toombak showed inter-variational differences depending on place of purchase. We found *Virgibacillus* was increased in the pre-prepared form while *Corynebacterium casei, Atopococus tabaci, Atopostipes suicloacalis, Oceanobacillus chironomi* and *Staphylococcus gallinarum* were the most abundant species in the ready to buy forms. PICRUSt analysis highlighted increased activity of metal transport systems in the ready to buy samples as well as an antibiotic transport system. SEM-EDX highlighted large non-homogenous, irregular particles with increased sodium, while pH of samples was in the alkaline range.

The final composition of Toombak is affected by its method of preparation and the end product has the potential to impart many negative consequences on the health of its users. TSNA levels observed in Toombak were some of the highest in the world while the micro-environment of Toombak supports a distinct microbiota profile.

Graphical abstract: Toombak smokeless tobacco, Sudan… The story from start to finish



A: Scanning electron microscopy outlining bacterial contamination in ready to buy Toombak. *Firmicutes* and *Actinobacteria* are the predominant phyla.

B: PCoA plot [pink = Khartoum North, blue = Omdurman] highlighting the significant β diversity of ready to buy Toombak which is C C: produced from the *Nicotiana Rustica*. Place of purchase affects the final composition of Toombak.

C Toombak dip in the oral cavity in the lower mucolabial fold.

D: Development of an oral squamous cell carcinoma in the right, buccal and retromolar region in a chronic Toombak user.

2.2 Introduction

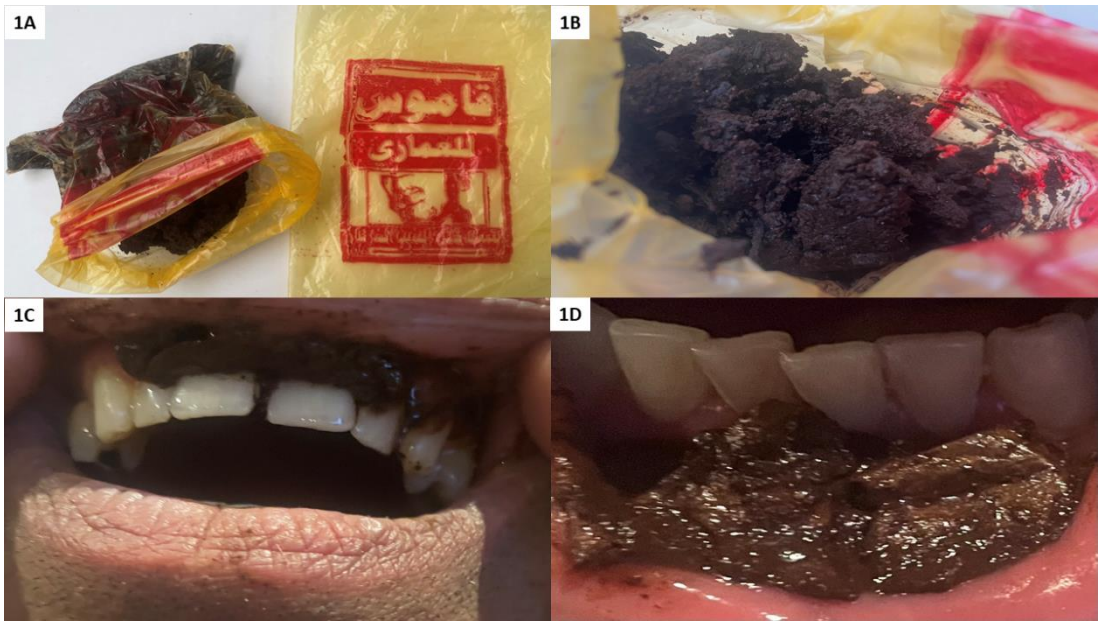
Smokeless tobacco refers to the use of unburnt or pyrolysis free tobacco, differing around the world in tobacco plant source, nicotine content and additives. Products are often applied orally [chewed, sucked, or placed passively to the gingivae] and include forms such as chewing tobacco, moist forms, dry snuff, and snus, while fine powdered types are administered nasally. It is thought that more than 300 million people use a form of smokeless tobacco worldwide, in over 70 countries that include Southeast Asia, many regions of Africa, the Middle East, Europe [Scandinavian countries], the Americas and Russia [1]. In Sudan, there are an estimated 4–10 million users of Toombak; a moist smokeless tobacco used primarily by males. Idris et al. [1998], concluded that 60 % of Sudanese households contain at least one family member with Toombak use, while it has been estimated that Toombak prevalence of use amongst adolescents, young adults and those above 60 years of age is; 34 %, 32 % and 47 % respectively [2].

Toombak is produced from the *Solanaceae* species, *Nicotiana Rustica,* a plant with characteristic yellow flowers, that contains up to nine times more nicotine compared to *Nicotiana tabacum*, that is more widely utilised in the production of tobacco products worldwide [3]. *Nicotiana Rustica* is also used to make smokeless tobacco products in Turkey [Maras powder] [4], South America [Rapé] [5], Vietnam [Thuốc lào] and Russia [Makhorka]. Plantation begins around November in western Sudan, leaves are cultivated from February to March, and air curing occurs during April and May. During this latter stage, temperatures can reach 45 °C–60 °C. Leaves change from yellow to brown in a process of natural or ‘compost’ like fermentation and the Toombak is then further prepared by milling to non-homogenous particles.

Toombak production is finalised by the addition of alkaline carbonates, flavourings, and other additives. Sodium bicarbonate is commonly incorporated to the final mix in order to improve taste and promote effective nicotine absorption into the bloodstream. There are limited regulatory laws for Toombak production and sale and so it is highly likely that it is contaminated by human, animal, soil, and other environmental factors, which have not been fully reported.

Toombak is sold in polythene bags [Figure 1A], is a dark brown colour with a moist coarse consistency and has a distinct pungent aroma [Figure 1B]. A regular Toombak user would place approximately 6−10 g of a dip of Toombak in various areas of the oral cavity but mainly the upper or lower mucolabial folds and gingivobuccal sulci [Figures 1C and 1D] as well as the floor of the mouth. The dip is often replaced around 10–15 times per day, usually when the nicotine effect becomes bland, and it may also be retained during sleep [6].

Figure 1:



1A: the smokeless tobacco, Toombak locally known as 'sultan alkeef' in Sudan is sold in common polythene bags from local vendors around the country. 1B: Toombak is of a brown coarse structure ranging from finely ground to macro-sized particles. 1C: a user placing the Toombak dip in the upper and 1D: the lower mucolabial folds.

Smokeless tobacco use including Toombak has been attributed to cardiovascular disease, male infertility, psychological disturbances, periodontal disease, premalignant lesions and oral, oro-pharyngeal, oesophageal and numerous other cancers [[7], [8], [9]]. 81% of oral cancers in Sudan are thought to occur amongst Toombak users and present at the site of Toombak placement [10]. In fact, in a recent systematic review it was reported that Sudan has the second highest oral cancer incidence in the Middle East, primarily attributed to Toombak use [11].

Current knowledge is lacking as to why the properties of Toombak adversely affect human health. This study was undertaken to better understand the physical, biological and chemical properties of Toombak, including its microbial-ecological niche and how these findings may reflect on its final composition and users health.

2.3 Materials and methods

Sample collection

Two types of Toombak were collected; pre-prepared and ready to buy samples. S01 was sourced from Darfur, western Sudan, just after completion of the fermentation process in order to evaluate the smokeless tobacco prior to the preparatory stages. Twenty samples of ready to buy Toombak [S2-S21], were then collected from different vendors of Toombak sale; S2-S11 from Khartoum North and S12-S21 from Omdurman located in the capital Khartoum, Sudan. These were chosen to represent common points of Toombak purchase in these two cities.

Ready to buy samples were initially stored in a similar fashion to how Toombak users would store their product, at room temperature in airtight polythene bags. Samples were then shipped to Cork, Ireland, less than one week after sample collection and refrigerated, to prevent any changes to the product. The pre-prepared form was also refrigerated.

Heavy metals

Analysis of nine heavy metals/major cations [mg/kg] was performed on two of the 21 samples; the pre-prepared and a ready to buy sample of Toombak by ALS Life Sciences Ltd, [ALS Czech Republic]. Samples were homogenized and mineralized by acids and hydrogen peroxide prior to analysis. Arsenic, cadmium, chromium, cobalt, copper, lead and nickel were analysed using mass spectrometry with inductively coupled plasma. Iron analysis followed the same protocol although atomic emission spectrometry was used while mercury was determined by atomic absorption spectrometry.

Metabolomics

Metabolomic analysis was performed on all samples of Toombak by MS-OMICS [Vedbaek, Denmark]. Samples were extracted by shaking [2 rpm/sec] for one hour in 100 mM ammonium acetate buffer. For quality control, a mixed pooled sample was

prepared by taking a small aliquot from each sample. This sample was analysed at regular intervals throughout the sequence. Matrix effects were tested for quantified compounds by spiking the quality control sample at a minimum of two levels. The volatile organic compound method is a gas chromatography – mass spectrometry method targeted to volatile compounds using a low-mid polarity column. Urethane and N-nitrosodimethylamine were analysed without derivatization by adding methyl tert-butyl ether [MTBE] to the samples and injecting the organic phase to the instrument.

Formaldehyde and acetaldehyde required derivatization which was performed using a protocol similar to the one described by Bao et al. [2014] [12]. The raw gas chromatography – mass spectrometry data were processed by software developed by MS-OMICS and collaborators which use a powerful PARAFAC2 model.

For liquid chromatography metabolites, samples were diluted ten times in eluent A prior to analysis. The data were analysed using both a targeted and an untargeted approach. The targeted approach was used to extract the response of compounds included in a standard list covering 85 compounds. For the untargeted approach, feature extraction was conducted using mzMine, an open-source software for mass spectrometry data processing.

Levels of the heavy metals arsenic, cadmium, chromium, lead, mercury and nickel as well as formaldehyde, acetaldehyde, and the tobacco specific nitrosamines [TSNAs] in Toombak were compared with the GothiaTek standard for Swedish snus and chewing tobacco production adopted by the members of the trade organisation, European Smokeless Tobacco Council [ESTOC] and is the regulatory manufacturing standard for the production of Swedish snus and chewing tobacco [13]. The GothiaTek standard ensures the levels of constituents in smokeless tobacco are present in amounts that do not pose harm to the user [14]. Where the components in Toombak could not be compared with this standard, they were compared with other findings in the literature.

Microbiome

Thirteen samples were analysed by metagenomic sequencing for microbiome composition. These were S01, the pre-prepared form, and the ready to buy samples; S2−11, from Khartoum North and S12 -S13 from Omdurman.

DNA extraction

Toombak samples and a positive and negative control, were subjected to the DNA extraction protocol provided by the manufacturer of the DNeasy Powersoil Kit [Qiagen, Hilden, Germany]. Eluted DNA was transferred to Eppendorf tubes and frozen at −80 °C for further analysis.

Amplicon sequencing

Amplicon PCR was carried out on the Toombak DNA extracted samples using 2x KAPA HiFi HotStart Ready Mix and universal forward and reverse primers, to target the V3 – V4 hypervariable regions of 16S rRNA selected for next generation sequencing based diversity studies.

The following thermal cycling conditions were used: 3 min of initial denaturation at 95 °C followed by 30 cycles each at 95 °C for 30 s, annealing at 55 °C for 30 s, elongation at 72 °C for 30 s, and then at 72 °C for 5 min. Electrophoresis using 1.5 % agarose gel with Invitrogen DNA loading dye [ThermoFisher Scientific], verified base pair size which was expected at 550 base pairs. All samples were positive on the agarose gel while the control samples were negative. This was followed by a clean-up stage using AMPure XP beads [Beckman Coulter] to purify the amplicon from free primers and primer dimer species.

Index PCR was then continued by attaching dual indices and Illumina sequencing adapters using Nextera XT index kits and the following PCR thermal cycling conditions were used: 3 min of initial denaturation at 95 °C followed by 8 cycles each at 95 °C for 30 s, annealing at 55 °C for 30 s, elongation at 72 °C for 30 s, and then at 72 °C for 5 min. AMPure XP beads [Beckman Coulter] were utilised for final clean-up, and the library quantified using the Qubit® 2.0 Fluorometer and Qubit dsDNA high sensitivity assay kit [ThermoFisher Scientific]. Samples were sequenced using the V3-V4 regions of the 16S rRNA gene on an Illumina MiSeq device 2 × 300 platform [Illumina, Inc. San Diego].

Bioinformatic analysis

The 300-base paired end FASTQ products generated from 16S rRNA sequencing were merged using fast length adjustment of short reads [FLASH] using default parameters [15]. QIIME's split_libraries_fastq.py script was used for demultiplexing and filtering of the FASTQ sequence data. Further quality filtering was performed using the USEARCH analysis tool. Briefly, single unique reads were removed following denoising, chimera removal and grouping into operational taxonomic units [OTUs] at 97 % similarity and was performed using USEARCH v7 [64 bit] [16]. Alignment of OTUs was performed using python nearest alignment space termination [PyNAST], a flexible tool for aligning sequences to a template alignment [17]. Further, assigning of taxonomic ranks was performed by using the Basic Local Alignment Search Tool [BLAST®], against the SILVA SSURef database release 132 [18]. The U.S National Library of Medicine, National Centre for Biotechnology Information was then accessed in order to assign species identity to those FASTA sequences with > 98 % percentage homogeneity with BLASTn [19].

Metagenomic statistical analysis

Statistical analysis with graphical outputs was undertaken using Microbiome Analyst R package. Data normalisation followed by compositional and functional profiling, comparative analysis, and Phylogenetic Investigation of Communities by Reconstruction of Unobserved States [PICRUSt] for the Toombak samples [20, 21] were performed.

Quantitative visualisation using stacked bar charts representing abundant phyla, families, and genera within all Toombak samples was achieved. Alpha [α] diversity was used to identify community profile and any significant variation in the microbiome of Toombak. Beta [β] diversity evaluated the relative abundance of OTUs and the clustering and outliers amongst samples. This was formulated using Bray – Curtis matrices and visualised by principal coordinate analysis [PCoA] in the ready to buy forms only. Correlations were assessed through pattern search between the two ready to buy groups while classic univariate analysis [ANOVA] evaluated the variation of microbiome between Khartoum North and Omdurman samples. DeSeq2

was also employed to assess differential abundance between the market groups using shrinkage estimation for dispersion and fold change to improve estimates of differential expression [22]. False detection rates [FDR] or q values were reported to limit false positive results [23].

Scanning electron microscopy with energy dispersive X-ray spectroscopy [SEM-EDX]

Structural analysis was achieved by scanning electron microscopy at 35, 800 and 4000 x magnification of two of the 21 samples; S01 and a ready to buy sample of Toombak. A form of snuff bought in the Republic of Ireland [McChrystals snuff], was also used as a comparison in this assessment. Samples were mounted onto aluminium stubs using double sided carbon tape. Samples were sputter-coated with a 5 nm layer of gold palladium [80:20] using a Quorum Q150 RES Sputter Coating System [Quorum Technologies, UK], before being examined using a JEOL JSM 5510 Scanning Electron Microscope [JEOL Ltd., Japan]. Digital electron micrographs were obtained of areas of interest. Energy dispersive X-ray spectroscopy [EDX] analysis was performed using an INCA x-sight EDX Detector [Oxford Instruments, Buckinghamshire, UK] on samples that were not sputter coated to determine the characteristic elements within the sample.

pH

pH was determined on all 21 samples by weighing equal parts of Toombak [3 g] at room temperature and adding 3 mL dH2O to each sample. The mixtures were vortexed for 5 min, decanted into separate Eppendorf tubes and centrifuged at 21,380 X g for 10 min. The supernatants were then transferred to fresh bijou tubes and analysed using a laboratory grade pH meter that was fully calibrated.

2.4 Results

Heavy metals

Arsenic, cadmium, lead, and mercury were at acceptable levels in the Toombak samples. Nickel levels were at the upper range of acceptable limits while chromium, cobalt and copper were high in the pre-prepared form. Iron was found to be the most abundant heavy metal in Toombak. The International Agency for Research on Cancer

[IARC] categorisation for heavy metal carcinogen classification is listed alongside the GothiaTek standard or references from the literature [Table 1].

Metabolomics of Toombak

Using the volatile organic compound method; Benzaldehyde, benzyl alcohol, phenylethyl alcohol, formaldehyde, acetaldehyde, propionaldehyde and 3 – methyl – valeric acid were detected in the Toombak samples while urethane and N-nitrosodimethylamine were not detected.

Both Benzaldehyde and benzyl alcohol were markedly increased in the pre-prepared sample S01, while in the ready to buy samples, lower concentrations were found of the former and variable concentrations of the latter. Phenylethyl alcohol was elevated in the pre-prepared sample as well as in samples from both Khartoum North S08, S09 and Omdurman S13, S14, S15 and S20, while 3 – methyl – valeric acid was low in all samples except in the ready to buy sample S21 from Omdurman where it was markedly increased.

Volatile aldehydes were increased in Toombak compared to European standards [Table 2]. Formaldehyde ranged from 8−17 mg/kg in samples from Omdurman and 7−12 mg/kg from those found in Khartoum North. Compared to the allowed levels of this aldehyde in Swedish snus; only two of the 21 samples of Toombak, contained formaldehyde levels below 7.5 mg/kg. Acetaldehyde and propionaldehyde were also elevated in all samples. Acetaldehyde was found to be in the range of 500–1300 mg/kg in Toombak.

Table 1: Heavy metal content in one pre-prepared and one ready to buy sample of Toombak.

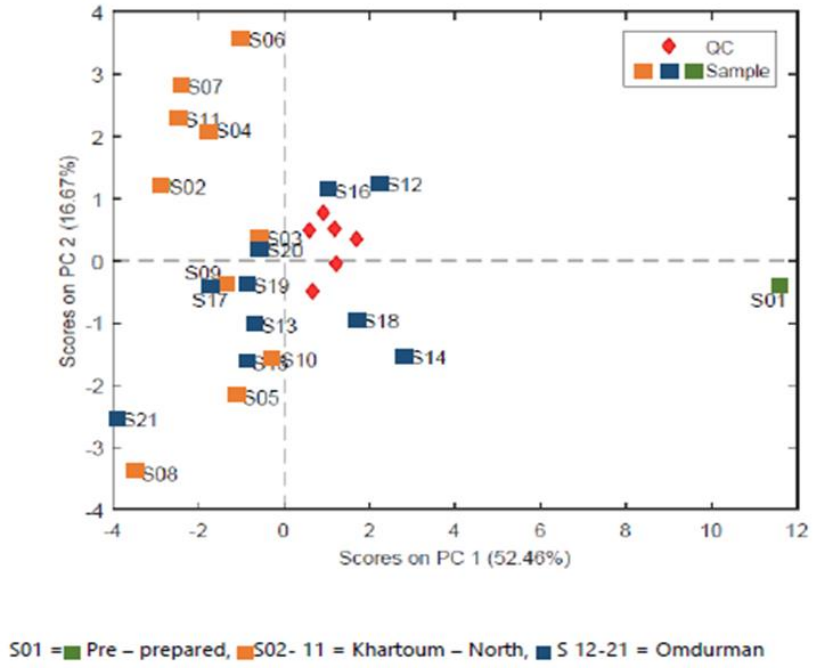
Metals/ major cations	**Limit of reporting**	**Pre-prepared sample [mg/kg]**	**Ready to buy sample [mg/kg]**	**GothiaTek standard limits [mg/kg]**	**Other reference**	**IARC carcinogen classification**
Arsenic	0.10	0.12	< 0.10	0.25	–	Group 1
Cadmium	0.04	0.34	0.09	0.5	–	Group 1
Chromium	0.20	1.86	1.50	1.5	–	Group 1
Cobalt	0.05	0.99	0.64	–	0.98 mg/kg [25]	Group 2B
Iron	1	1660	1270	–	8–18 mg/d* [26]	Group 2B
Copper	0.10	4.80	2.73	–	2-3 mg/d* [27]	Group 3
Lead	0.05	0.39	0.35	1.0	–	Group 2B
Mercury	0.003	0.014	0.004	0.02	–	Group 2B
Nickel	0.20	2.03	1.36	2.25	–	Group 1

Table 2: Levels of formaldehyde and acetaldehyde in Toombak samples compared to GothiaTek standards.

Compound (aldehyde)	Pre- prepared Toombak	Omdurman ready to buy Toombak range	Khartoum North ready to buy Toombak range	GothiaTek standard limits of aldehydes in Swedish snus
Formaldehyde (mg/kg)	25 mg/kg	8–17 mg/kg	7–12 mg/kg	7.5 mg/kg
Acetaldehyde (mg/kg)	900 mg/kg	500–1300 mg/kg	600–1200 mg/kg	25 mg/kg

The pre-prepared Toombak had a distinct metabolomic composition from all the remaining ready to buy samples as highlighted on the principal component analysis model [PCA] [Figure 2]. Quality control samples were tightly grouped together indicating analytical variance greatly exceeded biological variance. Quality control samples were also located close to the centre of the plot indicating that matrix effects were not an issue with these samples and that data were of high quality [24].

Figure 2:



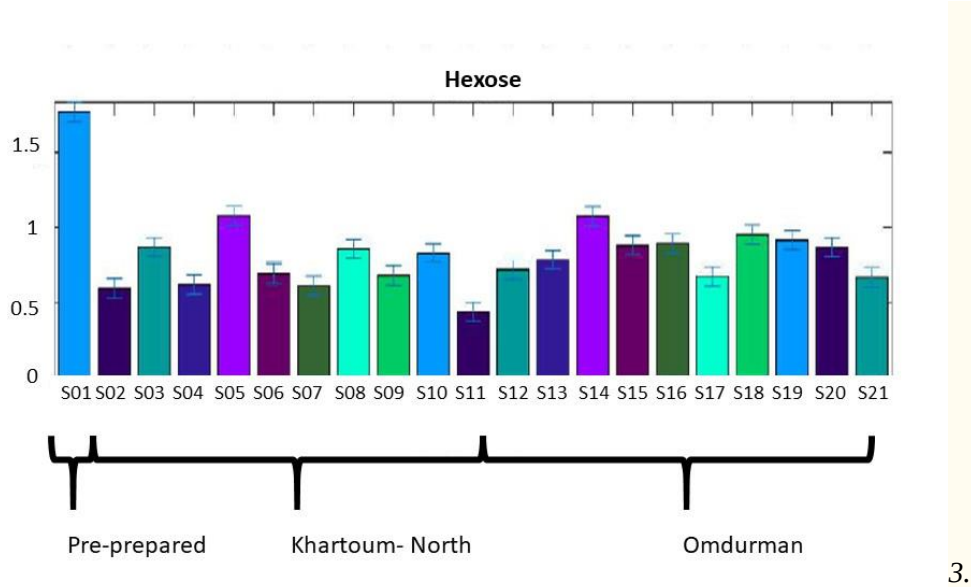
2: Multivariate score plot: PCA model describing two dimensional LC/MS data of Toombak. Data have been auto scaled. S01 = pre-prepared Toombak, S02 – S11 = Khartoum North samples and S12 –S21= Omdurman samples. Distinct variation between S01 and the ready to buy samples in the plot suggests handmade preparation of Toombak greatly modifies the final product. No major variation was found between the ready to buy groups.

Metabolomic findings were then compared for the 16 compounds included in the reduced dataset using separate bar charts for each of the compounds. These compounds were amino acids [valine, carnitine, citrulline, proline and alanine], nicotine, tobacco specific nitrosamines [[N-nitrosonornicotine [NNN], nicotine-derived nitrosamine ketone [NNK], N-nitrosoanabasine [NAB], N-nitrosoanatabine [NAT]], hexose, adenine, hypoxanthine, inosine, pantothenic acid and succinic acid. Additionally, 1058 features were extracted using mzMine.

Highest levels of adenine were found in the pre-prepared sample while ready to buy Toombak samples all had low levels. Hypoxanthine was lowest in S01 and highest in Khartoum North samples S05, S07, S08, and S09. In particular, inosine was negligible in the pre-prepared sample as well as in S08 from Khartoum North and Omdurman samples S16, S17, S18, S19 and S21, while it was markedly elevated in Khartoum North samples S03, S05 and S10 and Omdurman samples S13 and S20.

All samples contained the amino acids carnitine, citrulline, valine, alanine, and proline in varying amounts. Carnitine was found to be highly variable while citrulline was highest in the pre-prepared sample and S12 [Omdurman]. Valine was high in all samples while alanine was increased only in the pre-prepared form and one sample from Omdurman; S14. Proline was found to be elevated in the pre-prepared form and reduced in the remaining ready to buy samples. Hexose [Figure 3] was markedly elevated in the pre-prepared form [S01] but lower in all ready to buy samples while pantothenic acid showed a similar pattern. Succinic acid was highest in the pre-prepared form but was also elevated in the ready to buy samples.

Figure 3:



3.

3: Bar graph. Hexose monosaccharide in Toombak samples was found to be significantly elevated in the pre-preared sample S01, compared to the remaining ready to buy samples

Nicotine in the pre-prepared form was 31 mg/g but ranged lower; between 16−25 mg/g in the ready to buy samples. Mean levels of the individual TSNAs in the ready to buy samples of Toombak were found to be as follows: NAB, 10-40 mg/kg, NAT, 5-14mg/kg, NNK, 1200-4700 mg/kg and NNN, 60-130 mg/kg. The highest concentrations of nicotine [31 mg/g], NNN [160 mg/kg], NAT [28 mg/kg] and NAB [46 mg/kg] were found in the pre-prepared sample while concentrations of NNK were consistent throughout both pre-prepared and ready to buy Toombak. Cumulative NNN and NNK in the ready to buy samples of Toombak were further compared to the GothiaTek standard [Table 3] and were found to range between 1270−4830 mg/kg amongst samples.

Table 3: Levels of nicotine and the tobacco specific nitrosamines [TSNAs], NNK, NAB, NAT and NNN in the pre-prepared and ready to buy Toombak samples. NNN and NNK were cumulatively compared in mg/kg with GothiaTek standard maximum limits allowed.

Type	**Sample**	**Compound mg/g**					**GothiaTek standard NNN + NNK mg/kg = [0.95]**
		Nicotine	**NNK**	**NAB**	**NAT**	**NNN**	**NNN + NNK converted from mg/g to mg/kg in Toombak samples**
Pre-prepared Toombak	S01	31	3.5	0.046	0.028	0.16	3660
Khartoum North Toombak		**Nicotine**	**NNK**	**NAB**	**NAT**	**NNN**	
	S02	22	1.8	0.012	0.006	0.07	1870
	S03	21	2.6	0.017	0.007	0.09	2690
	S04	22	1.9	0.014	0.006	0.08	1980
	S05	19	2.4	0.017	0.008	0.08	2480
	S06	25	1.5	0.014	0.006	0.08	1580
	S07	24	1.2	0.015	0.006	0.08	1280
	S08	19	1.2	0.011	0.005	0.07	1270

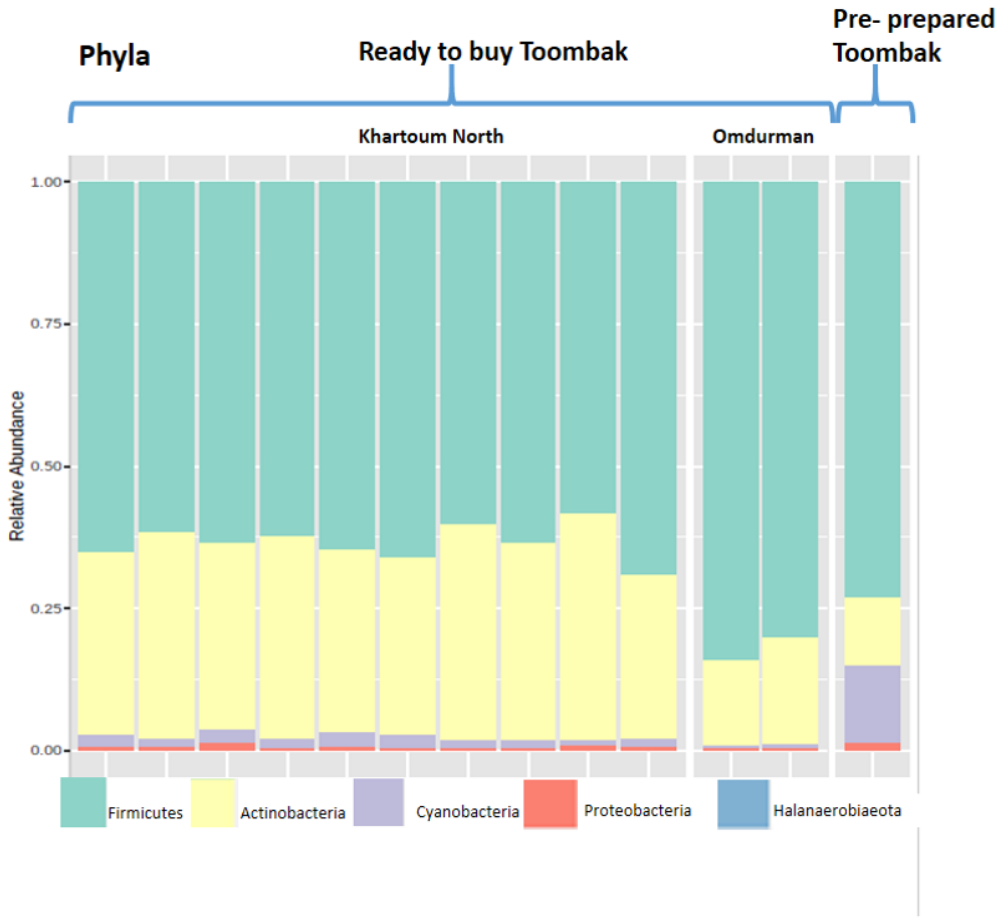
	S09	20	2.0	0.022	0.008	0.09	2090
	S10	22	2.3	0.021	0.011	0.10	2400
	S11	22	1.9	0.021	0.011	0.10	2000
Omdurman Toombak	**Sample**	**Nicotine**	**NNK**	**NAB**	**NAT**	**NNN**	**NNN + NNK converted from mg/g to mg/kg in Toombak samples**
	S12	20	4.7	0.037	0.011	0.13	4830
	S13	20	2.9	0.024	0.011	0.10	3000
	S14	22	2.7	0.029	0.014	0.12	2820
	S15	19	3.0	0.022	0.008	0.10	3100
	S16	20	3.6	0.025	0.007	0.11	3710
	S17	20	2.4	0.021	0.007	0.11	2510
	S18	23	3.4	0.040	0.013	0.12	3520
	S19	22	1.9	0.015	0.007	0.08	1980
	S20	23	2.0	0.026	0.010	0.11	2110
	S21	16	1.4	0.010	0.005	0.06	1460
Total Range of compounds		**16-25 mg/g**	**1.2-4.7 mg/g**	**0.010 – 0.040 mg/g**	**0.005 – 0.014 mg/g**	**0.06 – 0.13 mg/g**	**1260 – 4830 mg/kg**

amongst ready to buy samples							
LOD		**0.01**	**0.02**	**0.003**	**0.002**	**0.002**	

Microbiome of Toombak

Overall, 258 assigned OTUs or features from a total read count of 1,203,891 were obtained. The maximum count was 171,130 while the minimum count per sample was 61,860 with an average of 92,607 reads. A total of 66 low abundance features were removed, based on their prevalence being under 10 %, while a further eight low variance features were removed based on a standard deviation under 5%. The remaining OTUs or features after data filtering were 141. In total, 13 phyla, 12 classes, 47 orders, 85 families and 134 genera were characterised in Toombak. *Firmicute*s and *Actinobacteria* where the most abundant phyla amongst all Toombak samples. *Cyanobacteria* were high in the sample obtained from western Sudan [pre-prepared form], while *Proteobacteria* and *Halenaerobiaeota* were the least abundant phyla in all samples [Figure 4A].

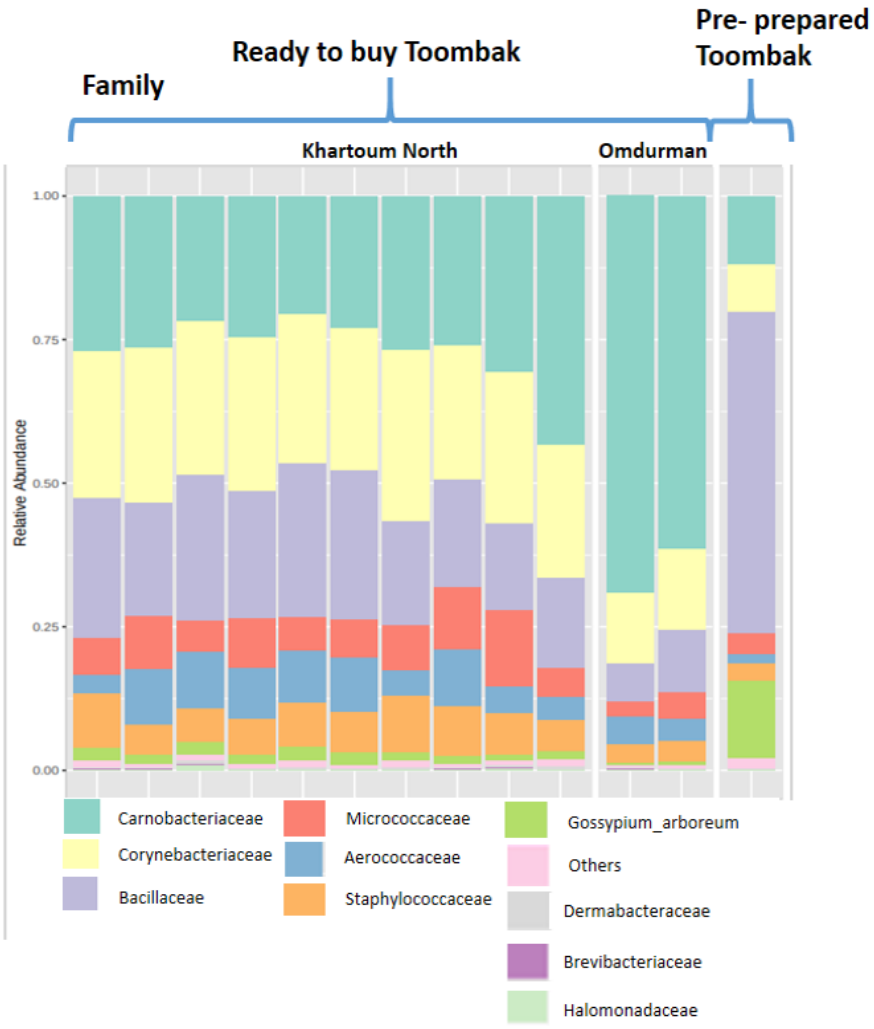
Figure 4A:



4A: Microbiome Analyst R programming of abundant phyla within Toombak samples. Firmicutes *are the most abundant phylum, followed by* Actinobacteria. *Pre-prepared samples harboured increased* Cyanobacteria*; due to the less degraded plant composition of the pre-prepared form.*

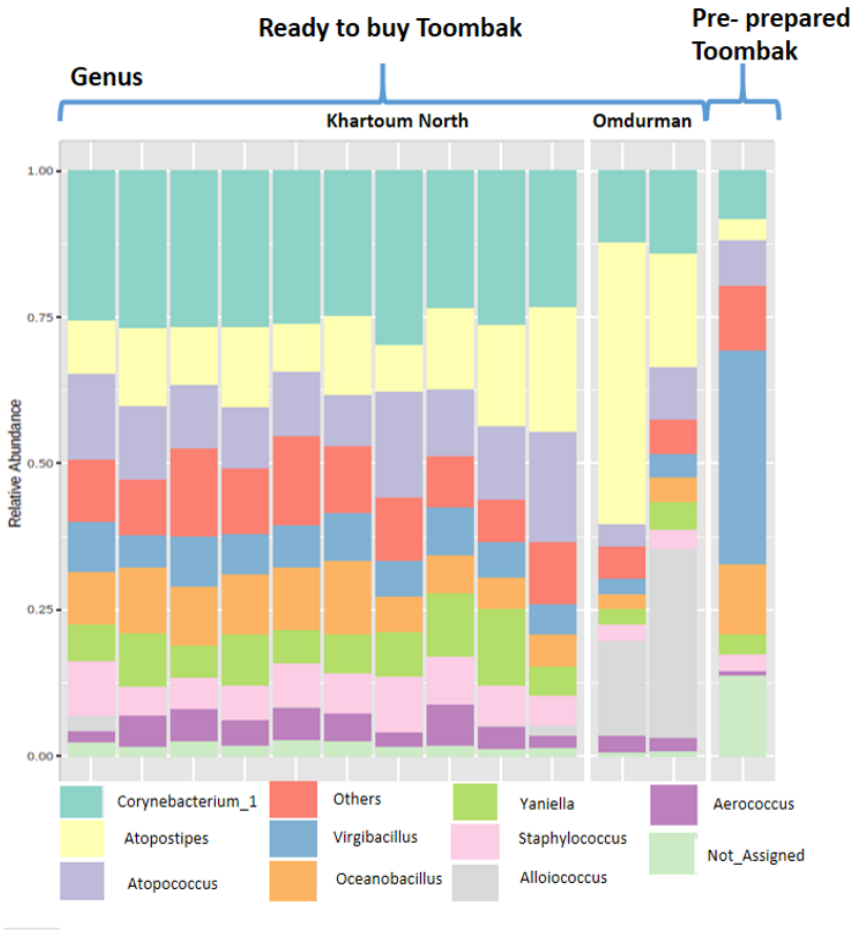
The families *Carnobacteriaceae* and *Corynebacteriaceae* dominated in ready to buy forms while *Bacillaceae* dominated in the pre-prepared Toombak [Figure 4B]. Meanwhile, figure 4C highlights the genera distribution in the ready to buy samples from Khartoum North and Omdurman as well as the pre-prepared form. Relative abundance amongst the ready to buy Toombak included *Corynebacterium_1* [24%], *Atopostipes* [16%], *Atopococcus* [12%], *Oceanobacillus* [8%], *Yaniella* [7%], *Virgibacillus* [7%], *Staphylococcus* [6%] *and Aerococcus* [5%].

Figure 4B:



4B: Stacked bar chart representing 16S rRNA compositional data of families within Toombak. Corynebacteriaceae and Carnobacteriaceae are predominant in the ready to buy forms while in the pre-prepared Toombak; Bacillaceae are markedly increased.

Figure 4C:



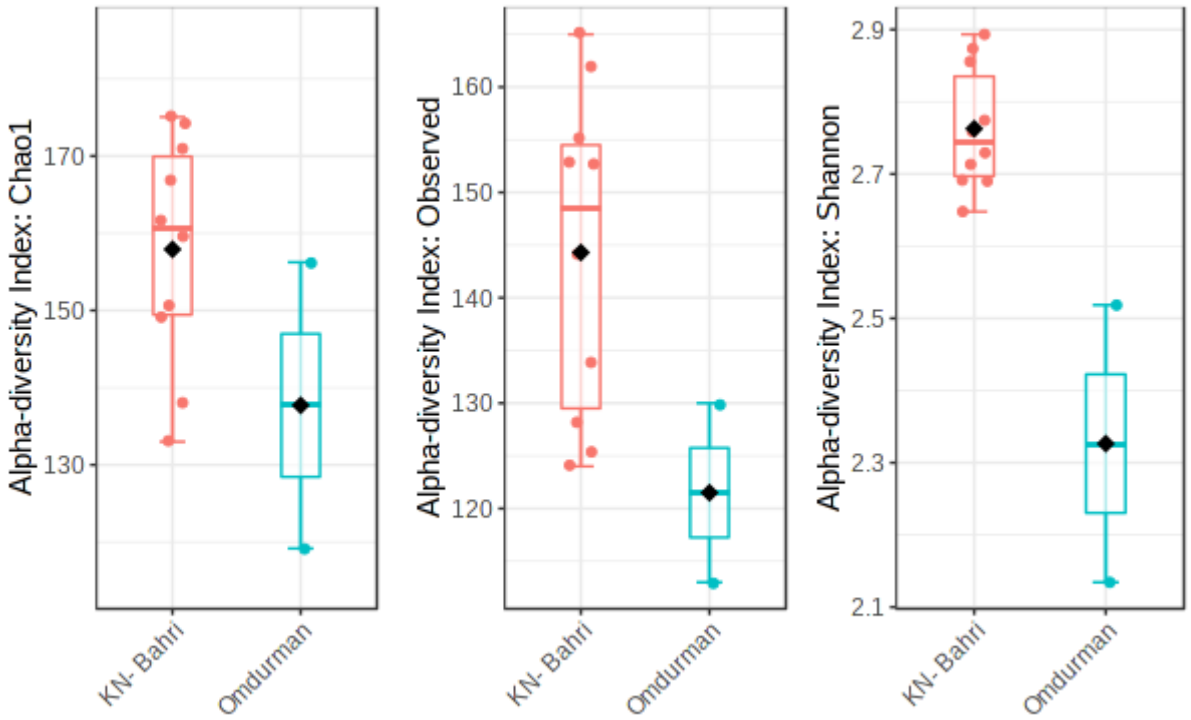
4C: Stacked bar chart of genera composition amongst Toombak samples. Corynebacterium_1, Atopostipes and Atopococcus are the most abundant genera in Khartoum North samples. Atopostipes and Alloiococcus were found to be abundant amongst Omdurman samples while Virgibacillus was the most abundant genus in the pre-prepared form. Oceanobacillus, Staphylococcus, Yaniella and Aerococcus are other genera identified in the ready to buy Toombak samples

When comparing both ready to buy groups, *Corynebacterium_1* was the most abundant genus in Toombak from Khartoum North [26 %], followed by *Atopostipes* [13%], *Atopococcus* [13%] and *Oceanobacillus* [9%], while *Atopostipes* was the most abundant genus in Toombak from Omdurman [34 %] followed by *Alloiococcus* [24 %] and *Corynebacterium_1* [13 %]. Other variations included *Staphylococcus* that

accounted for 6% of the composition of Toombak from Khartoum North but only 3% of Omdurman samples and *Atopococcus* which accounted for 13 % of Khartoum North samples but only 7% of Omdurman samples. On the other hand, *Alloiococcus* was highly distributed amongst Omdurman samples [24 %], while it was minimally detected in Toombak samples from Khartoum North. *Virgibacillus* was the most abundant genus in the pre-prepared form [q=3.6314e-8].

Alpha diversity was similar amongst ready to buy Toombak samples, attributed to OTU feature domination as indicated by a non-significant Chao1 [P = 0.46], observed richness [p = 0.16], and Shannon diversity index [p = 0.25] [Figure 5].

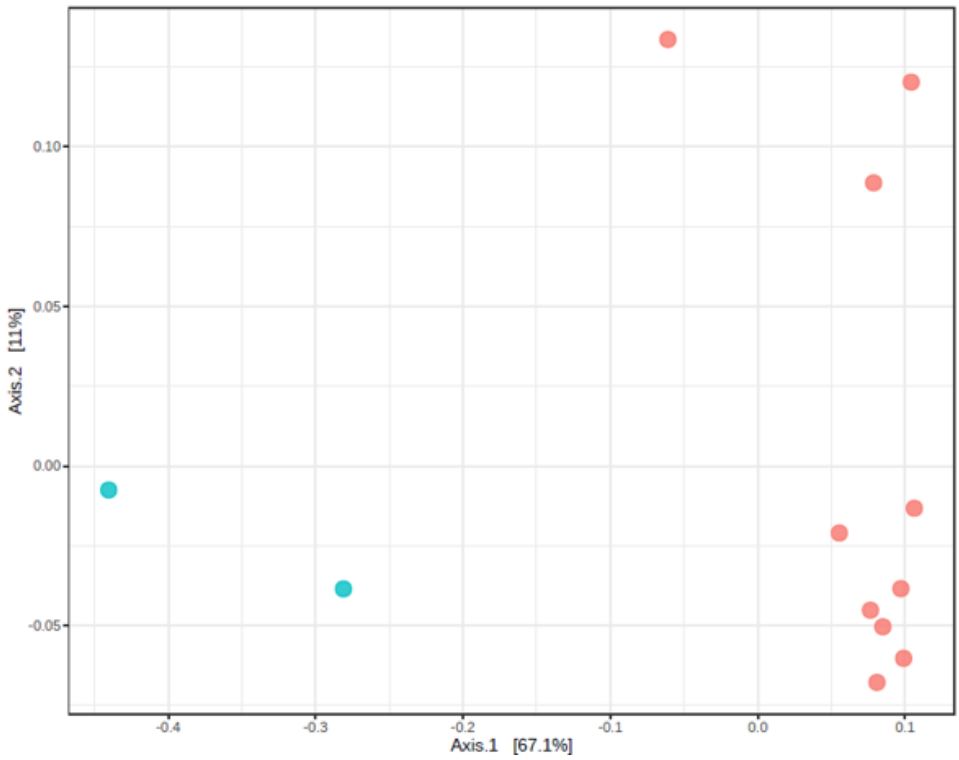
Figure 5:



5: α diversity as measured by Chao1 [p = 0.46], observed richness [p = 0.16], and Shannon diversity index [p = 0.25]. Non-significant α diversity highlights that the microbial ecological niche is the same amongst all ready to buy Toombak and is attributed to OTU feature domination. Shannon scores > 2 indicate an increased number of rare species.

Beta diversity, however, was statistically significant [$p < 0.025$] with an R^2 of 0.61. Samples from Khartoum North exhibited intra–group commonality compared to those from Omdurman [Figure 6].

Figure 6:



6: PCoA model highlighting β diversity of ready to buy Toombak in relation to place of purchase at feature level [OTUs]. Pink=Khartoum_North, Blue=Omdurman. Predominant inter-group microbiome variation of Toombak samples [dependant variable] due to location of purchase [independent variable]. Samples from Khartoum-North exhibit intra-group commonality compared to those from Omdurman. This variation is statistically significant [<0.025]. R2= 0.61, thus 61% of the microbiome in Toombak ready to buy samples can be related to place of purchase most likely affected by multi-local handlers and the social and geographical considerations that are unique between Khartoum-North and Omdurman.

Utilising BLASTn, OTUs were interrogated at species level in the ready to buy Toombak samples, for those sequences > 98 % identity to the NCBI database of rRNA bacterial and archaeic species identification. This revealed that *Corynebacterium*

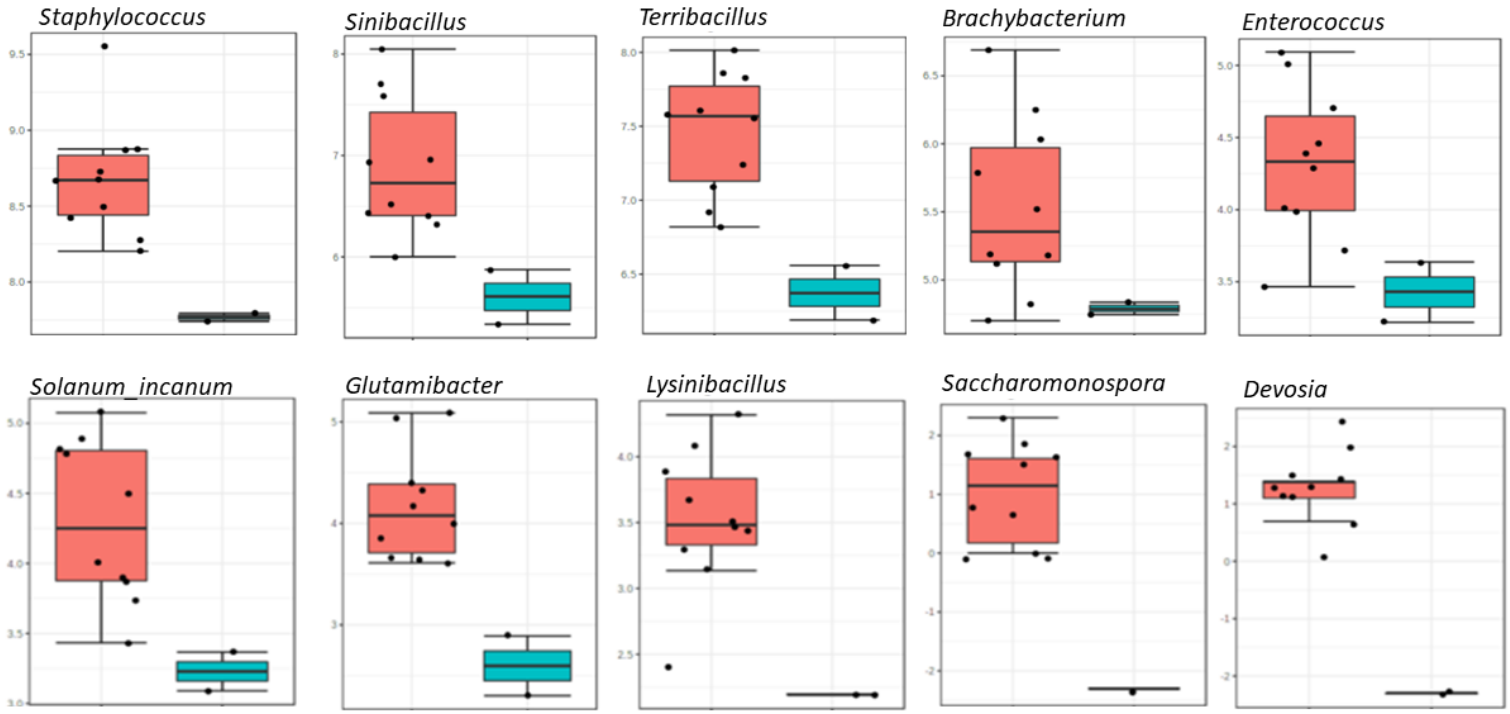
casei, [98.89 % identity][q=1.6417e-5], *Atopococcus tabaci* [99.34 % identity] [p=0.05], , *Atopostipes suicloacalis* [98.49 % identity], *Oceanobacillus chironomi* [98.49 % identity], *Staphylococcus gallinarum* [99.35 % identity] [p=0.01] , *Enteractinococcus lamae* [98.45 % identity] and *Facklamia tabacinasalis* [99.35 % identity] were some of the most represented species amongst the ready to buy Toombak samples.

Furthermore, the species *Corynebacterium casei* and the genus *Lentibacillus* were highly correlated with Toombak samples from Khartoum North, with correlation scores of 0.94 and 0.83, respectively. The species, *Atopostipes suicloacalis* and genus *Alloiococcus* were positively correlated with Omdurman samples with correlation scores of 0.93 and 0.71, respectively.

Classic univariate analysis [Anova] highlighted significant variations in genera abundance between the two Toombak ready to buy groups. Figure 7 highlights the most enriched genera amongst Khartoum North samples which included *Staphylococcus* [q = 0.002], *Sinibacillus* [q = 0.028], *Terribacillus* [q = 0.028], *Brachybacterium* [q = 0.046], *Enterococcus* [q = 0.032], *Glutamibacter* [q = 0.025], *Lysinibacillus* [q = 0.021], *Saccharomonospora* [q = 0.025] and *Devosia* [q = 0.004]. Shrub addition [*Solanum_incanum*] in the final ready to buy Toombak from Khartoum North was also apparent, most likely incorporated during the early production stages [q = 0.006].

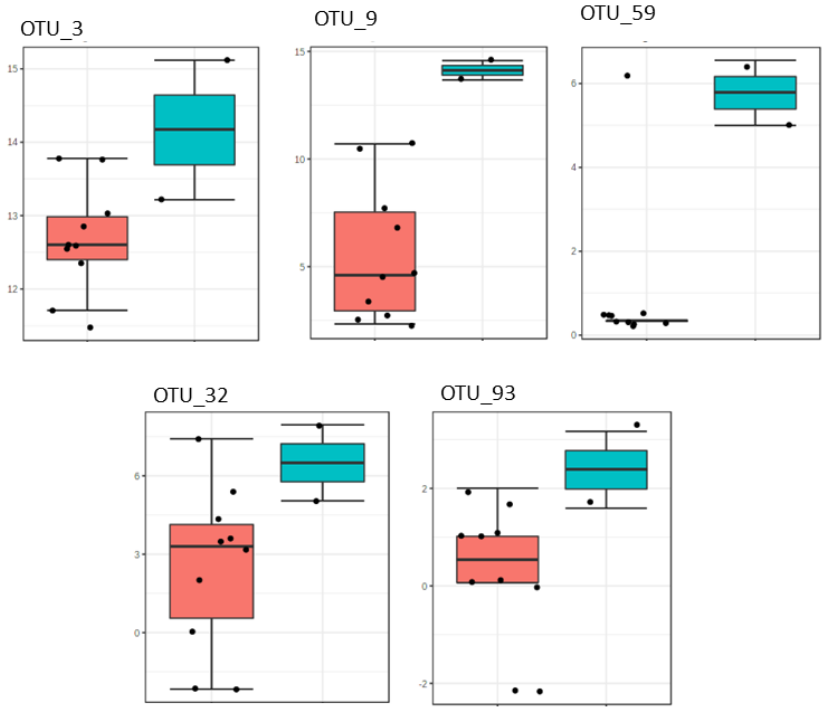
In ready to buy Toombak, it was found that *Staphylococcus gallinarum, Devosia albogilva* [99.77 % identity], *Enteractinococcus lamae*, *Lysinibacillus telephonicus*, [99.35 % identity], *Oceanobacillus chironomi*, *Saccharomonsospora viridis* [99.33 % identity], *Glutamibacter mysorens* [99% identity], *Terribacillus saccharophilus* [99.78 % identity] and *Terribacillus halophilus* [99.14 % identity], were significantly enriched in Khartoum North samples. DeSeq2 was further employed to assess differentially abundant OTUs particularly amongst Omdurman samples. The genus *Alloiococcus* and the three species *Atopostipes suicloacalis, Alkalibacterium iburiense* [99.57 % identity] and *Serratia quinivorans* [99.14 % identity] were increased in those samples from Omdurman [Figure 8].

Figure 7:


Staphylococcus
Sinibacillus
Terribacillus
Brachybacterium
Enterococcus
Solanum_incanum
Glutamibacter
Lysinibacillus
Saccharomonospora
Devosia

7: Pink=Khartoum-North, blue=Omdurman. Classic univariate analysis [Anova], log scaled, highlights significant variation in microbiome between the two Toombak ready to buy groups, FDR adjusted p values or q values. Staphylococcus [q=0.002], Sinibacillus [q=0.028], Terribacillus [q=0.028], Brachybacterium [q=0.046], Enterococcus [q=0.032], Glutamibacter [q=0.025], Lysinibacillus [q=0.021], Saccharomonospora [q=0.025] and Devosia [q=0.004] were the most abundant genera in Toombak obtained from Khartoum-North. Shrub contamination [Solanum_incanum] in the final ready to buy Toombak from Khartoum-North was also found [q=0.006] and reflected the varied production techniques of Toombak.

Figure 8:

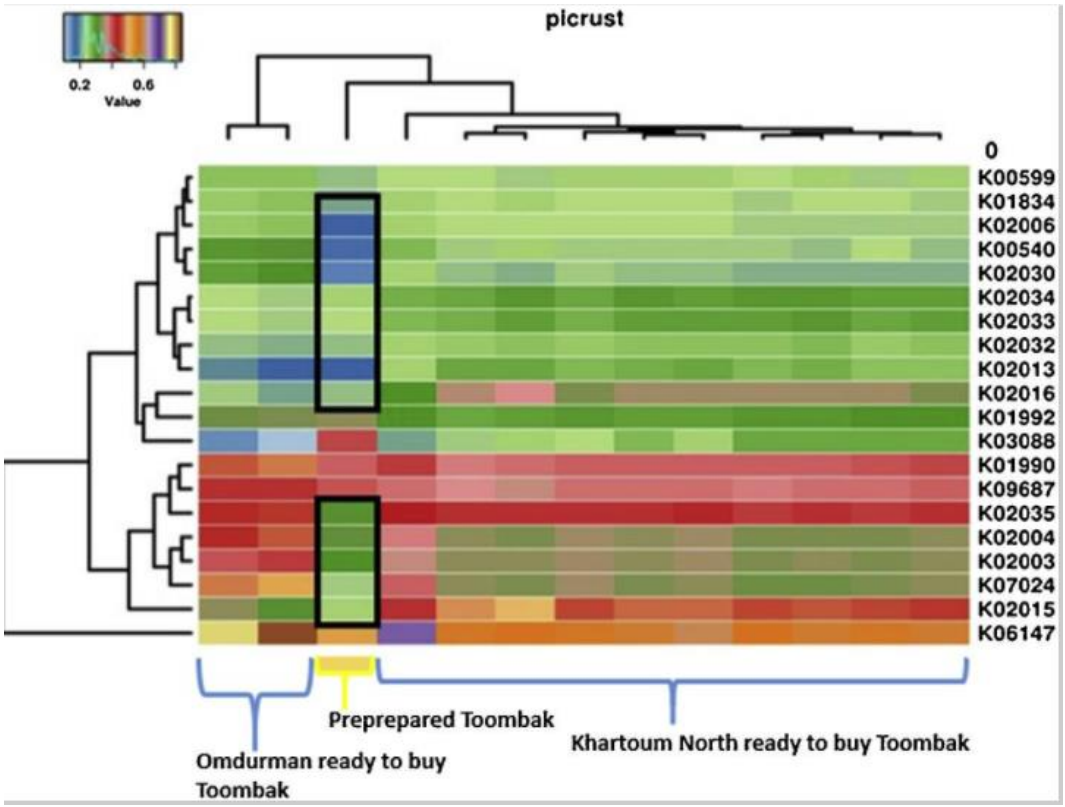


8: DESeq2 was used to assess differentially abundant OTUs amongst Omdurman samples. Pink =Khartoum-North, blue = Omdurman. DeSeq2 determines the mean expression of various shrinkage estimates and fold changes to improve final interpretability, while also correcting dispersion estimates that may be too low through modelling or under sampling. OTU_3 was the most significantly aligned species utilising BLASTn and was found to be 98.49% identical to Atopostipes suicloacalis. Omdurman samples also contained increased abundances of Alkalibacterium iburiense, Serratia quinivorans and the genus Alloiococcus.

PICRUSt assimilations were determined using the Kyoto Encyclopaedia for Genes and Genomes [KEGG] pathways to assess activity for metagenome functions amongst ready to buy samples and pre-prepared Toombak and profiled using a heatmap [Figure 9]. Ready to buy Toombak samples had an abundance of functions in the ATP binding proteins, metal transport mechanisms that included cobalt/nickel transport [K02006], iron complex transports [K02013 and K02015], and peptide/nickel transport system substrate binding protein [K02035]. The ATP binding protein antibiotic transport system [K09687] was highly expressed in all samples while the F420H[2] dependent quinone reductase [K00540] pathway was present amongst the ready to buy samples.

Signalling and cellular process pathways in Toombak included the polar amino acid transport system substrate binding protein [K02030], putative ABC transport system permease proteins [K02003 and K02004], found to be low in the pre-prepared sample and the ABC 2 type transport system [K01990], found to be high in all samples. Finally, sucrose 6 phosphate [K07024], associated with starch and sucrose metabolism was reduced in the pre-prepared form.

Figure 9:



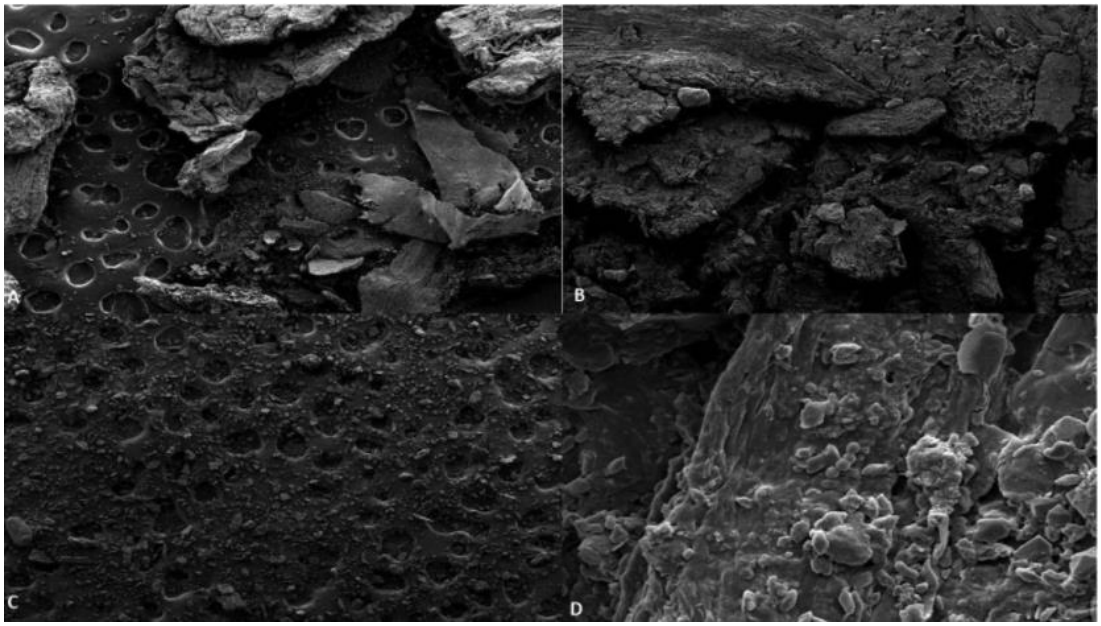
9: PICRUSt data assimilation of the top 20 KEGG pathways in pre-prepared and ready to buy Toombak with visualisation of abundance activity using heatmap profiling. Increased peptide/nickel [K02035], cobalt/ nickel [K02006], and iron transport systems [K02013 and K02015] amongst ready to buy Toombak as well as active antibiotic transport systems [K09687] amongst all samples were identified.

Microscopic analysis

At low magnification [35x], the large non-homogenous and irregular particles were highlighted in the Toombak samples [Figures 10A and 10B] compared to a pasteurised regulated European tobacco powder with a uniform structure [Figure 10C]. A ready to

buy sample of Toombak at 4000x magnification yielded evidence of bacterial cocci and rod contamination [Figure 10D].

Figure 10:



10A: pre-prepared Toombak before handmade preparation with large, wooden and leaf like comonents at 35X magnification. 10B: ready to buy Toombak highlighting a non-homogenous large particle size product that is more closely packed together and may be quite abrasive to the oral mucosa [35X]. 10C: evenly dispersed European snuff particles at 35X and 10D: ready to buy Toombak with abundance of bacterial rods and cocci at 4000x magnification.

Energy dispersive X ray Spectroscopy

Elemental composition was analysed amongst 26 spectra within the Toombak samples. Carbon, oxygen, magnesium, aluminium, silicon, sulphur, chlorine, calcium, molybdenum, beryllium, sodium, and potassium were all found in the samples and were calculated in weight %. Oxygen and carbon were the most predominant elements in Toombak and European standard, however, calcium, potassium, sodium, and silicon were also present.

While the pre-prepared Toombak sample had only a maximum of 0.55 % sodium, ready to buy Toombak contained a maximum of 4.93 %. Potassium was increased in the pre-prepared form [max 3.17 %], while a high percentage of calcium was present in the ready to buy sample [max 5.67 %]. Table 4 summarises the main elements in

Toombak samples compared to the control. While the weight % of elements in Toombak were more variable and of increased total weight % [max 0.44 %–5.67 %], the control sample exhibited consistent levels and a lower total weight % of the most common six elements [max 0.21 %–2.52 %].

Table 4: The main elements found in the pre–prepared Toombak and one ready to buy sample compared to a control [European smokeless tobacco] using energy dispersive X ray spectroscopy.

	Pre-prepared Toombak		Ready to buy Toombak		Control	
Weight %	**Min %**	**Max %**	**Min %**	**Max %**	**Min %**	**Max %**
Elements						
Calcium	0.56 %	1.08 %	0.36 %	5.67 %	0.98 %	2.52 %
Potassium	0.98 %	3.17 %	0.40 %	2.23 %	1.06 %	2.11 %
Aluminium	0.50 %	2.03 %	0.55 %	0.71 %	0.14 %	0.21 %
Sodium	0.21 %	0.55 %	2.08 %	4.93 %	1.13 %	2.27 %
Magnesium	0.20 %	0.90 %	0.24 %	0.44 %	0.28 %	0.72 %
Silicon	0.57 %	1.99 %	0.80 %	2.09 %	0.20 %	0.74 %

pH of the original sample was 8.73, while pH of the ready to buy samples ranged from 9.08–9.92.

2.5 Discussion

Compared to a European regulated product, Toombak was a course, non-homogenous smokeless tobacco containing visible bacterial cocci and rods under microscopy. SEM-EDX showed high elemental sodium in the ready to buy Toombak samples, but other elements could also be found. All the ready to buy samples were at pH 9 or above attributed to the addition of sodium bicarbonate, primarily to enhance nicotine bioavailability.

The pre-prepared and ready to buy Toombak samples had similar heavy metal concentrations, of which there were four cations classified as group 1 carcinogens, four cations classified as group 2 carcinogens, and one cation classified as a group 3 carcinogen in the samples. Six of the nine cation concentrations were available for comparison with the GothiaTek standard and were found to be of acceptable values. The remaining elements, cobalt, copper and iron were compared with other references [[25], [26], [27]] [Table 1]. Arsenic, cadmium, lead, and mercury were found to be of acceptable values in the Toombak. *Nicotiana Rustica* has been previously shown to accumulate sulphur which may support arsenic and cadmium detoxification [28], however nickel levels were closer to the upper range of acceptable limits.

Chromium, cobalt, and copper were high in the pre-prepared form while iron was by far the most abundant heavy metal in Toombak, at 1270 mg/kg in the ready to buy sample. Therefore, an average consumer using 90 g of Toombak [6 g of Toombak X 15 uses per day] would be exposed to an average of 114.3 mg of iron a day, while the reported acceptable dietary intake of iron is 8−18 mg per day [26].

Such levels of iron in Toombak are higher than the iron content in home-brewed beer [48−62 mg/l] prepared in south and central Africa in iron pots, the consumption of which has been associated with the condition Africa iron overload/Bantu siderosis. Manifestations of this disease include liver damage, diabetes, osteoporosis and cardiac abnormalities [29]. Interestingly, Toombak is stored and prepared for long periods in stainless steel pots containing at least 50 % iron and whether this undertaking has a

direct effect on the iron concentration found in Toombak requires further investigation. Other sources of heavy metals include the use of pesticides such as fungicides and insecticides, fertilisers, sludge irrigation with polluted water and settled aerosols [30].

True bioavailability of heavy metals in Toombak or in any other smokeless tobacco, can be difficult to accurately determine. Bioavailability may be as low as 6% and as high as 100 % in some forms of smokeless tobacco [31]. Metals such as cadmium, cobalt and nickel have been proven to be quite efficient at being extracted by artificial saliva and thus could potentially pose greater harm to Toombak users [25]. The oral, oesophageal and gut linings can allow for absorption of cations but biological damage depends on the total concentration of the metal in the smokeless tobacco, the proximal transfer by epithelial contact, saliva and digestive juices as well as the individual health status [32]. Heavy metals also exhibit the property of bioaccumulation in the body over time, and thus a chronic Toombak user may be at risk of long-term chronic disease or side effects associated with heavy metal exposure.

Cadmium exposure in humans has been associated with respiratory, skeletal [osteomalacia and osteoporosis], renal and cardiovascular disease. Precancerous lesions, oral ulceration and progressive fibrosis of the oral tissues as well as liver damage have been associated with copper exposure [33] while the development of leukaemia, cancers of the bladder, kidney, and skin, neurological disturbances and renal failure are associated with arsenic toxicity. Neurological and haemoglobin synthesis disturbances can occur from mercury exposure. Users of Toombak are at particular risk of neurotoxic effects due to starting the product at a young age [34].

A diverse range of metabolomic compounds were isolated from Toombak samples with variations between the pre-prepared and ready to buy Toombak [Figure 2]. This highlights that the different stages of Toombak processing and storage can create a significant change to the final product. Formaldehyde, acetaldehyde and propionaldehyde were all detected in the Toombak samples. Only two of the 21 Toombak samples were found to be below the recommended level set out by the GothiaTek standards, acceptable for formaldehyde while acetaldehyde levels were 20–52 times higher in all the Toombak samples [Table 2]. Aldehydes are mutagenic and

carcinogenic and can be found naturally in smokeless tobacco or may be added. They are formed by the incomplete degradation of tobacco leaves as well as being a metabolic end-product of many microorganisms that are numerous in Toombak. Formaldehyde is a group 1 IARC carcinogen while acetaldehyde is now classed as a group 2B carcinogen. Formaldehyde has been associated with cancers of the nasopharynx, nasal sinuses as well as the oral cavity, brain, pancreas and non-solid cancers which include leukaemia, and multiple myeloma [35]. Acetaldehyde can affect both the oesophageal and gut mucosa [36] while animal models have shown that cardiovascular tissues are highly sensitive to aldehydes [37]. Aldehydes have also been shown to cause direct DNA damage in the buccal cells of cigarette smokers [38] and can activate the Jun/AP-1 pathway in oral keratinocytes leading to the transcription of oncogenes [39].

It is unclear if aldehydes are added to Toombak as a preservative or a flavouring as is done with some other forms of smokeless tobacco [40]. Most worryingly, acetaldehyde in itself is potently addictive, and can act in synergy with nicotine to increase the addictive potential of the latter [41].

Propionaldehyde was found in all Toombak samples. It is more likely introduced from drinking water during its preparation but may also be added. It is used in the manufacture of plastics and utilised as a disinfectant and preservative. In small amounts [140 μg/day], both the FDA and WHO have approved propionaldehyde as a synthetic flavouring, known to possess a fruity odour. Propionaldehyde has been shown however to increase arterial blood pressure and heart rate by sympathomimetic activities [42].

Nicotine levels are high in the *Nicotiana Rustica* plant as opposed to other Nicotiana species due to an additionally expressed putrescine methyltransferase gene which allows for its efficient production, while nicotine transport from the root to the shoot is more efficient [28]. Nicotine levels in the ready to buy Toombak samples in this study ranged between 16–25 mg/g [average 20.5 mg/g] but were highest in the pre-prepared form [31 mg/g] [43] [Table 3]. Thus, in a Toombak user, utilising 90 g of the ready to buy product a day [6 g × 15 times/day], nicotine exposure would be 1440 – 2250 mg/g per day [average 1845 mg/g]. Nicotine levels in Toombak in the literature

are highly variable and range from 5.16 to 102.4 mg/g [44]. It is absorbed by passive diffusion into the blood, where the more alkaline the pH, the more unionised nicotine that becomes available and can be easily absorbed by epithelial linings [45]. Addition of water to Toombak also improves nicotine absorption and permeability through the oral mucosa as well as better transfer to saliva. Short term effects of nicotine exposure from smokeless tobacco include increased heart-rate [within the first 15 min], nausea and vomiting, while long term effects include the development of cardiovascular disease such as atherosclerosis [46] through increased low density lipoproteins and decreased high density lipoproteins [47]. Nicotine has also been associated with delayed wound healing, prevention of apoptosis of cancer cells and the promotion of reactive oxygen species and oxidative damage to cells [48].

Mean levels of the individual TSNAs amongst ready to buy Toombak samples were all found to be high [NAB; 10-40 mg/kg, NAT; 5-14 mg/kg, NNK; 1200-4700 mg/kg and NNN; 60-130 mg/kg]. NNN and NNK levels in Toombak were compared to the maximum limits of these compounds in Swedish snus [NNN + NNK =0.95 mg/kg]. Concentrations of NNN and NNK in ready to buy Toombak ranged from 1270−4830 mg/kg, values thousands fold higher than the maximum limits of these compounds in Swedish snus [Table 3], a trend also reported by Idris et al. [1991] [49]. Mean levels of NNN and NNK in Indian forms of smokeless tobacco are reported to be 22.9 mg/kg and 2.6 mg/kg, respectively [50], while it has been proposed by the FDA that NNN levels in smokeless tobacco sold in the USA should not exceed 1 mg/kg [51]. Thus, TSNA concentrations in Toombak may be the highest found amongst all smokeless tobacco products in the world. These are carcinogenic compounds produced when nitrate is converted to nitrite by both aerobic and anaerobic metabolic routes that involve nitrate or nitrite reductase as the two key enzymes. This is followed by nitrosation which involves a chemical reaction between nitrite and nicotine [52,53].

TSNAs are absent in the green leafy tobacco plant and their accumulation occurs during the post cultivation period of Toombak production, harvesting, fermentation, storage and even after purchase. TSNA formation is also closely related to the triad of nicotine, moisture content and elevated temperatures. Moist forms of smokeless tobacco with added water such as Toombak, promote bacterial contamination and their growth [54]. The high environmental temperatures then allow for the chemical

reduction of nitrate to nitrite, where the latter is originally elevated in Toombak due to the high nicotinic alkaloid content in the *Nicotiana Rustica* plant.

The metabolism and activation of NNN and NNK in particular, lead to the pyridyloxobutylation of DNA, where in the setting of proto-oncogenes and tumour suppressor genes, can cause activation and inactivation of these critical genes, respectively, and thus tumour growth [55,56]. Furthermore, NNN and NNK are thought to act in a synergistic manner, binding to nicotinic acetylcholine receptors, further facilitating tumour development [57]. NNN is a powerful oral and oesophageal carcinogen in laboratory animals as well as in humans [58, 59], and together, NNN and NNK have been strongly linked with the development of oral cancer when administered to rats [60]. Another consideration is that the enzymatic activity of saliva facilitates the release and extraction of TSNAs in the oral cavity [61] and the TSNA content in saliva of Toombak users has been shown to be elevated [62]. The NNK specific metabolite, 4-[methylnitrosamino]-1-[3-pyridyl]-1-butanol [NNAL], excreted in urine is highly elevated amongst Toombak users [63] and is supported by the findings of the largely increased TSNA content reported in this paper.

Adenine is a purine nucleobase that was found in high amounts in the pre–prepared sample and at lower levels in the ready to buy Toombak, most likely reflecting lack of its degradation at this stage. Hypoxanthine and inosine were found at higher concentrations in the ready to buy samples compared to the pre-prepared form, suggesting the deamination of adenine and purine metabolism that are a natural metabolic process of plant metabolism. Amino acids in Toombak accumulate in parallel with the curing stage, where breakdown of tobacco proteins and hydrolysis of pigments occur by enzymes and microorganisms releasing such amino acids [64]. Citrulline was elevated in the pre-prepared sample and lower in the ready to buy Toombak and is derived from the catabolism of arginine by lactic acid bacteria as well as being a nitrogenous precursor of urethane, a group 2A IARC carcinogen. No urethane was detected in any of the Toombak samples. Urethane content has been poorly linked to the process of fermentation where its formation is associated with increased temperatures during pasteurisation and presence of added ethanol in smokeless tobacco, however these features are not part of Toombak preparation [65]. Carnitine presence in Toombak may provide bacteria with thermal and osmotic

protection [66] promoting their survival but its presence was found to be highly variable. Proline, found in all samples in Toombak, may induce yeast to hyphal form in the oral cavity and promote their pathogenic capability [67] while alanine and valine are not known to cause harm.

Benzylaldehyde and benzyl alcohol, aroma compounds in *Nicotiana Rustica*, were highest in the pre-prepared sample, but variable in the ready to buy Toombak and contribute to the tobacco flavour [68, 69]. Phenylethyl alcohol, which has a rose odour, is found naturally amongst various human nutritional consumables [70] but is also known to be added to Gutka, a type of smokeless tobacco used in India [71]. All samples of Toombak contained phenylethyl alcohol, however, levels were variable where six of the 21 samples had high concentrations including the pre–prepared Toombak, while the remaining samples had reduced or negligible amounts. This may suggest that this compound occurs both naturally and is added during the manufacturing process. Phenylethyl alcohol may have some selective inhibitory action on gram negative bacteria [72] thus may be a contributing factor to Toombak microorganism composition. The flavour of Toombak may also arise from the presence of proline and hexose which contribute to the Maillard reaction in tobacco, a non-enzymatic reaction that leads to the development of flavour compounds [73].

Bacteria can transform glucose and other sugars into succinic acid during fermentation [74] while *Lactobacillus* and *Bacillus* strains can produce succinic acid from citric acid in particular [75,76]. Citric acid is a prominent component of *Nicotiana Rustica* after curing while Toombak contained abundant *Bacillus* [68]. Interestingly, *Corynebacterium* and *Enterococcus* species are some of the main studied succinic acid producers, both found in Toombak [77]. Succinic acid contributes to the taste of Toombak, by adding a bitter salt-like flavour and is also high in some fermented beverages [e.g sake from Japan] but is not known to harm human health. It also has many other uses in industry and is either produced chemically or through fermentation by fungi and bacteria. 3 – Methyl – valeric acid, found in Toombak, has a strong, pungent, fruity and cheesy flavour [78] and is formed from the alkaline hydrolysis of glycolipids found in Nicotiana species including *Nicotiana Rustica* [79]. Pantothenic acid [vitamin B5] most likely originates from the plant structure with no harmful addition to users. Hexose [glucose and fructose] was

elevated only in the pre-prepared sample and was considerably reduced in all the ready to buy Toombak samples [Figure 3], a finding supported by the air curing process of Toombak, which contributes to the low sugar levels. Hexose may also contribute to the taste of Toombak. Low sugar content may however, allow bacterial growth and prevent acid metabolites from occurring, thus keeping Toombak at alkaline pH [80,81]. At least, in part, the low sugar content of Toombak should not contribute to the risk of caries development from extrinsic sugars, amongst Toombak users.

In total, 13 phyla, 12 classes, 47 orders, 85 families and 134 genera were characterised in Toombak of which *Firmicutes* and *Actinobacteria* were the predominant phyla amongst ready to buy and pre-prepared forms, although in the latter, *Actinobacteria* were markedly reduced while *Cyanobacteria* were increased. *Proteobacteria* were the least abundant in all forms of Toombak. *Proteobacteria* are more adapted to, and thus are generally found in cold environments, explaining their low abundance in Toombak obtained from Sudan, a country of a hot and temperate climate [82]. *Bacillaceae* family were enriched in the pre-prepared form while *Virgibacillus* was the most abundant genus [Figures 4A, 4B and 4C].

The microbiome of Toombak has been previously reported by three authors [ElHebshi, Smyth and Tyx et al.] with the studies reporting on one to five Toombak samples. ElHebshi et al. [2017], analysed one Toombak sample where *Facklamia, Desemzia, Atopostipes, Lysinibacillus* and *Corynebacterium* were the most abundant genera found, while *Facklamia tabacinasalis, Atopostipes suicloacalis, Lysinibacillus chungkukjangi* and *Desemzia incerta* predominated in the Toombak sample [83]. Tyx et al. [2016], analysed two Toombak samples and found a high relative abundance of *Corynebacteriaceae* and *Staphylococcaceae* and a low α diversity [84], while Smyth et al. [2017], analysed five Toombak samples and one alkalinising agent and found *Enteractinococcus* and *Corynebacterium* to be significantly increased [85]. Our results are aligned with some of these previous data, in particular that *Corynebacterium* is one of the most abundant genus in Toombak. This study is the first to report Toombak microbiome composition with respect to inter-locational variations, Khartoum North and Omdurman. Although *Corynebacterium* are enriched amongst Toombak including the species *Corynebacterium casei*, they are not the

dominant genus or species, respectively. *Atopostipes suicloacalis, Atopococcus tabaci, Oceanobacillus chironomi, Staphylococcus gallinarum, Enteractinococcus lamae, Facklamia tabacinasalis* as well as *Alloiococcus* genera also predominantly defined the final composition of Toombak. Furthermore, *Staphylococcus cohnii* was found to be highly abundant amongst Toombak samples analysed by Smyth et al. [2017] [85], while *Staphylococcus gallinarum* was the most abundant species in our samples; known to be highly salt tolerant [86].

We found a high Shannon diversity index score [>1] indicating that Toombak has an increased number of microorganisms, particularly rare species. This trend is similar amongst samples since non-significant findings were obtained when comparing both ready to buy groups. Chao1 was found to be similar between the two ready to buy groups due to certain OTUs dominating in the samples [Figure 5]. This reflects the natural plant composition and alkalinic niche of Toombak, as confirmed by our pH findings [87]. Therefore, abundant bacteria in Toombak can survive alkalinity or even prefer it, such as the *Bacilli* class. These likely become predominant in the setting of elevated temperatures during Toombak production [fermentation and storage]. Their ability to produce endospores allows them to be highly resistant to harsh conditions while *Bacilli* can also drive alkalinity and improve nicotine absorption, contributing to the addictive nature of Toombak. *Bacilli* isolated from American forms of smokeless tobacco have been shown to injure the oral mucosa of hamster cheek via the kallikrein/kinin metabolic pathway and thus may directly impact on chronic inflammation in the mouth [88].

The alkalinic nature of Toombak is further appreciated with the presence of alkaphiles such as *Oceanobacillus chironomi*, a facultative alkaphile and *Alkalibacterium* known to possess superior cellulose hydrolytic and carbohydrate fermentation properties [89], while *Alkalibacterium iburiense* identified in increased abundance amongst Omdurman samples is an obligate alkaphile. This allows for a self-sustainable fermentation system within Toombak over long periods and past the original air-curing period [90]. Furthermore, the contribution of bacteria to the potently elevated TSNA levels in Toombak are an important addition to the role of the microbiome. *Corynebacterium* and *Staphylococcaceae* species are known to be

potent nitrate reducing bacteria as well as efficient nitrogen metabolisers [91], and may contribute to the high TSNA content of Toombak.

This study is the first to assess β diversity amongst numerous Toombak samples, particularly the socio – geographical dynamics which was found to significantly affect [$p < 0.025$] the final composition of the ready to buy groups [Figure 6]. The microbiome of the ready to buy Toombak is affected by the preparatory techniques, continuous unsterile human handling of the product and poor storage qualities, that allow for microbiome changes that may yet remain underestimated in the implications for human health. Furthermore, Toombak contained several animal and insect microbes including *Enteractinoccocus lamae* isolated from animal faeces [92], *Staphylococcus gallinarum* first isolated from chickens [93] and *Oceanobacillus chironomi*, isolated from chironomid egg mass [94]. *Lysinibacillus telephonicus* has been isolated from the screen of a mobile phone [95]. *Alloiococcus* is regarded as a commensal of the external auditory meatus, nasopharynx and nasal cavity but is also associated with development of otitis media [[96], [97], [98]] and was found to be abundant in samples from Omdurman. Health concerns arise from long term Toombak use due to the potential gateway for entry of novel infection causing bacteria as well as the presence of antibiotic resistant genes amongst them. Metagenomic pathways of ready to buy samples were active for metal transport systems particularly peptide/nickel, cobalt/nickel and iron transport compared to the pre-prepared form. These active transport systems may reflect the increased metal bioavailability in Toombak and active metal transport systems have been associated with carcinogenic smokeless tobacco products [83]. The K09687 pathway, expressed in all samples is an antibiotic transport system while F420H[2], a dependant quinone reductase pathway found to be active in the ready to buy samples may protect against bactericidal agents [99] [Figure 9]. Further analysis is required through whole genome sequencing and may be an important area in the future to better understand which species may relate to specific metagenomic pathways within Toombak that may have an impact on important diseases such as cancer and infection.

The physical structure of Toombak is a suitable niche for the survival of bacteria, fungi and the harbouring of other contaminants. The rough surface of Toombak can encourage chronic mucosal abrasion and epithelial cell damage and can promote

inflammation and entry of opportunistic pathogens into the body. Compared to a European sample of smokeless tobacco, Toombak had a non-homogenous consistency and high levels of a number of elements. The SEM-EDX also highlights the large, non-homogenous structure of Toombak [Figures 10A and 10B], compared to a regulated European sample [Figure 10C] and provides visual evidence of bacterial rods and cocci at high magnification [Figure 10D]. Oxygen and carbon were the most predominant organic elements in Toombak, however, calcium, potassium, sodium, silicon, and aluminium were also abundant. Magnesium, beryllium, sulphur, and chlorine were less abundant within the samples. Sodium was present at 0.55 weight % in the pre-prepared form while it was 4.93 % in the ready to buy sample, suggesting that the addition of sodium bicarbonate may be the causative factor for this increase. Such elements can have hazardous local and systemic effects on users, particularly chronic abrasion as well as allergic and inflammatory predisposition to disease [100]. Sodium compounds may also lead to inflammation and cancer development [7].

All Toombak samples were in the alkaline pH range. The pre-prepared Toombak was of lower pH value, 8.73, compared to the remaining ready to buy samples that were above pH 9. The alkalinity of Toombak is a result of the cultural addition of sodium bicarbonate in order to convert the total nicotine to a free or unprotonated form which is more readily absorbed in the body. This allows Toombak to be highly addictive as nicotine becomes more potently available and is rapidly transported to the central nervous system. The addition of alkaline products such as ash, ammonium or calcium bicarbonate, slaked lime, and sodium citrate, can also be observed in other forms of smokeless tobacco.

Conclusion

In this study, we have shown that the Sudanese smokeless tobacco, Toombak, contains numerous potential biohazardous components. Chromium, cobalt and copper were found to be high in the pre-prepared sample while iron concentration was alarmingly high in both pre-prepared and ready to buy Toombak. This may be a hidden cause of disease that includes liver damage, diabetes, osteoporosis, and cardiac disease.

High levels of volatile aldehydes were also found in all the ready to buy Toombak samples. Aldehydes are both carcinogenic and addictive and may be produced naturally, added in Toombak or present as contaminants from water. High levels of heavy metals and aldehydes in Toombak are an alarming cause of concern. Furthermore, samples of Toombak were found to contain some of the highest TSNA content in smokeless tobacco in the world, most likely a contribution of many factors. These include the high content of nicotinic alkaloid in the *Nicotiana Rustica* plant, the elevated temperatures used during production, prolonged storage, and the high abundance of nitrate-reducing bacteria. TSNAs are among the major cancer-causing agents in Toombak, leading to mutational DNA damage of cells. Ready to buy Toombak was found not to contain a high extrinsic sugar content while flavour-related compounds are likely naturally developed in Toombak.

Toombak microbiome consisted of a core set of microbiota, consisting of the two main phyla *Actinobacteria* and *Firmicutes.* The most abundant genera within the ready to buy Toombak were *Corynebacterium_1, Atopostipes, Atopococcus, Oceanobacillus* and *Staphylococcus,* while *Virgibacillus* was the most abundant genus in the pre-prepared form. BLASTn revealed *Corynebacterium casei* as the most

abundant species in Toombak from Khartoum North samples, while *Atopostipes suicloacalis* was the most abundant species in Toombak from Omdurman. Other species included *Staphylococcus gallinarum, Atopococcus tabaci, Enteractinococcus lamae, Oceanobacillus chironomi* and *Facklamia tabacinasalis*. α Diversity was similiar amongst all Toombak samples [i.e., the same genera were present], however their distribution and relative abundance were variable, dependant on place of purchase as shown by a significant β diversity and DeSeq2 analysis. PICRUSt data showed metagenomic activity related to enriched peptide/nickel, cobalt/nickel and iron transport pathways in the ready to buy samples, an active antibiotic transport system and numerous active KEGG pathways related to survival in diverse environments.

This paper is the first to represent a wide scope analysis of numerous Toombak samples, a form of smokeless tobacco used extensively in Sudan incorporating metabolomic, metagenomic, microscopic, heavy metal and pH analysis. While smokeless tobacco sold in Europe and the USA is considerably regulated, Toombak is produced via individual manufacturing processes in Sudan. Use of the product is often initiated at a young age due to social difficulties and its low cost. Toombak harbours extensive biohazardous properties and it is prudent that urgent new and extensive measures be taken by the relevant government ministries in the Sudan to reassess its production laws as well as to take steps to reduce its use amongst the population.

2.6 References

1. Siddiqi K., Vidyasagaran A.L., Readshaw A., Croucher R. A policy perspective on the global use of smokeless tobacco. *Curr. Addict. Rep.* 2017; 4:503–510.

2. Idris A.M., Ibrahim Y.E., Warnakulasuriya K.A., Cooper D.J., Johnson N.W., Nilsen R. Toombak use and cigarette smoking in the Sudan: estimates of prevalence in the Nile state. *Prev. Med.* 1998; 27:597–603.

3. Roberts K. 2008. Nicotiana Sp .https://web.archive.org/web/20080906115744/ http://artsci.wustl.edu/ ~gjfritz/Nicotiana_sp.html. Accessed 07/06/2020.

4. Güven A., Köksal N., Büyükbeşe M.A., Cetinkaya A., Sökmen G., Aksu E., Cağlayan C.E. Effects of using a different kind of smokeless tobacco on cardiac parameters: "Maraş Powder" *Anadolu Kardiyol. Derg.* 2003; 3:230–235.

5. Stanfill S.B., Oliveira da Silva A.L., Lisko J.G., Lawler T.S., Kuklenyik P., Tyx R.E., Peuchen E.H., Richter P., Watson C.H. Comprehensive chemical characterization of Rapé tobacco products: nicotine, un-ionized nicotine, tobacco-specific N'-nitrosamines, polycyclic aromatic hydrocarbons, and flavor constituents. *Food Chem. Toxicol.* 2015; 82:50–58.

6. Idris A.M., Ibrahim S.O., Vasstrand E.N., Johannessen A.C., Lillehaug J.R., Magnusson B., Wallström M., Hirsch J.M., Nilsen R. The Swedish snus and the Sudanese toombak: are they different? *Oral Oncol.* 1998; 34:558–566.

7. Boffetta P., Hecht S., Gray N., Gupta P., Straif K. Smokeless tobacco and cancer. *Lancet Oncol.* 2008; 9:667–675.

8. Critchley J.A., Unal B. Health effects associated with smokeless tobacco: a systematic review. *Thorax.* 2003; 58:435–443. [

9. Sunanda P., Panda B., Dash C., Ray P.K., Padhy R.N., Routray P.K. Prevalence of abnormal spermatozoa in tobacco chewing sub-fertile males. *J. Hum. Reprod. Sci.* 2014; 7:136–142.

10. Elbeshir E.I., Abeen H.A., Idris A.M., Abbas K. Snuff dipping and oral cancer in Sudan: a retrospective study. *Br. J. Oral Maxillofac. Surg.* 1989; 27:243–248.

11. Patil S., Arakeri G., Alamir A.W.H., Patil S., Awan K.H., Baeshen H., Raj T., Fonseca F.P., Brennan P.A. Is toombak a risk factor for oral leukoplakia and oral squamous cell carcinoma? A systematic review. *J. Oral Pathol. Med.* 2020; 49:103–109.

12. Bao M., Joza P.J., Masters A., Rickert W.S. 2014. Analysis of Selected Carbonyl Compounds in Tobacco Samples by Using Pentafluorobenzylhydroxylamine Derivatization and Gas Chromatography-Mass Spectrometry; p. 86. 26.

13. Rutqvist L.E., Curvall M., Hassler T., Ringberger T., Wahlberg I. Swedish snus and the GothiaTek® standard. *Harm Reduct. J.* 2011; 8:11.

14. SwedishMatch . 2020. January 15th 2020, January 15th Gothiatek Standard.https://www.swedishmatch.com/Snus-and-health/ GOTHIATEK/ GOTHIATEK-standard/ Accessed.

15. Magoč T., Salzberg S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011; 27:2957–2963.

16. Edgar R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010; 26:2460–2461.

17. Caporaso J.G., Bittinger K., Bushman F.D., DeSantis T.Z., Andersen G.L., Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics.* 2010; 26:266–267.

18. Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., Glöckner F.O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013;41: D590–6.

19. McGinnis S., Madden T.L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* 2004;32: W20–5.

20. Langille M.G.I., Zaneveld J., Caporaso J.G., McDonald D., Knights D., Reyes J.A., Clemente J.C., Burkepile D.E., Vega Thurber R.L., Knight R., Beiko R.G., Huttenhower C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 2013; 31:814–821.

21. Dhariwal A., Chong J., Habib S., King I.L., Agellon L.B., Xia J. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* 2017;45: W180–W188.

22. Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15:550.

23. Dong M., Li L., Chen M., Kusalik A., Xu W. Predictive analysis methods for human microbiome data with application to Parkinson's disease. *PLoS One.* 2020;15

24. Broadhurst D., Goodacre R., Reinke S.N., Kuligowski J., Wilson I.D., Lewis M.R., Dunn W.B. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics.* 2018; 14:72.

25. Pappas R.S., Stanfill S.B., Watson C.H., Ashley D.L. Analysis of toxic metals in commercial moist snuff and Alaskan iqmik. *J. Anal. Toxicol.* 2008; 32:281–291.

26. Trumbo P., Yates A.A., Schlicker S., Poos M. Dietary reference intakes: vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc. *J. Am. Diet. Assoc.* 2001; 101:294–301.

27. Turnlund J.R. Copper nutriture, bioavailability, and the influence of dietary factors. *J. Am. Diet. Assoc.* 1988; 88:303–308.

28. Sierro N., Battey J.N.D., Bovet L., Liedschulte V., Ouadi S., Thomas J., Broye H., Laparra H., Vuarnoz A., Lang G., Goepfert S., Peitsch M.C., Ivanov N.V. The impact of genome evolution on the allotetraploid Nicotiana rustica – an intriguing story of enhanced alkaloid production. *BMC Genomics.* 2018; 19:855.

29. Walker A.R., Arvidsson U.B. Iron overload in the south african bantu. *Trans. R. Soc. Trop. Med. Hyg.* 1953; 47:536–548.

30. Dhaware D., Deshpande A., Khandekar R.N., Chowgule R. Determination of toxic metals in Indian smokeless tobacco products. *ScientificWorldJournal.* 2009; 9:1140–1147.

31. Prabhakar V., Jayakrishnan G., Nair S.V., Ranganathan B. Determination of trace metals, moisture, pH and assessment of potential toxicity of selected smokeless tobacco products. *Indian J. Pharm. Sci.* 2013; 75:262–269.

32. WHO study group on tobacco product regulation: report on the scientific basis of tobacco product regulation: fourth report of a WHO study group. [WHO Technical report series; 967].

33. Trivedy C., Meghji S., Warnakulasuriya K.A., Johnson N.W., Harris M. Copper stimulates human oral fibroblasts in vitro: a role in the pathogenesis of oral submucous fibrosis. *J. Oral Pathol. Med.* 2001; 30:465–470.

34. Jarup L. Hazards of heavy metal contamination. *Br. Med. Bull.* 2003; 68:167–182.

35. Formaldehyde, https://monographs.iarc.who.int/wp-content/ uploads/2018/06/ mono100F-29.pdf.

36. Seitz H.K., Stickel F. Acetaldehyde as an underestimated risk factor for cancer development: role of genetics in ethanol metabolism. *Genes Nutr.* 2010; 5:121–128.

37. Keith R.J., Fetterman J.L., Riggs D.W., O'Toole T., Nystoriak J.L., Holbrook M., Lorkiewicz P., Bhatnagar A., DeFilippis A.P., Hamburg N.M. Protocol to assess the impact of tobacco-induced volatile organic compounds on cardiovascular risk in a cross- sectional cohort: cardiovascular Injury due to Tobacco Use study. *BMJ Open.* 2018;8

38. M-w Weng, Lee H.-W., Park S.-H., Hu Y., Wang H.-T., Chen L.-C., Rom W.N., Huang W.C., Lepor H., Wu X.-R., Yang C.S., Tang M.-s. Aldehydes are the predominant forces inducing DNA damage and inhibiting DNA repair in tobacco smoke carcinogenesis. *Proc. Natl. Acad. Sci.* 2018;115: E6152–E6161.

39. Timmons S.R., Nwankwo J.O., Domann F.E. Acetaldehyde activates Jun/AP-1 expression and DNA binding activity in human oral keratinocytes. *Oral Oncol.* 2002; 38:281–290.

40. Kostygina G., Ling P.M. Tobacco industry use of flavourings to promote smokeless tobacco products. *Tob. Control.* 2016;25: ii40–ii49.

41. Belluzzi J.D., Wang R., Leslie F.M. Acetaldehyde enhances acquisition of nicotine self-administration in adolescent rats. *Neuropsychopharmacology.* 2005; 30:705–712.

42. Stanek J.G.S. 2008. Toxicological Review of Propionaldehydewww.epa.gov/iris . Washington, DC.

43. Sami A., Stanton C., Ross P., Ryan T., Elimairi I. Ultra- structure of Toombak; smokeless tobacco of Sudan and its effects on oral and systemic health. *Access Microbiology.* 2020:2.

44. Idris A.M., Prokopczyk B., Hoffmann D. Toombak: a major risk factor for cancer of the oral cavity in Sudan. *Prev. Med.* 1994; 23:832–839.

45. Singh I., Singh A., Kour R., Singh A., Singh R., Bali A. Is sodium carbonate in snuff a causative factor for oral mucosal lesions: a cross-sectional analysis. *J. Int. Soc. Prev. Community Dent.* 2018; 8:339–342.

46. Fant R.V., Henningfield J.E., Nelson R.A., Pickworth W.B. Pharmacokinetics and pharmacodynamics of moist snuff in humans. *Tob. Control.* 1999; 8:387–392.

47. Cluette-Brown J., Mulligan J., Doyle K., Hagan S., Osmolski T., Hojnacki J. Oral nicotine induces an atherogenic lipoprotein profile. *Proc. Soc. Exp. Biol. Med.* 1986; 182:409–413.

48. Campain J.A. Nicotine: potentially a multifunctional Carcinogen? *Toxicol. Sci.* 2004; 79:1–3.

49. Idris A.M., Nair J., Ohshima H., Friesen M., Brouet I., Faustman E.M., Bartsch H. Unusually high levels of carcinogenic tobacco-specific nitrosamines in Sudan snuff [toombak] *Carcinogenesis.* 1991; 12:1115–1118.

50. Warnakulasuriya S., Straif K. Carcinogenicity of smokeless tobacco: evidence from studies in humans & experimental animals. *Indian J. Med. Res.* 2018; 148:681–686.

51. Berman M.L., Hatsukami D.K. Reducing tobacco-related harm: FDA's proposed product standard for smokeless tobacco. *Tob. Control.* 2018; 27:352–354.

52. Mehra R., Mohanty V., Balappanavar A.Y., Kapoor S. Bacterial contamination of packaged smokeless tobacco sold in India. *Tob. Prev. Cessat.* 2020:6.

53. Priyanja J.K.M. Diversity study of nitrate reducing Bacteria from soil Samples-A metagenomics approach. *J. Comput. Sci. Syst. Biol.* 2015:8.

54. Han J., Sanad Y.M., Deck J., Sutherland J.B., Li Z., Walters M.J., Duran N., Holman M.R., Foley S.L. Bacterial populations associated with smokeless tobacco products. *Appl. Environ. Microbiol.* 2016; 82:6273–6283.

55. Bakdash A. Shammah [Smokeless tobacco] and public health. *Asian Pac. J. Cancer Prev.* 2017; 18:1183–1190.

56. Wang M., Cheng G., Sturla S.J., Shi Y., McIntee E.J., Villalta P.W., Upadhyaya P., Hecht S.S. Identification of adducts formed by pyridyloxobutylation of deoxyguanosine and DNA by 4-[acetoxymethylnitrosamino]-1-[3-pyridyl]-1-butanone, a chemically activated form of tobacco specific carcinogens. *Chem. Res. Toxicol.* 2003; 16:616–626.

57. Xue J., Yang S., Seng S. Mechanisms of Cancer induction by tobacco specific NNK and NNN. *Cancers [Basel]* 2014; 6:1138–1156.

58. Yuan J.M., Knezevich A.D., Wang R., Gao Y.T., Hecht S.S., Stepanov I. Urinary levels of the tobacco-specific carcinogen N'-nitrosonornicotine and its glucuronide are strongly associated with esophageal cancer risk in smokers. *Carcinogenesis.* 2011; 32:1366–1371.

59. Stepanov I., Sebero E., Wang R., Gao Y.T., Hecht S.S., Yuan J.M. Tobacco-specific N-nitrosamine exposures and cancer risk in the Shanghai Cohort Study: remarkable coherence with rat tumor sites. *Int. J. Cancer.* 2014; 134:2278–2283.

60. Hecht S.S., Rivenson A., Braley J., DiBello J., Adams J.D., Hoffmann D. Induction of oral cavity tumors in F344 rats by tobacco-specific nitrosamines and snuff. *Cancer Res.* 1986; 46:4162–4166.

61. Prokopczyk B., Wu M., Cox J.E., Hoffmann D. Bioavailability of tobacco-specific N-nitrosamines to the snuff dipper. *Carcinogenesis.* 1992; 13:863–866.

62. Idris A.M., Nair J., Friesen M., Ohshima H., Brouet I., Faustman E.M., Bartsch H. Carcinogenic tobacco-specific nitrosamines are present at unusually high levels in the saliva of oral snuff users in Sudan. *Carcinogenesis.* 1992; 13:1001–1005.

63. Murphy S.E., Carmella S.G., Idris A.M., Hoffmann D. Uptake and metabolism of carcinogenic levels of tobacco-specific nitrosamines by Sudanese snuff dippers. *Cancer Epidemiol. Biomarkers Prev.* 1994; 3:423–428.

64. Yang S.S., Smetena I. Determination of free amino acids in tobacco by HPLC with fluorescence detection and precolumn derivatization *Chromatographia.* 1993; 37:593–598.

65. McAdam K., Vas C., Kimpton H., Faizi A., Liu C., Porter A., Synnerdahl T., Karlsson P., Rodu B. Ethyl carbamate in Swedish and American smokeless tobacco products and some factors affecting its concentration. *Chem. Cent. J.* 2018; 12:86.

66. Meadows J.A., Wargo M.J. Carnitine in bacterial physiology and metabolism. *Microbiology.* 2015; 161:1161–1174.

67. Ali A.K.S. Tobacco exract induces yeast to Hyphal form transition in the human pathogen, Candida albicans. *Am J Clin Microbiol Antimicrob.* 2018; 1:1013.

68. Popova V., Ivanova T., Stoyanova A., Nikolova V., Docheva M., Hristeva T., Damyanova S., Nikolov N. Chemical constituents in leaves and aroma products of Nicotiana rustica L. Tobacco. *Int. J. Food Stud.* 2020:9.

69. LaVoie E.J., Tucciarone P., Kagan M., Adams J.D., Hoffmann D. Analyses of steam distillates and aqueous extracts of smokeless tobacco. *J. Agric. Food Chem.* 1989; 37:154–157.

70. Burdock G. CRC Press; Boca Raton: 2010. Fenaroli's Handbook of Flavor Ingredients.

71. Kellian C. Drugs in american society: an encyclopedia of history, politics, culture, and the law. *Ref. Rev. Inc. Aslib Book Guide.* 2015; 29:455.

72. Lilley B.D., Brewer J.H. The selective antibacterial action of phenylethyl alcohol. *J. Am. Pharm. Assoc.* 1953; 42:6–8.

73. Ali H., Pätzold R., Brueckner H. Determination of L- and D-amino acids in smokeless tobacco products and tobacco. *Food Chem.* 2006; 99:803–812.

74. Harrison R., Todd P., Rudge S., Petrides D. 2015. Bioseparations Science and Engineering.

75. Kaneuchi C., Seki M., Komagata K. Production of succinic Acid from citric Acid and related acids by lactobacillus strains. *Appl. Environ. Microbiol.* 1988; 54:3053–3056.

76. Bruce W.F. The decomposition of citric acid by bacillus aertrycke. *J. Biol. Chem.* 1934; 107:119–129.

77. Song H., Lee S.Y. Production of succinic acid by bacterial fermentation. *Enzyme Microb. Technol.* 2006; 39:352–361.

78. Bedoukian research. 2018. Technical Data Sheet - 3-METHYL PENTANOIC ACID.http://search.bedoukian.com/flavorfragrance/ff_product.asp?method=POP&id=477

79. Matsuzaki T., Shinozaki Y., Suhara S., Ninomiya M., Shigematsu H., Koiwai A. Isolation of glycolipids from the surface lipids of Nicotiana bigelovii and their distribution in Nicotiana species. *Agric. Biol. Chem.* 1989; 53:3079–3082.

80. Andersen R.A., Fleming P.D., Hamilton-Kemp T.R., Hildebrand D.F. pH Changes in smokeless tobaccos undergoing nitrosation during prolonged storage: effects of moisture, temperature, and duration. *J. Agric. Food Chem.* 1993; 41:968–972.

81. Banožić M., Jokic S., Ackar D., Blažić M., Šubarić D. Carbohydrates-key players in tobacco aroma formation and quality determination. *Molecules.* 2020; 25:1734.

82. Yadav A.N., Verma D.P., Kumar M., Pal K., Dey R., Gupta A., Padaria J., Gujar G., Kumar S., Suman A., Prasanna R., Saxena A. Diversity and phylogenetic profiling of niche-specific Bacilli from extreme environments of India. *Ann. Microbiol.* 2015; 65:611–629. doi: 10.1007/s13213.

83. Al-Hebshi N.N., Alharbi F.A., Mahri M., Chen T. Differences in the bacteriome of smokeless tobacco products with different oral carcinogenicity: compositional and predicted functional analysis. *Genes [Basel]* 2017:8.

84. Tyx R.E., Stanfill S.B., Keong L.M., Rivera A.J., Satten G.A., Watson C.H. Characterization of bacterial communities in selected smokeless tobacco products using 16S rDNA analysis. *PLoS One.* 2016;11

85. Smyth E.M., Kulkarni P., Claye E., Stanfill S., Tyx R., Maddox C., Mongodin E.F., Sapkota A.R. Smokeless tobacco products harbor diverse bacterial microbiota that differ across products and brands. *Appl. Microbiol. Biotechnol.* 2017; 101:5391–5403.

86. Jia Y., Niu C.-T., Lu Z.-M., Zhang X.-J., Chai L.-J., Shi J.-S., Xu Z.-H., Li Q., Elkins C.A. A bottom-up approach to develop a synthetic microbial community model: application for efficient reduced-salt broad bean paste fermentation. *Appl. Environ. Microbiol.* 2020;86: e00306–20.

87. Qian X.-B., Chen T., Xu Y.-P., Chen L., Sun F.-X., Lu M.-P., Liu Y.-X. A guide to human microbiome research: study design, sample collection, and bioinformatics analysis. *Chin. Med. J.* 2020:133.

88. Rubinstein I., Pedersen G.W. Bacillus species are present in chewing tobacco sold in the United States and evoke plasma exudation from the oral mucosa. *Clin. Diagn. Lab. Immunol.* 2002; 9:1057–1060.

89. Yumoto I., Hirota K., Nodasaka Y., Tokiwa Y., Nakajima K. Alkalibacterium indicireducens sp. nov., an obligate alkaliphile that reduces indigo dye. *Int. J. Syst. Evol. Microbiol.* 2008; 58:901–905.

90. Okamoto T., Aino K., Narihiro T., Matsuyama H., Yumoto I. Analysis of microbiota involved in the aged natural fermentation of indigo. *World J. Microbiol. Biotechnol.* 2017; 33:70.

91. Walter B., Hänssler E., Kalinowski J., Burkovski A. Nitrogen metabolism and nitrogen control in corynebacteria: variations of a common theme. *J. Mol. Microbiol. Biotechnol.* 2007; 12:131–138.

92. Chen X., Li G.-D., Li Q.-Y., Hu C.-J., Qiu S.-M., Jiang Y., Jiang C.-L., Han L., Huang X.-S. Enteractinococcus lamae sp. Nov. And Enteractinococcus viverrae sp. nov., isolated from animal faeces. *Antonie Van Leeuwenhoek.* 2015; 108:1477–1483.

93. Devriese L., Poutrel B., Kilpper-Bälz R., Schleifer K. Staphylococcus gallinarum and Staphylococcus caprae, two new species from animals. *Int. J. Syst. Evol. Microbiol.* 1983; 33:480–486.

94. Raats D., Halpern M. Oceanobacillus chironomi sp. nov., a halotolerant and facultatively alkaliphilic species isolated from a chironomid egg mass. *Int. J. Syst. Evol. Microbiol.* 2007; 57:255–259.

95. Rahi P., Kurli R., Khairnar M., Jagtap S., Pansare A.N., Dastager S.G., Shouche Y.S. Description of Lysinibacillus telephonicus sp. nov., isolated from the screen of a cellular phone. *Int. J. Syst. Evol. Microbiol.* 2017; 67:2289–2295.

96. Lappan R., Jamieson S.E., Peacock C.S. Reviewing the pathogenic potential of the otitis-associated Bacteria Alloiococcus otitidis and Turicella otitidis. *Front. Cell. Infect. Microbiol.* 2020:10.

97. Harimaya A., Takada R., Somekawa Y., Fujii N., Himi T. High frequency of Alloiococcus otitidis in the nasopharynx and in the middle ear cavity of otitis-prone children. *Int. J. Pediatr. Otorhinolaryngol.* 2006; 70:1009–1014.

98. Proctor D.M., Relman D.A. The landscape ecology and microbiota of the human nose, Mouth, and throat. *Cell Host Microbe.* 2017; 21:421–432.

99. Gurumurthy M., Rao M., Mukherjee T., Rao S.P.S., Boshoff H.I., Dick T., Barry C.E., 3rd, Manjunatha U.H. A novel F420-dependent antioxidant mechanism protects Mycobacterium tuberculosis against oxidative stress and bactericidal agents. *Mol. Microbiol.* 2013; 87:744–755.

100. Halstead M.M., Watson C.H., Pappas R.S. Electron microscopic analysis of surface inorganic substances on oral and combustible tobacco products. *J. Anal. Toxicol.* 2015; 39:698–701.

2 – [B] The Mycobiome of the Sudanese smokeless tobacco Toombak

Amel Sami 1,2, Imad Elimairi 2, *, C. Anthony Ryan 4, R. Paul Ross 1, and Catherine Stanton 1,3

1 APC Microbiome Ireland, School of Microbiology, University College Cork, Cork T12 YN60, Ireland.
2 Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, National Ribat University, Nile Street, Khartoum 1111, Sudan
3 Teagasc Food Research Centre, Moorepark, Fermoy, Cork P61 C996, Ireland
4 Department of Paediatrics and Child Health, University College Cork, Cork T12 DFK4, Ireland.

This chapter has been submitted to *Fungal Ecology* journal, Elsevier

2.1 Abstract

Background

Toombak is a smokeless tobacco produced and utilised in Sudan from *Nicotiana* tobacco plant species and often utilised by males in the oral cavity. The final product is a roughened dark brown non-homogenous tobacco with high pH and an abundance of bacterial genera including but not limited to, *Corynebacterium, Staphylococcus* and *Atopostipes*. The mycobiome or fungal composition of Toombak however has not been previously explored.

Materials and methods

Sixteen samples of ready to buy Toombak were purchased from two geographical locations in Sudan: Omdurman and Khartoum-North. Samples were analysed through internal transcriber [ITS] gene sequencing. DNA extraction was followed by amplicon sequencing and processing of reads by DADA2 in R. Fungal taxonomy was assigned to resulting Amplicon Sequence Variants [ASVs].

Statistical analysis included fungal relative abundance, core microbiome detection, species detection, Metagenomeseq [false detection rates], One-way Anova significant variation analysis and LEfSe plotting utilising Microbiome Analyst.

Results

Toombak was found to be dominated in abundance by the phylum *Ascomycota* [86-93%] and the genus *Aspergillus* [85.54%]. *Aspergillus heterocaryoticus* [43.27%] and *austwickii* [39.48%] were the most prevalent species in Toombak. Khartoum-North samples exhibited a higher alpha diversity with an increased number of positively correlated genera from this location compared to Omdurman. One-way Anova further distinguished Khartoum-North samples to be significantly increased in the genera *Aspergillus, Epicoccum, Alternaria* and *Thielavia*. Samples from Omdurman showed tobacco leaf inferiorities with blight associated *Itersonilia perplexans* and plant pathogen *Ramularia collo-cygni* while Khartoum-North samples were positively correlated with the phytotoxin producing genus, *Macrophomina*. Toombak from

Khartoum-North was further positively correlated with *Neotestudina*, one of the causative agents for the Sudanese endemic disease Eumycetoma.

Conclusion

Toombak has a rich mycobiome that is impacted by its open and unregulated processes of production. The fungal mycobiota in Toombak can be a source of alteration for normal mycobiome health and may relate to negative health consequences for long-term users. *Aspergillus* abundance in Toombak likely disorganises the normal oral microbiota flora and exposes users to a range of mycotoxins that may contribute to the increased risk of oral cancer development while Toombak users may also be at an increased risk of the development of fungal infections that include but are not limited to head and neck *Aspergillosis* and *Eumycetoma.*

2.2 Introduction

The abundance of fungi amongst smokeless tobacco products is varied in the literature with reports ranging from 1-30% [1, 2]. While several studies have been achieved on the microbiome and its related bacterial composition, there remains an inherent lack of current literature supporting the understanding of the fungal community in this form of tobacco [3] [4]. Nevertheless, several studies have delineated that the mycobiome of smokeless tobaccos from around the world [2, 5, 6] includes *Aspergillus, Pichia, Sterigmatomyces* and *Mortierella [7]* [5] as well as *Debaryomyces, Geotrichum, Mucor, Sepedomium, Penicillium* and *Trichophyton* [8]. The storing of smokeless tobacco may further promote a differing fungal genera that includes enrichment with *Xeromyces* and *Rhodotorula* [9]. The mycobiome of smokeless tobacco is further impacted by high temperatures, humidity, long-term storage, lack of sunlight and human and environmental contamination [10-12].

The mycobiome of smokeless tobacco likely plays an important role in the modulation of the normal human-mycobiome relationship through genomic, metabolic and proteomic processes that can have a substantial effect on host immune integrity [13] [14]. Fungal associated elements in smokeless tobacco may include spores, cells wall components such as flagella and toxins that can impact on local and systemic body responses [15]. Fungi for example, are able to breakdown nicotine [albeit with less efficacy than bacteria] [16, 17] playing a role in the sustainability of harmful constituents in smokeless tobacco such as carcinogenic nitrosamines [18].

Studies also increasingly report that fungi and their biological components are associated with the development of chronic inflammation including oral squamous cell carcinoma [OSCC] development [14, 15]. The mycobiome of smokeless tobaccos likely is involved in the development of soft tissue damage and impinges on the fungal community of the surrounding OSCC. It has been shown that there is a varied fungal distribution surrounding OSCC tumours compared to normal oral tissue with some studies reporting an atypical loss of *Candida albicans* [19] while others have found an enrichment of this species [20].

In this study, we analyse the mycobiome of the Sudanese smokeless tobacco, Toombak, a smokeless tobacco produced in western Sudan from the *Nicotiana Rustica* tobacco plant. The Toombak is produced through open-air fermentation and the addition of sodium bicarbonate. It is often stored for unknown but often lengthy periods of time in extensive heat. A 'dip' is often replaced by users several times a day. While we have explored the bacterial genomics and metabolics, physical structural, heavy metal content and pH of Toombak in a previous paper [3], here we assess the mycobiome of Toombak, a novel area in the understanding of how Toombak may further contribute to the oral and systemic health of Sudanese users. We further explore how the mycobiome of Toomba is impacted by geographic location at the cities Omdurman [the most populated city in the country with a lower standard of living] and Khartoum-North [21].

2.3 Materials and methods

Sample collection and location

Sixteen samples of ready to buy Toombak were purchased from Sudan. Seven of these were purchased from Omdurman [west] while nine samples were retrieved from Khartoum-North [east], separated by the river Nile [22]. Toombak sold in these markets is often prepared by different persons and in varied manufacturing methods. Samples were initially kept at room temperature in airtight polythene bags [Figure 1] after which they were then shipped to Cork, Ireland, less than one week after sample collection. Here they were refrigerated, to prevent any changes to the product.

Figure 1: The physical nature of Toombak

1: Toombak is a dark brown and pungently odourful smokeless tobacco produced in Sudan. It is sold in polythene bags and often stored at hot room temperature.

Mycobiome analysis

DNA extraction

DNA extraction of samples was achieved by the DNeasy Powersoil Pro Kit [QIAGEN®, Hilden, Germany]. Eluted DNA was transferred to sterile Eppendorf tubes and frozen at - 80^{o}C for further analysis.

Amplicon sequencing

ITS gene sequencing was prepared with a similar method to the Illumina Metagenomic Sequencing Library Preparation protocol for 16S rRNA gene library preparation and followed the methods described by Walsh et al [23]. PCR was performed on DNA extracted samples using AMPure XP beads [Beckman Coulter] and primers that were specific to the ITS1-ITS2 regions of the ITS gene with Illuminas overhang adaptors [ITSF1 primer 5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTGGTCATTTAGAGG AAGTAA 3′ and ITS2 primer 5′

GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTGCGTTCTTCATCG ATGC 3′].

PCR was then performed in a 50 μL reaction mixture containing 10 ng of template DNA and 2x KAPA HiFi HotStart Ready Mix. The following thermal cycling conditions were used: 3 min of initial denaturation at 95°C; 30 cycles each of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 30 s and the last step at 72°C for 5 min.

PCR clean-up was achieved with AMPure XP beads [Beckman Coulter] to purify the 16S V3 and V4 amplicon from free primers and primer dimer species. Electrophoresis was performed on the PCR-amplified products using 1.5% agarose gel with Invitrogen DNA loading dye [ThermoFisher Scientific] to verify the expected size at 550 base pairs. The gel was then visualised under 300 nm ultraviolet light. Index PCR was then continued by attaching dual indices, and Illumina sequencing adapters using Nextera XT index kits, and the following PCR thermal cycling conditions were used: 3 min of initial denaturation at 95°C; 8 cycles each of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 30 s; the last step at 72°C for 5 min. AMPure XP beads [Beckman Coulter] were utilised for final clean-up, and the final library was quantified using the Qubit® 2.0 Fluorometer and the Qubit dsDNA HS Assay kit [ThermoFisher Scientific]. Samples were then normalised, and the final library run on an Agilent high sensitivity chip and quantified by qPCR using the KAPA Illumina Library quantification kit.

The resulting sequencing reads were variable in length and were processed using DADA2 [v1.18.0] in R, following DADA2 pipeline workflow for ITS sequences [24, 25]. The reads were filtered to remove sequences with ambiguous ‘N’ bases. This was followed by primer removal from the reads using Cutadapt [v3.5] to remove primers as described in the DADA2 ITS pipeline workflow [26]. The quality of the reads was inspected, and the reads were further filtered and trimmed to retain reads that were a minimum of 50 base pair long with default parameters. Dereplicating and merging of paired-end reads was done following the workflow, following which chimeras were removed and taxonomy was assigned to the resulting ASVs [Amplicon Sequence Variants] using the UNITE ITS database.

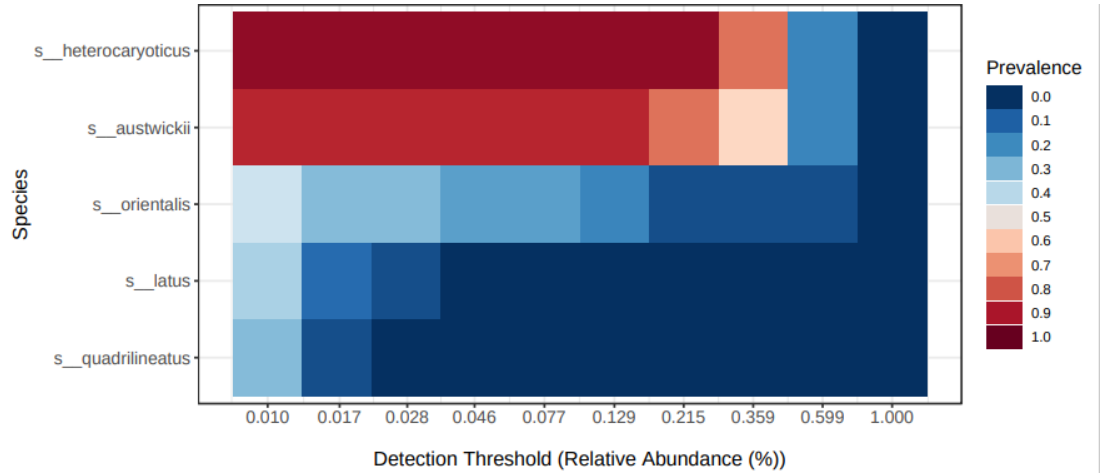
Statistical analysis

The resulting ASV table was used for statistical analysis with Microbiome Analyst [www.microbiomeanalyst.ca] to achieve fungal relative abundance, core microbiome delineation, species detection, Metagenomeseq through false detection rates or q-value and One-way Anova significant variation analysis and LEfSe plotting. Data were not rarefied but normalised utilising total sum scaling [TSS] [27].

2.4 Results

A total of 3474917 reads could be obtained with 1531 ASVs. Toombak from both geographic locations was found to be enriched in the phylum *Ascomycota* [86-93%] while *Basidiomycota* were more abundant amongst Omdurman samples [3.3%] compared to Khartoum-North [0.7%]. Other phyla in Toombak included *Chytridiomycota* and *Mucoromycota*. Toombak was found to be enriched in the genus *Aspergillus* [85.54%] followed by *Rhizopus* [4.46%] where only the species *Rhizopus arrhizus* could be detected. *Aspergillus heterocaryoticus* [43.27%] and *austwickii* [39.48%] dominated the samples with a high detection threshold in the core mycobiome entity profiling for *Aspergillus* species. *Aspergillus orientalis, latus* and *quadrilineatus* were also a part of Toombak mycobiome, albeit with lower relative abundance [Figure 2].

Figure 2: Core mycobiome entity profiling of Aspergillus species in Toombak



2: Core mycobiome entity profiling of Aspergillus species in Toombak highlights the species Aspergillus heterocaryoticus and austwickii to be the most abundant in samples, while

Aspergillus orientalis, latus and quadrilineatus were low core mycobiome species counterparts.

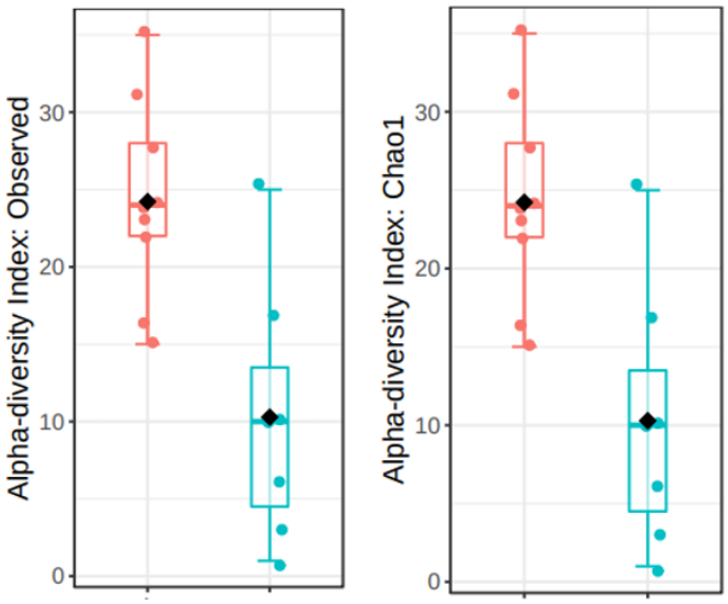
Low relative abundance features in Toombak are were found to include but were not limited to the genera *Issatchenkia, Giberella, Cladosporium* and *Candida* [Table 1].

Table 1. Low relative abundance of fungal genera and species in Toombak

Genera	**Relative abundance [%]**	**Species detected**
Issatchenkia	1.70%	*Issatchenkia orientalis*
Gibberella	1.08%	*Gibberella intricans*
Cladosporium	0.57%	*Cladosporium delicatum* [0.38%]
Epicoccum	0.42%	
Cyberlindnera	0.35%	*Cyberlindnera fabianii*
Thielavia	0.35%	*Thielavia subthermophila*
Malassezia	0.28%	*Malassezia restricta* [0.26%]
Exserohilum	0.22%	
Candida	0.21%	
Aspergillus species detected but in low abundance in Toombak		*Aspergillus minisclerotigenes* [1%]
		Aspergillus quadrilineatus [0.71%]
		Aspergillus latus [0.65%]

There was a significantly higher Chao1 and observed richness [p=0.003] alpha [α] diversity in samples from Khartoum-North compared to Omdurman samples [Figure 3].

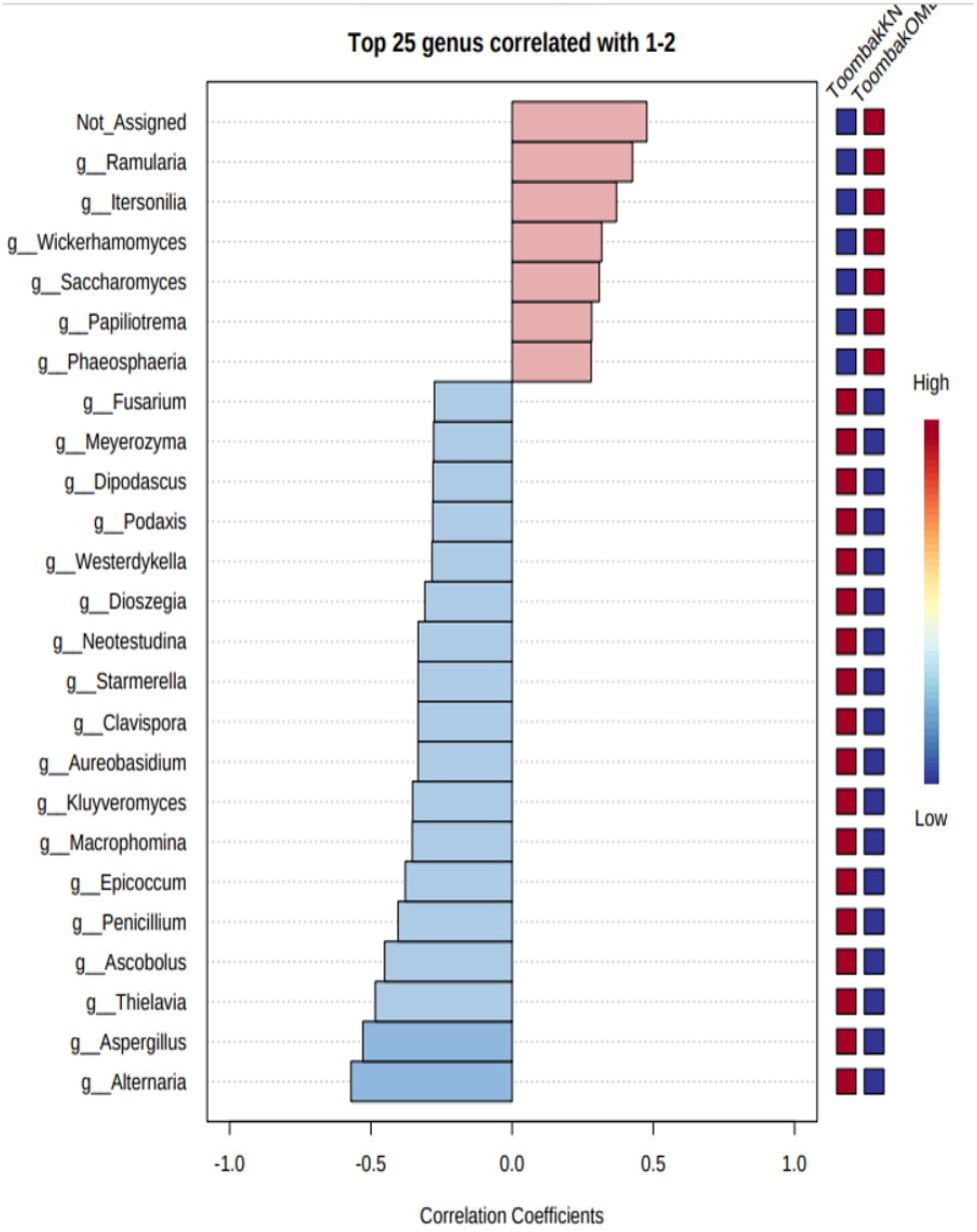
Figure 3: α diversity amongst Toombak samples [Pink=Khartoum-North, Blue=Omdurman]



3: α diversity amongst Toombak samples was found to be higher in Khartoum-North purchases. There was a significantly higher Chao1 and observed richness [p=0.003] α diversity in samples from Khartoum-North compared to Omdurman samples.

Utilising pattern search, the genera *Ramularia, Itersonilia, Wickerhamomyces, Saccharomyces, Papiliotrema* and *Phaeosphaeria* were positively correlated with samples from Omdurman while the genera *Alternaria, Aspergillus, Thielavia, Ascobolus, Penicillium, Epicoccum, Macrophomina, Kluyveromyces, Aureobasidium, Clavispora, Starmerella, Neotestudina, Dioszegia, Westerdykella, Podaxis, Dipodascus, Meyerozyma* and *Fusarium* were positively correlated with samples from Khartoum-North [Figure 4].

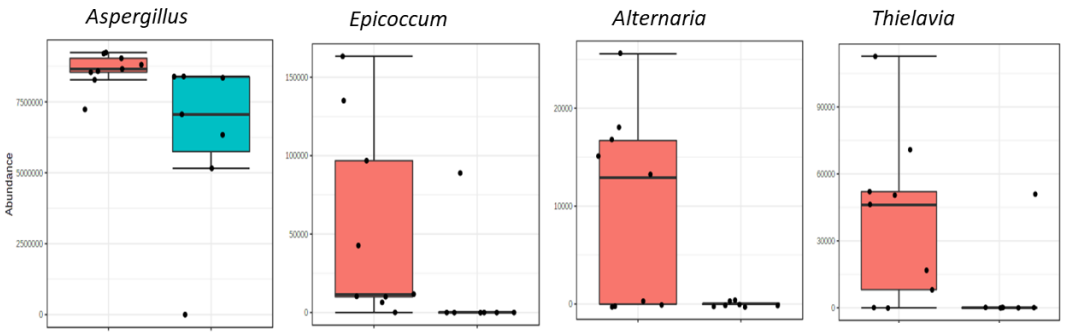
Figure 4: Pattern search of the top 25 genera affiliations in Khartoum-North and Omdurman Toombak samples



4: Pattern search of top 25 correlates of mycobiomic genera found in Omdurman and Khartoum-North samples. The genera Ramularia, Itersonilia, Wickerhamomyces, Saccharomyces, Papiliotrema and Phaeosphaeria were found to be positively correlated and high in Toombak from Omdurman while the genera Alternaria, Aspergillus, Thielavia, Ascobolus, Penicillium, Epicoccum, Macrophomina, Kluyveromyces, Aureobasidium, Clavispora, Starmerella, Neotestudina, Dioszegia, Westerdykella, Podaxis, Dipodascus, Meyerozyma and Fusarium were associated with Toombak from Khartoum-North.

One-way Anova highlighted a significant abundance in the genera *Aspergillus* [p=0.006], *Epicoccum* [p=0.01], *Alternaria* [p=0.02] and *Thielavia* [p=0.02] amongst Khartoum-North samples [Figure 5]. The species *Candida albicans* [p=0.008] and *Alternaria angustiovoidea* [p=0.017] in Khartoum-North samples were also increased. While *Aspergillus* was abundant in Toombak from both locations, LEfSe plotting indicated a higher discriminant for Khartoum-North LDA score [-6.08]. LEfSe plotting further distinguished the genera, *Epicoccum* [LDA score = -4.3], *Alternaria* [LDA score = -3.69] and *Thielavia* ⌈LDA score = -4.21] with Khartoum-North samples.

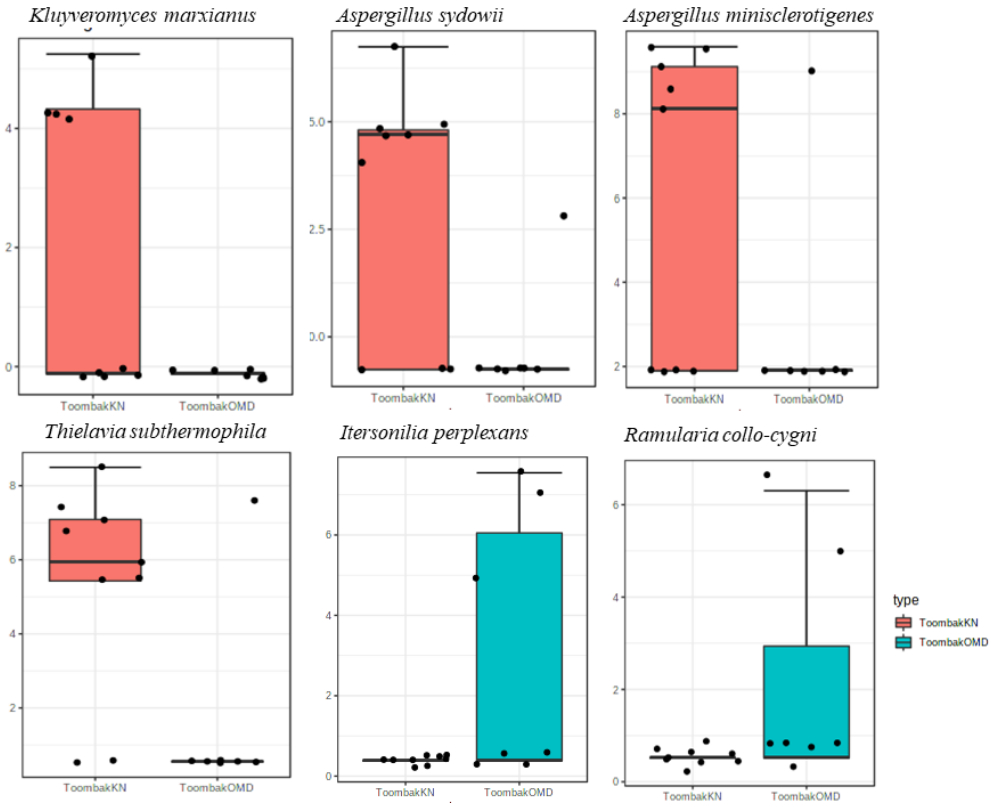
Figure 5: One-way Anova with the most significantly varied genera between samples



5: Pink=Khartoum-North, Blue=Omdurman. One-way Anova highlighted a significant increased abundance in the genera Aspergillus [p=0.006], Epicoccum [p=0.01], Alternaria [p=0.02] and Thielavia [p=0.02] amongst Khartoum-North samples.

Incorporating Metagenomeseq to continue detection of species variations amongst samples with false detection rates [q value], we found *Itersonilia perplexans* [q=1.0213e-8] and *Ramularia collo-cygni* [q=0.01], were significantly increased in samples from Omdurman while *Kluyveromyces marxianus* [q=1.856e-5], *Aspergillus sydowii* [q=3.561e-4], *Aspergillus minisclerotigenes* [q=0.001] and *Thielavia subthermophila* [q=0.001] were significantly increased in samples from Khartoum-North [Figure 6].

Figure 6: Species variations



6: Log fold abundances using Metagenomeseq in Omdurman [blue] and Khartoum-North [pink] samples of Toombak. Khartoum-North exhibited variations with significant increases in Kluyveromyces marxianus [q=1.856e-5], Aspergillus sydowii [q=3.561e-4] and minisclerotigenes [q=0.001], and Thielavia subthermophila [q=0.001] while Omdurman samples were found to have increased Itersonilia perplexans [q=1.0213e-8] and Ramularia collo-cygni [q-0.01].

2.5 Discussion

Through this study, the mycobiome of 16 Sudanese smokeless tobacco samples [Toombak] was analysed using next generation sequencing and was found to harbour an array of fungal genera and species but with a core mycobiome predominantly enriched in *Aspergillus*. Compositional variations were also found to exist between samples relating to the place of purchases, Khartoum-North and Omdurman.

The mycobiome of smokeless tobacco is said to share similarities to the mycoflora of tobacco leaves and along with the microbiome, has an important role to play in its final constituency [28]. Some studies have advocated that fungi play a more vital role in the early stages of fermentation [1, 8, 29] [4] but a lesser role to play overall compared to the bacterial flora of smokeless tobacco [30]. On the other hand the fungal composition may also provide specific traits to smokeless tobacco such as effecting the flavour of these products [31].

In Toombak, the stalks along with the leaves of the tobacco plant are often incorporated into smokeless tobacco production, allowing a wider niche of fungal communities to be a component of the final product [10]. Toombak was found to be predominantly enriched in the phyla *Ascomycota* and *Basidiomycota*, similar to findings from other smokeless tobacco products from around the world [5]. Predominance of *Ascomycota* is more strongly associated with dry forms of smokeless tobacco rather than moist forms such as Toombak [2].

Similar to our findings, *Aspergillus* is the most predominant genus in smokeless tobacco products from India and the USA [6] while in a study of Nigerian smokeless tobacco, *Aspergillus* was found to be of 27.8% abundance [32]. Other studies have however found *Aspergillus* to harbour a reduced compositional importance in the mycobiome of smokeless tobacco [5]. *Aspergillus* is likely a natural part of the smokeless tobacco mycobiome that develops during tobacco leaf decomposition, aging, production and storage [6, 18] but can also become part of the smokeless tobacco mycobiota due to additives and contamination [33].

Interestingly, reports have also conveyed that *Nicotiana* tobacco plant extracts harbour antifungal effects against some *Aspergillus* species [34]. High nicotine levels, such as those in Toombak have also been shown to have inhibitory effects on some fungi such as *Candida albicans* and *Cryptococcus neoformans* [35].

The high abundance of *Aspergillus* found in Toombak samples likely exposes users of this product to mycotoxins that may include aflatoxins, patulin, and ochratoxin A. Mycotoxins may be a potentiating risk factor for the development of oral premalignant lesions and squamous cell carcinoma amongst Toombak users and may also be the cause for the development of systemic tumours [hepatocellular carcinoma] and infectious diseases [Aspergillosis] [6, 36].

Furthermore, *Aspergillus* is able to reduce nitrate to nitrite likely contributing to the high concentrations of the class 1 carcinogenic compounds, tobacco specific nitrosamines, which have been reported to be potently elevated in Toombak [5]. The alkalinic pH of Toombak is also likely to affect mycotoxin concentrations [33]. In this study, although *Aspergillus* was abundant in all samples, it was more abundant in Khartoum-North samples.

Aspergillus heterocaryoticus and *austwickii* were sequenced for the first time in this study from Toombak samples while other *Aspergillus* species were also identified [Figure 2], possibly relating to the storage of Toombak [37]. In other studies, *Aspergillus fumigatus* has been detected from the smokeless tobacco, Gutkha [7] while in smokeless tobacco products from Pakistan, *Aspergillus* species detected included *Aspergillus flavus and terreus* [6].

Aspergillus heterocaryoticus has been detected as part of the mycobiome of fermented soil during the production of Chinese liquor [38], while *Aspergillus austwickii* is a recently described species and part of the *Aspergillus flavus* clades. Problematically, both *Aspergillus austwickii* and *minisclerotigenes*, found in Toombak can produce numerous and potent fungal extrolites that include the aflatoxins, B1, B2, G1 and G2, cyclopiazonic acid, [a mycotoxin that can suppress both immune and haematopoietic cells [39]], aflatrem and paspalinine [tremorgenic mycotoxins] [37].

Aspergillus quadrilineatus, in the Toombak mycobiome, produces mycotoxins that can lead to systemic dysfunctions such as increased heart and respiratory rates, somnolence and diarrhoea [40]. In the head and neck region, this species has also been reported to be a cause for fungal sinusitis [41], while Sterigmatocystin, a carcinogenic, is also produced by *Aspergillus quadrilineatus* [42]. Furthermore, *Aspergillus quadrilineatus* has the potential to inhibit the growth of *Candida albicans*, a normal inhabitant of the oral cavity [43] thus likely contributing to oral dysbiosis. *Aspergillus*

sydowii, is alkaliphilic and a halophilic species appropriate to the Toombak environment [44, 45], while *Aspergillus latus* can avoid immune cell detection by macrophages and neutrophils and is resistant to antifungal medication [46].

In a study of smokeless tobacco products from Nigeria, *Aspergillus niger* and *flavus*, as well as other species that included *Rhizopus stolonifer* and *Alternaria alternata* were reported in samples [47]. In this study, there was a differing *Rhizopus* species presence with a high abundance of *Rhizopus arrhizus [Rhizopus oryzae]* [48]. *Rhizopus arrhizus* can survive high temperatures and is likely to have a role to play during the early fermentation of Toombak [49].

Several low abundance fungal genera and species were found in Toombak that included *Issatchenkia orientalis* [1.7%], *Gibberella intricans* [1.08% and *Cladosporium delicatum* [0.38%] [Table 1]. These genera likely add to an over-representation of fungal species in the mouth with subsequent altered local host mucosal responses to the normal oral mycobiome [50].

We further found variations between the Toombak sampled from the two cities: Omdurman and Khartoum-North. In Omdurman samples, presence of the genus, *Wickerhamomyces*, a mosquito symbiont may allude to environmental contamination while those genera that reflect tobacco leaf inferiority include the blight causing species, *Itersonilia perplexans* [51] and the leaf pathogen, *Ramularia collo-cygni* [52]. Toombak from Omdurman may also harbour an increase in the flavour developing fungi [*Papiliotrema]* [53] while Khartoum-North samples were positively correlated with phytotoxin producing genera [*Macrophomina]* [54]. One-way Anova further distinguished Khartoum-North samples to be significantly increased in the genera *Aspergillus, Epicoccum, Alternaria* and *Thielavia* [Figure 5]. The genera *Alternaria* and *Epicoccum* have consistently been reported to be hypersensitivity allergens that can stimulate IgE release in the body [55] [56], while *Thielavia* can grow rapidly in high temperatures [57].

Ascobolus, Fusarium, Penicillium, Starmerella and *Westerdykella,* found in Toombak have also been detected in Indian smokeless tobaccos [5] while *Neotestudina*, one of the causative agents of Eumycetoma in Sudan was a unique finding to the Toombak smokeless tobacco [58]. Mycetoma is an endemic disease of Sudan characterised by granulomatous inflammation and deformity to affected body parts [59]. Although it is

unlikely that Toombak allows for direct Mycetoma infection in the body, it may have an important and unknown role to play in the transfer of organisms associated with the disease and the geographical spatial distribution of Mycetoma in Sudan [60].

The species, *Kluyveromyces marxianus, Thielavia subthermophila, Candida albicans* and *Alternaria angustiovoidea* were found also to be significantly increased in Toombak from Khartoum-North [Figure 6] *Kluyveromyces marxianus* is a yeast isolated from many environments including plants and fermented foods [61] and in Sudan has been described in the Sudanese fermented camel milk, Garis [62] and the sour milk fermented product, Rob [63].

Conclusion

Our findings provide new evidence into the mycobiomic biocomposition of Toombak and helps complete the understanding of its microbiological nature. Although *Aspergillus* has been previously reported in smokeless tobaccos, its high abundance in Toombak is a unique finding in this study while for the first time, new *Aspergillus* species have been detected in this product. Indeed the mycobiome of Toombak poses a considerable human pathogenic risk [64] and can hamper local mucosal immunity likely in a chronic manner [65].

This study highlights the unparalleled nature of Toombak compared to other smokeless tobaccos throughout the world and in tandem, questions the predispositions to disease states that may accompany Toombak use. *Aspergillus, Penicillium, Alternaria* and *Fusarium* detected in Toombak for example may be the source of previously undiagnosed allergic disease [5] while *Neotestudina*, one of the causative agents of the endemic disease Eumycetoma, through this study, was found to reside in Toombak.

Delineating the mycobiome of Toombak thus allows future research to expand upon new pathways in the investigation, diagnosis, treatment, and management of oral and clinical diseases that may be adversely linked to the use of this smokeless tobacco while our findings also help set the future direction for Toombak production

2.6 References

1. Tyx R, Rivera A, Keong L, Stanfill S. 2020. An exploration of smokeless tobacco product nucleic acids: a combined metagenome and metatranscriptome analysis. Applied microbiology and biotechnology 104:751-763.
2. Rivera AJ, Tyx RE, Keong LM, Stanfill SB, Watson CH. 2020. Microbial communities and gene contributions in smokeless tobacco products. Appl Microbiol Biotechnol 104:10613-10629.
3. Sami A, Elimairi I, Patangia D, Watkins C, Ryan CA, Ross RP, Stanton C. 2021. The ultra-structural, metabolomic and metagenomic characterisation of the sudanese smokeless tobacco 'Toombak'. Toxicology Reports 8:1498-1512.
4. Tyx R, Rivera A, Stanfill S, Zaatari G, Watson C. 2022. Shotgun metagenome sequencing of a Sudanese toombak snuff tobacco: genetic attributes of a high tobacco-specific nitrosamine containing smokeless tobacco product. Letters in Applied Microbiology 74:444-451.
5. Sajid M, Srivastava S, Yadav RK, Singh H, Singh S, Bharadwaj M. 2022. Composition and Ecological Functionality of Fungal Communities Associated with Smokeless Tobacco Products Mainly Consumed in India. Microbiology Spectrum 10:e02273-21.
6. Saleem S, Naz SA, Shafique M, Jabeen N, Ahsan SW. 2018. Fungal contamination in smokeless tobacco products traditionally consumed in Pakistan. J Pak Med Assoc 68:1471-1477.
7. Shetty PJ, Hegde V. 2015. Assessment of pathogenic micro-organisms in commercially available smokeless tobacco products. Avicenna Journal of Medicine 5:54-55.
8. Vishwakarma A, Verma D. 2021. Microorganisms: crucial players of smokeless tobacco for several health attributes. Applied Microbiology and Biotechnology 105:6123-6132.
9. Jia-Xi ZHOU L-FY, Xiao-Min ZHANG, Yuan-Fei LOU, Li-Yuan ZHANG, Yang YANG, Xiao ZOU. 2018. Analysis of fungal community composition on the stored tobacco surfaces. Mycosystema 37[4]:434-443.

10. Stanfill SB. 2020. Reducing carcinogen levels in smokeless tobacco products, p 167-187, Smokeless Tobacco Products. Elsevier.
11. Alsaimary IE, Jassim HA, Mezal TJ. Bacterial contamination of various tobacco types.
12. Larsson L, Szponar B, Ridha B, Pehrson C, Dutkiewicz J, Krysińska-Traczyk E, Sitkowska J. 2008. Identification of bacterial and fungal components in tobacco and tobacco smoke. Tobacco Induced Diseases 4.
13. Khan S, Sinha R, Dixit A. 2022. Microbial "OMICS" in Oral Cancer, p 149-161, Microbes and Oral Squamous Cell Carcinoma: A Network Spanning Infection and Inflammation. Springer.
14. Belvoncikova P, Splichalova P, Videnska P, Gardlik R. 2022. The Human Mycobiome: Colonization, Composition and the Role in Health and Disease. Journal of Fungi 8:1046.
15. Pauly JL, Paszkiewicz G. 2011. Cigarette smoke, bacteria, mold, microbial toxins, and chronic lung inflammation. Journal of oncology 2011.
16. Lisuma JB, Mbega ER, Ndakidemi PA. 2019. Influence of nicotine released in soils to the growth of subsequent maize crop, soil bacteria and fungi. Int J Agric Biol 22:1-12.
17. Tyx RE, Rivera AJ, Satten GA, Keong LM, Kuklenyik P, Lee GE, Lawler TS, Kimbrell JB, Stanfill SB, Valentin-Blasini L. 2022. Associations between microbial communities and key chemical constituents in US domestic moist snuff. PloS one 17:e0267104.
18. Vondracek M, Xi Z, Larsson P, Baker V, Mace K, Pfeifer A, Tjälve H, Donato MT, Gomez-Lechon MJ, Grafström RC. 2001. Cytochrome P450 expression and related metabolism in human buccal mucosa. Carcinogenesis 22:481-488.
19. He S, Chakraborty R, Ranganathan S. 2023. Metaproteomic Analysis of an Oral Squamous Cell Carcinoma Dataset Suggests Diagnostic Potential of the Mycobiome. International Journal of Molecular Sciences 24:1050.
20. Perera M, Al-Hebshi NN, Perera I, Ipe D, Ulett GC, Speicher DJ, Chen T, Johnson NW. 2017. A dysbiotic mycobiome dominated by Candida albicans is identified within oral squamous-cell carcinomas. Journal of oral microbiology 9:1385369-1385369.
21. Zerboni A, Brandolini F, Mariani GS, Perego A, Salvatori S, Usai D, Pelfini M, Williams MAJ. 2021. The Khartoum-Omdurman conurbation: a growing

megacity at the confluence of the Blue and White Nile Rivers. Journal of Maps 17:227-240.

22. Ibrahim A. 2015. Analysis of Risk Factors Associated with Seroprevalence of Neospora caninum in Dairy Animals from Khartoum State, Sudan ARTICLE INFO ABSTRACT Article history. Sudan Journal of Science and Technology Journal homepage: http://jstsustechedu/ 16:1-9.
23. Walsh AM, Crispie F, Kilcawley K, O'Sullivan O, O'Sullivan MG, Claesson MJ, Cotter PD. 2016. Microbial Succession and Flavor Production in the Fermented Dairy Beverage Kefir. mSystems 1:e00052-16.
24. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13:581-583.
25. Team RC. 2014. R: A language and environment for statistical computing. MSOR connections 1.
26. Martin M. 2011. CUTADAPT removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17.
27. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. 2017. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Research 45:W180 - W188.
28. Kaul P, Saxena N, Sharma P. 1990. Mycoflora of tobacco leaves and tobacco products. International Journal of Ecology and Environmental Sciences 16:179-185.
29. Rivera A, Tyx R. 2021. Microbiology of the american smokeless tobacco. Applied microbiology and biotechnology 105:4843-4853.
30. Di Giacomo M, Paolino M, Silvestro D, Vigliotta G, Imperi F, Visca P, Alifano P, Parente D. 2007. Microbial community structure and dynamics of dark fire-cured tobacco fermentation. Appl Environ Microbiol 73:825-37.
31. Shen H, Meiling Y, Yao L, Yangyuan T, Duobin M, Changtong L, Zhifei C, Yingjie F. 2021. Study on the difference of chemical composition and microbiome of different flavor types of reconstituted tobacco leaf concentrate. Journal of Light Industry 36.

32. Okechi R, Oparaugo J, Azuwike C, Nnokwe J, Nneamaka C, Ezenekwe F. 2014. A Survey on the Microbial Contaminants of Snuff Sold in Local Markets in Imo State, Nigeria. 3:366-373.
33. Zitomer N, Rybak ME, Li Z, Walters MJ, Holman MR. 2015. Determination of aflatoxin B1 in smokeless tobacco products by use of UHPLC-MS/MS. Journal of agricultural and food chemistry 63:9131-9138.
34. Udoye BO, Okeke SI. Phytochemical Properties, Pharmacological Activities and Ethnomedicinal Uses of Nicotiana tabacum L.-A Comprehensive Review.
35. Pavia CS, Pierre A, Nowakowski J. 2000. Antimicrobial activity of nicotine against a spectrum of bacterial and fungal pathogens. J Med Microbiol 49:675-6.
36. Patra B, Deep SK, Rosalin R, Pradhan SN. 2022. Flavored Food Additives on the Leaves of Piper betle L.: A Human Health Perspective. Applied Biochemistry and Biotechnology:1-23.
37. Frisvad JC, Hubka V, Ezekiel CN, Hong SB, Nováková A, Chen AJ, Arzanlou M, Larsen TO, Sklenář F, Mahakarnchanakul W, Samson RA, Houbraken J. 2019. Taxonomy of Aspergillus section Flavi and their production of aflatoxins, ochratoxins and other mycotoxins. Stud Mycol 93:1-63.
38. Pu S, Yan S. 2022. Fungal Diversity Profiles in Pit Mud Samples from Chinese Strong-Flavour Liquor Pit. Foods 11.
39. Hymery N, Masson F, Barbier G, Coton E. 2014. Cytotoxicity and immunotoxicity of cyclopiazonic acid on human cells. Toxicology in vitro 28:940-947.
40. Gbodi TA, Nwude N, Aliu YO, Ikediobi CO, Chineme CO. 1991. Acute toxicity of crude extracts of Aspergillus quadrilineatus isolated from acha [Digitaria exilis Stapf]. Vet Hum Toxicol 33:27-31.
41. Polacheck I, Nagler A, Okon E, Drakos P, Plaskowitz J, Kwon-Chung K. 1992. Aspergillus quadrilineatus, a new causative agent of fungal sinusitis. Journal of Clinical Microbiology 30:3290-3293.
42. Gbodi TA. 1993. Isolation, purification, toxicity and some physicochemical properties of mycotoxins produced by aspergillus quadrilineatus isolated from acha [Digitaria exilis stapf] in Plateau State of Nigeria. Vet Hum Toxicol 35:207-12.

43. Hwang JY, Chung B, Kwon OS, Park SC, Cho E, Oh DC, Shin J, Oh KB. 2021. Inhibitory Effects of Epipolythiodioxopiperazine Fungal Metabolites on Isocitrate Lyase in the Glyoxylate Cycle of Candida albicans. Mar Drugs 19.
44. Nishikawa A, Senba H, Kimura Y, Yokota S, Doi M, Takenaka S. 2022. Isolation and characterization of a salt-tolerant γ-glutamyl transpeptidase from xerophilic Aspergillus sydowii. 3 Biotech 12:253.
45. Jiménez-Gómez I, Valdés-Muñoz G, Moreno-Ulloa A, Pérez-Llano Y, Moreno-Perlín T, Silva-Jiménez H, Barreto-Curiel F, Sánchez-Carbente MDR, Folch-Mallol JL, Gunde-Cimerman N, Lago-Lestón A, Batista-García RA. 2022. Surviving in the Brine: A Multi-Omics Approach for Understanding the Physiology of the Halophile Fungus Aspergillus sydowii at Saturated NaCl Concentration. Front Microbiol 13:840408.
46. Steenwyk JL, Lind AL, Ries LNA, Dos Reis TF, Silva LP, Almeida F, Bastos RW, Fraga da Silva TFC, Bonato VLD, Pessoni AM, Rodrigues F, Raja HA, Knowles SL, Oberlies NH, Lagrou K, Goldman GH, Rokas A. 2020. Pathogenic Allodiploid Hybrids of Aspergillus Fungi. Curr Biol 30:2495-2507.e7.
47. Onuorah S, Michael O. 2016. Microbial Contamination Of Locally-Prepared Snuff Sold At Eke-Awka Market, Anmbra State, Nigeria. American Journal of Life Science Researches 4:74-77.
48. Vebliza Y, Sjamsuridzal W, Oetari A, Santoso I, Roosheroe I. 2018. Re-identification of five strains of Rhizopus arrhizus from tempeh based on ITS regions of rDNA sequence data, vol 2023.
49. Chen Q-L, Cai L, Wang H-C, Cai L-T, Goodwin P, Ma J, Wang F, Li Z. 2020. Fungal Composition and Diversity of the Tobacco Leaf Phyllosphere During Curing of Leaves. Frontiers in Microbiology 11.
50. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM. 2010. Characterization of the oral fungal microbiome [mycobiome] in healthy individuals. PLoS Pathog 6:e1000713.
51. Koike ST, Tjosvold SA. 2001. A Blight Disease of Dill in California Caused by Itersonilia perplexans. Plant Dis 85:802.
52. Havis ND, Brown JK, Clemente G, Frei P, Jedryczka M, Kaczmarek J, Kaczmarek M, Matusinsky P, McGrann GR, Pereyra S, Piotrowska M, Sghyer

H, Tellier A, Hess M. 2015. Ramularia collo-cygni--An Emerging Pathogen of Barley Crops. Phytopathology 105:895-904.

53. Zheng T, Zhang Q, Li P, Wu X, Liu Y, Yang Z, Li D, Zhang J, Du G. 2022. Analysis of Microbial Community, Volatile Flavor Compounds, and Flavor of Cigar Tobacco Leaves From Different Regions. Frontiers in Microbiology 13.
54. Dhar TK, Siddiqui KA, Ali E. 1982. Structure of phaseolinone, a novel phytotoxin from Macrophomina phaseolina. Tetrahedron letters 23:5459-5462.
55. Caillaud D, Keirsbulck M, Leger C, Leynaert B, Group OMAW. 2022. Outdoor Mold and Respiratory Health: State of Science of Epidemiological Studies. The Journal of Allergy and Clinical Immunology: In Practice 10:768-784. e3.
56. POITER P, BERMAN D, TOERIEN A, MALHERBE D, WEINBERG E. 1991. Clinical significance identification in the. SAMJ 79:19.
57. Badali H, Chander J, Gupta A, Rani H, Punia RS, De Hoog GS, Meis JF. 2011. Fatal cerebral phaeohyphomycosis in an immunocompetent individual due to Thielavia subthermophila. J Clin Microbiol 49:2336-41.
58. Ahmed A, van de Sande W, Verbrugh H, Fahal A, van Belkum A. 2003. Madurella mycetomatis strains from mycetoma lesions in Sudanese patients are clonal. Journal of clinical microbiology 41:4537-4541.
59. Fahal A, Mahgoub ES, Hassan AME, Abdel-Rahman ME. 2015. Mycetoma in the Sudan: an update from the mycetoma research centre, University of Khartoum, Sudan. PLoS neglected tropical diseases 9:e0003679.
60. Hassan R, Simpson H, Cano J, Bakhiet S, Ganawa E, Argaw D, Newport MJ, Deribe K, Fahal AH. 2021. Modelling the spatial distribution of mycetoma in Sudan. Trans R Soc Trop Med Hyg 115:1144-1152.
61. Karim A, Gerliani N, Aïder M. 2020. Kluyveromyces marxianus: An emerging yeast cell factory for applications in food and biotechnology. Int J Food Microbiol 333:108818.
62. Abdelgadir W, Nielsen DS, Hamad S, Jakobsen M. 2008. A traditional Sudanese fermented camel's milk product, Gariss, as a habitat of Streptococcus infantarius subsp. infantarius. International journal of food microbiology 127:215-219.

63. Ahmed AMK. ISOLATION AND IDENTIFICATION OF YEASTS ASSOCIATED WITH SUDANESE SOUR MILK" ROB" FERMENTATIONUniversity of Khartoum.
64. Chattopadhyay S, Malayil L, Mongodin EF, Sapkota AR. 2021. A roadmap from unknowns to knowns: Advancing our understanding of the microbiomes of commercially available tobacco products. Applied Microbiology and Biotechnology 105:2633-2645.
65. Daniels TE, Chou L, Greenspan JS, Grady DG, Hauck WW, Greene JC, Ernster VL. 1992. Reduction of Langerhans cells in smokeless tobacco-associated oral mucosal lesions. Journal of oral pathology & medicine 21:100-104.

2 – [C] Sudanese Toombak smokeless tobacco users harbour significantly altered long-term cortisol body production

Amel Sami 1,2, Imad Elimairi 2, *, Catherine Stanton 1,3, R. Paul Ross 1, and C. Anthony Ryan 4

1 APC Microbiome Ireland, School of Microbiology, University College Cork, Cork T12 YN60, Ireland.
2 Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, National Ribat University, Nile Street, Khartoum 1111, Sudan
3 Teagasc Food Research Centre, Moorepark, Fermoy, Cork P61 C996, Ireland
4 Department of Paediatrics and Child Health, University College Cork, Cork T12 DFK4, Ireland.

This chapter has been published as Sami A, Elimairi I, Anthony Ryan C, Paul Ross R, Stanton C. Sudanese Toombak smokeless tobacco users harbour significantly altered long-term cortisol body production. Steroids. 2023 Feb 2; 193:109189. Doi: 10.1016/j.steroids.2023.109189 [See Appendix 4]

2.1 Abstract:

The Sudanese, in particular its male population, are known to utilise a smokeless tobacco product [Toombak] which is placed in the oral cavity and can be replaced several times a day. Toombak has been shown to harm human health and is highly addictive. The effect on body cortisol response over a retrospective period in users of this product has not been previously explored. In addition, psycho-dependency scores of Toombak users have not been analysed.

In this study, 37 male subjects, age 18–45 years were recruited, of which 18 were non-users of Toombak and 19 were Toombak users. One hair sample was collected from each user and non-user of Toombak. Each hair sample [n=37] was placed in a pre-prepared long piece of foil with two labels on either side marked: 'scalp-side' and 'distant-side'. Cortisol was extracted by mincing 10 mg of 'scalp-side' hair, not exceeding 3 cm, with methanol addition, incubation, and sonication. Cortisol was measured using the enzyme-linked immunosorbent assay kit [Enzo Life Sciences, UK]. The amount of hair cortisol in the samples was determined using spectrophotometry at wavelength 405 nm measured in pg/ml and visualised with a four parametric logistic curve. Toombak users were further asked to complete the Fagerstrom Test for Nicotine Dependence-Smokeless Tobacco questionnaire [FTND-ST] comprising of six questions. Scores of > 5 indicated a significant dependence, while a score of < 4 marked low to moderate dependence.

The mean concentration of hair cortisol in Toombak users [9.7 pg/ml] was significantly lower [p=0.023] compared to non-users [19.4 pg/ml], with total concentrations ranging from 2.1 to 55.6 pg/ml. FTND–ST scores ranged from 4 to 9, with high levels of psycho-dependency [score > 5] and nicotine tolerance found in 85 % of Toombak users. Cortisol body release in Sudanese smokeless tobacco users was found to be significantly altered. While low cortisol levels do lead to anxiolytic effects, in the long-term, this can allow for increased susceptibility to low cortisol-associated diseases

2.2 Introduction

Cortisol is a glucocorticoid hormone that is secreted from the adrenal cortex. It is vital in gluconeogenesis and in mediating immunity and reducing inflammation [1]. Cortisol release occurs in a circadian pattern and is regulated by the hypothalamus pituitary adrenal axis. Corticotropin-releasing hormone is released from the paraventricular nucleus of the hypothalamus after which adrenocorticotropic hormone is secreted from the pituitary gland. This mechanism allows the secretion of cortisol to occur through a negative feedback loop whereby cortisol inhibits the release of corticotropin-releasing hormone, deactivating the hypothalamus pituitary adrenal axis [1]. Free cortisol diffuses into hair from the follicular capillaries and is incorporated and maintained in the hair structure due to its lipophilicity. As a result, hair samples remain stable at room temperature and are not affected by washing or drying [2].

Cortisol measurements from hair have shown promising research and clinical uses, particularly in the field of psychoneuroendocrinology [3]. Hair cortisol analysis is a non-invasive method of adrenal gland function in the body [4]. It reflects well, the hypothalamus pituitary adrenal axis [5] and average blood cortisol levels over a broad retrospective period [6]. Utilising hair for cortisol hormone analysis also offers the advantage of the measurements being unaffected by the pulsatile release of cortisol and the psychological and physical mediators on cortisol release such as exercise, trauma, and nicotine usage [7,8]. In healthy individuals, hair cortisol levels have been found to range between 17.7 and 153.2 pg/mg [9], 12–163 pg/ml [10] and 2–51.63 pg/ml [6]. The median hair cortisol concentration through automated methods was found to be 55 pg/mg [11]. Hair cortisol concentration gradually increases with age with incremental increases of 0.10 log pg/ mg each year [12]. In disease states such as the overproduction of cortisol, known as Cushing's syndrome, hair cortisol levels have been reported to be as high as 266 +/- 738.4 pmol/g [13].

The main difficulties in utilising hair as a cortisol hormone assessment method include a general need for a more standardised extraction methodology. The composition of hair also varies globally [such as African hair, which grows slower] making hair cortisol concentrations less comparable between different populations [14, 15]. Hair cortisol can also be damaged by UV radiation [natural sunlight] and chemical treatments [16]. Other samples from the human body from which cortisol can be

extracted include urine, faeces, blood, saliva and nails [17]. Urine cortisol measurements while not affected by the circadian rhythm, only reflect a cortisol picture of days rather than the months that hair cortisol measurements can provide. Salivary cortisol is said to reflect only about 70% of cortisol concentration in the body, while studies on nail cortisol extraction have produced similar results to hair cortisol studies [18–21].

Toombak is a form of smokeless tobacco used primarily by Sudanese males. Users of Toombak generally exhibit chronicity of use of this product starting from a young age. Toombak has been shown to contain high levels of nicotine and other toxic compounds that can harm human health and allow it to be addictive [22]. Nicotine in Toombak likely produces an anxiolytic effect [23] although long-term exposure has been associated with anxiogenic traits [24]. The Fagerstrom Test for Nicotine Dependence questionnaire [also known as the Fagerstrom tolerance test] is one of the most commonly utilised questionnaires measuring nicotine dependence and addiction levels amongst tobacco users and was adapted for smokeless tobacco use by Ebbert et al. [2006] [25] into the Fagerstrom Test for Nicotine Dependence for Smokeless Tobacco [FTND-ST] questionnaire. It can reliably analyse nicotine dependence and has shown a good association with the physical dependence associated with smokeless tobacco use [26–28]. The FTND-ST questionnaire employs the question ‘time to the first use of tobacco after waking’, a validated measure of dependence for smokeless tobacco users [29]. This questionnaire has also accurately portrayed withdrawal and quitting difficulties experienced by smokeless tobacco users [30]. In one study on Nepalese smokeless tobacco users, the FTND-ST questionnaire highlighted an 80% moderate to severe nicotine dependence amongst the study group [31,32]. In this study, the effects on cortisol body production were assessed in chronic Sudanese Toombak users over a retrospective period. The psycho-dependency of Toombak use was measured using the FTND-ST questionnaire.

2.3 Materials and methods

Sample collection, segmentation of hair and storage

Ethical approval was obtained from the National Ribat University, Sudan ethical committee. Hair samples were collected in Omdurman, Sudan, in cooperation with a barber grooming facility. They were obtained from the posterior side [the vertex] of

the head in those who had not attended a hair grooming visit for at least three months. Informed consent from participants was achieved both verbally and written. Participant's hair samples were collected during their routine visit. In addition, a recent medical evaluation was achieved, while those utilising Toombak also completed the FTND-ST questionnaire. Inclusion criteria included healthy individuals with average anthropometric measures, who only utilised smokeless tobacco and were not prescribed exogenous corticosteroids within the last three months. Other information retrieved included the age of the participants, length of Toombak use and frequency of use. Thirty-seven hair samples were collected: 18 from non-users of Toombak and 19 from Toombak users [n=37]. All participants were males and were aged between 18 and 45 years of age. Each hair sample collected was placed in a pre-prepared long piece of foil with two marked labels: 'scalp-side' and 'distant-side'. Participant's hair samples were secured to the appropriate positions in the foil.

Cortisol extraction

A standard method for quantifying hair cortisol concentration was utilised as outlined by Greff et al. [2019] [7]. Only the 'scalp-side' or proximal hair was chosen. Hair samples were weighed [10 mg] and did not exceed 3 cm in length [to reflect a 3-month retrospective cortisol period]. Samples were minced using sterile scissors and placed in sterilised Eppendorf tubes. Where the hair was extensively curly, all attempts were made to straighten the hair with a clean comb before cutting from the appropriate section. Methanol [1 ml] was then added to each Eppendorf tube in a fume hood and incubated at 52 ◦C for 16 h. The following day, the Eppendorf tubes were sonicated for 30 min in a covered bath sonicator [Clifton™ ultrasonic bath] and placed on a heated shaker [Multi-Therm™] at 52 ◦C, with shaking at 200 rpm for 16 h. On day three, the methanol from each hair sample was transferred into a new Eppendorf tube and evaporated to dryness using a sample concentrator [Genevac™ miVac FisherScientific] for five hours at 37 ◦C. Lastly, the residue was reconstituted in 250 μL of phosphate buffer saline, vortexed and stored at − 30 ◦C.

Measurement and analysis of cortisol concentration

Cortisol levels were analysed in duplicate, using a commercial enzyme-linked immunosorbent assay [ELISA] kit [Enzo Life Sciences, UK] in accordance with manufacturers' instructions. Briefly, the kit competitively uses a monoclonal antibody

to bind cortisol. Seven cortisol standards were prepared at concentrations of 10,000, 5,000, 2,500, 1,250, 625, 313 and 156 pg/ml. 100 µL of standards and samples were pipetted into the appropriate wells. Duplicate sampling was performed as the Enzo Life Sciences kit has a high cross-reactivity with the cortisol hormone [100%] thus obtaining accurate results. Furthermore, hair cortisol [unlike saliva or plasma cortisol], is not affected by the daily variations in cortisol concentration that may require a triple average or triplicate analysis. 50 µL of yellow antibody was pipetted into each well [except the blank, total activity and NSB wells]. 50 µL of blue conjugate was pipetted into each well except the total activity and blank wells. Samples were incubated at room temperature on a plate shaker for two hours at 500 rpm. The contents of the wells were emptied and 400 µL of wash solution was added to each well, with this step repeated three times. Next, 200 µL of p-nitrophenyl phosphate in buffer solution was added to the wells and samples were incubated again at room temperature for one hour. 50 µL of stop solution was finally added to every well.

The amount of hair cortisol in the samples was then determined using a 96-well spectrophotometer [Biotek Synergy HT spectrophotometer, BioTek® Instruments, Inc, Vermont, USA] using Gen5TM software and read at wavelength 405 nm with a correction between 570 and 590 nm. The mean optical density of the blank wells was subtracted from all readings. The detection limit was determined as the concentration of cortisol measured at two standard deviations from the zero along the standard curve. Sample recoveries for tissue culture media, saliva and urine were reported for the Enzo Life Sciences kit while hair cortisol recovery percentages were determined from the literature.

FTND-ST questionnaire

The 19 Toombak users were asked to fill in an Arabic translated FTND-ST questionnaire which contained six questions [Table 3]. Answers to question 1 were evaluated as either; ‘within 5 min’, ‘6–30 min’, ‘31–60 min’ or ‘after 60 min’, to question 2 as ‘always’, ‘sometimes’ or ‘never’, to question 3 as ‘the first one in the morning’ or ‘any other’, to question 4 as ‘more than three’, ‘two-three’ or ‘one’, to question 5 as ‘yes’ or ‘no’ and to question 6 as ‘yes’ or ‘no’. The answers provided were assessed on a point scale ranging from 0 to 3. The total score was then calculated and evaluated using the scoring instructions and included adding up responses to all

items. A score of > 5 indicated a significant dependence, while a score of < 4 showed a low to moderate dependence.

Statistical analysis

To achieve regression models of hair cortisol concentrations, the four-parameter logistic curve [4PL] curve fitting programme was performed with the 'MyAssays' website [https://www.MyAssays.com-Data Analysis Tools and Services for Bioassays]. The concentrations of the samples were first determined as corrected background measurements [raw measurement minus the average of the blank group] in relation to the seven standards utilising the 4PL curve.

The coefficient of variation, expressed in percentage [CV %] was also determined for each sample to evaluate the strength of duplicate analysis as well as the precision of the kit. The CV % were calculated as the percentage of the standard deviations of the replicate measurements divided by their mean and were expected to be ≤ 20% for optimum deviations between replicates. The concentration averages of the samples were then further calculated in correspondence to the weight/volume dilution expressions. Goodness of fit was measured using R^2, the mean square error, the sum of the squares of the residuals and the standard deviation of the residuals.

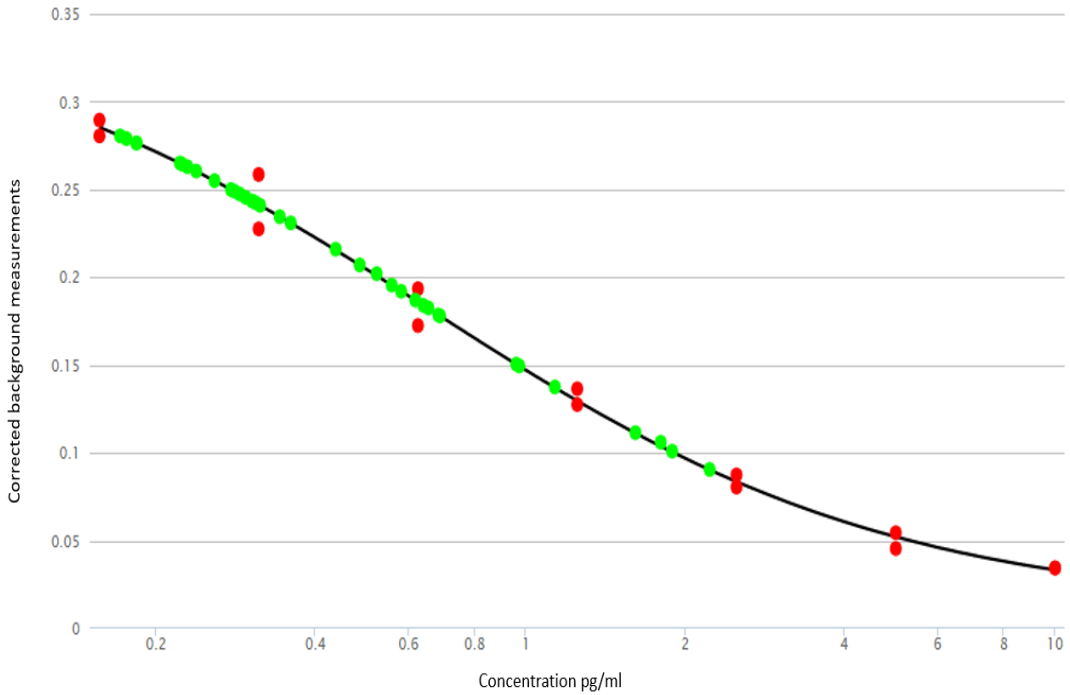
Data were also analysed using the IBM Corp. IBM SPSS Statistics for Windows Version 28.0 [Armonk, NY: IBM Corp]. Descriptive statistics were utilised to evaluate the data while the means of the two study cohorts were compared using independent samples T tests and the significant p value obtained. Correlations of FTND-ST scores with hair cortisol concentrations were further expressed through GGraph syntax utilising subgroup plotting.

2.4 Results

Cortisol assays

Figure 1 highlights a 4PL curve measuring the corrected background values of all samples.

Figure 1:



1: 4PL sigmoidal curve of hair cortisol samples bounded by the Y axis defined by the a [0.011] and d [0.35] coefficients. Optimum goodness of fit R^2 = 0.99 [close to 1]. Standards [red] highlight reproducibility and curve strength along the 4PL curve with samples [green] in range, measured in corrected background values.

These were found to be in range of the standards and were between 0.101 and 0.905 pg/ml [Table 1].

Table 1: Corrected background values were found to range between 0.101-0.905 pg/ml while 92% of samples had a CV% ≤ 20% indicating that samples had acceptable deviation between replicates. Furthermore, the intra-assay cortisol concentration of the Enzo Life Sciences kit was found to be 6.6%.

Sample number	Corrected background values	[CV] %
1	0.276	12.2
2	0.112	10.2
3	0.183	13.5
4	0.286	10.7
5	0.248	9.91
6	0.150	5.26
7	0.905	7.11
8	0.250	1.01
9	0.231	10.1
10	0.101	14.1
11	0.243	9.66
12	0.151	1.75
13	0.246	1.96
14	0.276	2.44
15	0.138	12.7
16	0.179	3.38
17	0.192	0.848
18	0.279	8.74
19	0.281	2.54
20	0.263	14.2
21	0.216	18.4
22	0.244	7.77
23	0.207	4.32
24	0.178	7.6
25	0.255	5.17
26	0.313	52.3*
27	0.106	30.3*
28	0.184	4.22
29	0.265	9.96
30	0.235	1.87
31	0.202	24.7*
32	0.187	7.6
33	0.241	8.63
34	0.264	8.82
35	0.196	5.1

36	0.261	12.9
37	0.249	6.99
		Intra-assay medium cortisol concentrations: 6.6%

**A CV% >20% amongst samples 26,27 and 31 may indicate that wider deviations amongst these sample replicates may exist.*

The CV % for all duplicate samples were found to be ≤ 20% for 92% of samples. Three samples, 26, 27 and 31 had a CV % of > 20%. The intra-assay cortisol concentration of the Enzo Life Sciences kit was found to be 6.6%. R^2, the goodness of fit, was found to be closer to 1, indicating optimum goodness of fit [0.99] of samples. Furthermore, the mean square error [0.0000638], the sum of the squares of the residuals [0.0008939] and the standard deviation of the residuals [0.009455] were all found to be closer to 0 indicating that samples also achieved a high goodness of fit. Furthermore, the detection limit was found to be 0.040 pg/ml while the cortisol sample recovery percentages from tissue culture media, saliva and urine were found to be 106.6%, 97.4% and 102.2% respectively.

While the percentage of hair cortisol sample recovery was not listed in the manual, the literature supports an average hair cortisol recovery of 88% [9]. Average concentrations of hair cortisol measurements in relation to Toombak use, FTND-ST scores and age were found to be between 2.1 and 55.6 pg/ml [Table 2].

Table 2: Summary of the participants of users and non-users of Toombak from Sudan in relation to Toombak use, FTND-ST scores and age. Hair cortisol analysis is expressed in concentration average [pg/ml], a weight/volume dilution of 25 and were found to range between 2.1- 55.6 pg/ml

Participant number	Toombak status	Concentration averages of hair cortisol [average pg/ml]	FTND-ST scores	Age
1	Y	4.59	6	28
2	N	40.25	-	22
3	N	16.36	-	25
4	Y	3.90	7	35
5	N	7.18	-	40
6	N	24.28	-	45
7	N	55.60	-	28
8	Y	6.92	4	25
9	Y	8.98	5	40
10	Y	47.23	6	18

11	Y	7.70	6	40
12	N	23.9	-	50
13	Y	7.37	5	39
14	N	4.58	-	25
15	N	28.36	-	32
16	N	17.08	-	30
17	Y	14.53	4	25
18	N	4.39	-	29
19	Y	4.27	5	29
20	Y	5.73	4	40
21	Y	10.92	6	37
22	Y	7.59	5	27
23	N	12.13	-	25
24	N	17.2	-	30
25	N	6.44	-	35
26	Y	2.13	7	26
27	N	44.96	-	40
28	N	16.0	-	25
29	Y	5.54	5	40
30	N	8.56	-	30
31	Y	13.0	8	28
32	Y	15.4	6	23
33	N	7.85	-	25
34	Y	5.58	9	35
35	N	13.94	-	30
36	Y	5.95	9	25
37	Y	7.03	5	38

The mean concentration of cortisol amongst hair samples in non-users of Toombak was found to be higher [19.4 pg/ ml] compared to Toombak users [9.7 pg/ml]. This difference was statistically significant [two-tailed p=0.023].

Questionnaire scores

FTND–ST scores ranged from 4 to 9. 85% of Toombak users were found to have a FTND-ST score of > 5, reflecting a high dependency on the product, while 15% had a low to moderate dependency score of < 4. Table 3 highlights the dependency questions asked to those utilising Toombak.

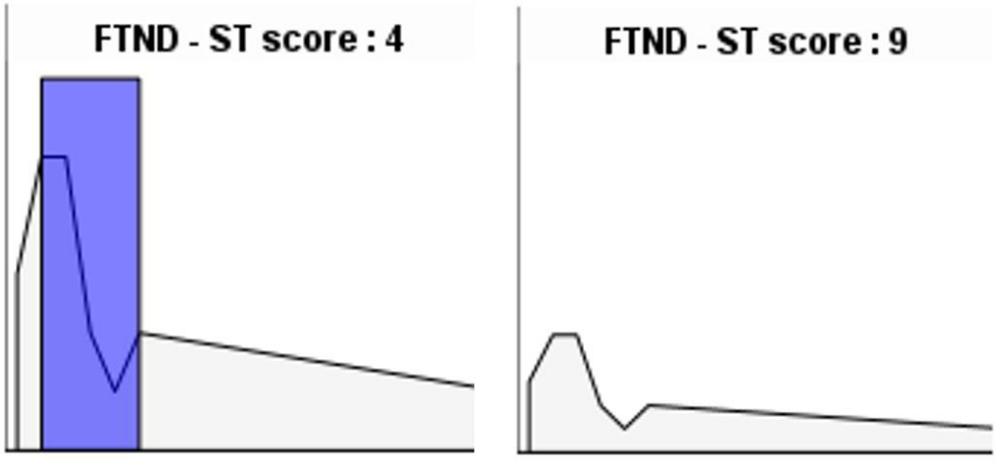
Table 3: Fagerstrom Test for Nicotine Dependence-Smokeless Tobacco [FTND- ST] questions for Toombak users.

Q1. How soon after you wake up do you place your first dip?
Q2. How often do you intentionally swallow tobacco juice?
Q3. Which chew would you hate to give up most?
Q4. How many cans/pouches per week do you use?
Q5. Do you chew more frequently during the first hours after awakening than during the rest of the day?
Q6. Do you chew if you are so ill that you are in bed most of the day?

Toombak use in years ranged from 2 to 12 years while the age of initiation ranged from 15 to 34 years. The frequency of use per day was between 15 and 25 dip replacements/day at an average of 7–10 g per use. Of the 19 participants, four [21%] Toombak users placed a Toombak dip in the oral cavity within 5 min of waking up, eight [42%] Toombak users utilised the tobacco within 6–30 min of waking up while seven [37%] users could use it within 31–60 min of waking up. Fourteen users [74%] never intentionally swallowed Toombak tobacco juice, while five users [26%] sometimes with intention swallowed the juice of the Toombak dip. When participants were asked which dip they hated to give up most, 16 users [84%] did not want to give up the first-morning dip, while only three users [16%] showed no preference for what time the dip was used. Thirteen users [68%] purchased two to three pouches [around 100 g each] of Toombak per week, while six users [32%] purchased more than three pouches. Seventeen users [89%] utilised Toombak more frequently during the first hours after awakening, while only two users [11%] used Toombak more frequently at other times of the day. Twelve Toombak users [63%] continued to place Toombak even if they were ill, while seven users [37%] did not use Toombak while they were bedridden. FTND-ST score correlations with hair cortisol concentrations indicated an inverse relationship between the two variables, such that the lower the dependency score [FTND-ST score of 4], the higher the area under the curve for average cortisol production for that subgroup. On the other hand, the greater the dependency score

[FTND-ST score of 9], the lower the average cortisol production for that subgroup [Figure 2].

Figure 2: Toombak subgroup configuration



2: Subgroups defined by FTND scores where each score is analysed in relation to the cortisol concentration averages [pg/ml]. The lowest [4] and highest [9] FTND- ST scores are compared. Blue= the pronounced inverse relationship with loss of area under the curve of cortisol response as dependency on Toombak heightens. Those with low FTND -ST levels [4], have a higher subgroup/cortisol relationship compared to those with high dependency scores [9] which show blunting.

2.5 Discussion

For the first time, long-term body cortisol level in Sudanese Toombak users is reported in this study. We have shown that chronic use of this form of smokeless tobacco leads to a significant hypocortisolism state compared to normalised cortisol levels in persons who do not utilise Toombak. Furthermore, those utilising Toombak were shown to have high psycho-dependency scores [85%] due to smokeless tobacco use and a nicotine-associated tolerance related to chronic exposure to Toombak.

A stress response is reflected by the activity of the hypothalamus pituitary adrenal axis, a system responsible for both the release and inhibition of cortisol. The hypothalamus secretes corticotropin-releasing hormone. It stimulates the pituitary gland to secrete adrenocorticotropin hormone, which is then carried in the peripheral circulation to allow the release of cortisol during stress from the adrenal cortex.

Nicotine initially promotes the activity of the hypothalamus pituitary adrenal axis and stimulates an increase in cortisol release [33, 34]. However, upon continuous and chronic use of Toombak, this response is altered [35]. It becomes less responsive to the stimulation of nicotine leading to the lower levels of cortisol release seen here amongst long term Toombak users [1, 36 37].

Addiction to other substances, such as to alcohol, gambling, food and exercise, have also shown similar patterns of low cortisol responses in the body [38]. Other factors that could underlie low cortisol release in Toombak users include the reduced synthesis of corticotropin-releasing hormone, low free cortisol, decreased cortisol response at receptor sites and the increased sensitivity to the negative feedback loop of cortisol release [39].

Nicotine increases the release of dopamine which has been found to reduce cortisol levels [40] while blunted cortisol levels may also be attributed to the desensitisation of serotonin receptors [41]. Indeed, high-nicotine containing products have been shown to elicit a diminishing 'high' after continuous use [42]. Dependency on smokeless tobacco has been shown to reduce body cortisol production [43], cause an over-reliance on these products [44] and allow for the increased maintenance of this habit once initiated [38]. Smokeless tobacco users have shown some of the highest Fagerstrom nicotine dependency scores compared to cigarette smokers and those who utilise both products [28], indicating that nicotine levels in these users are likely to give rise to many negative consequences. High-dependency smokers have also shown similarly blunted cortisol responses, with some studies associating this trend to the failure of adrenocorticotropin-releasing hormone in particular [45].

One reason for this is that smokeless tobacco use has been shown to cause potent downregulation of the cyclin-dependant kinase 2 and cyclin E regulatory molecules [46], which have an essential role to play in the production of adrenocorticotropic-releasing hormone. Indeed, treatments targeting cyclin E inhibition, such as Seliciclib are known to be effective in the management of hypercortisolism [Cushing's disease] [47]. Lipid peroxidation induced by chronic nicotine and tobacco specific nitrosamine exposure [both elevated in Toombak], may also reduce cortisol levels.

Long-term smokeless tobacco use can further induce oxidative stress in cells and result in tissue damage [48] in many organs, including the adrenal glands. Those with

Toombak use are likely exposed to the neurological and psychological effects of low cortisol levels. These include impaired stress responses, poor responses to environmental challenges, an increased likelihood of aggressive behaviour, panic attacks, suicidal thoughts, reduced motivation, increased risk of burnout and the development of chronic fatigue syndrome [38, 49–52].

A factor to consider is also the early age initiation that many Toombak users describe. This likely imposes a similar effect on the body to early lifetime adversities, now strongly linked to blunted stress responses in later life [53, 54]. Utilising the FTND–ST questionnaire in this study, it was found that many users had difficulty quitting Toombak, with a lack of control associated with this smokeless tobacco use. This feeling has also been shown to blunt the cortisol response system [36].

The hypocortisolism state reflected in long-term Toombak users can further be said to resemble those with generalised anxiety and post-traumatic stress disorders who are found to exhibit low hair cortisol levels [55]. Increases in the symptoms of post-traumatic stress disorder have also been associated with cortisol hyposecretion [12 56]. Blunted cortisol responses have been recognised in depression, including major depressive disorder [50]. Low cortisol can allow for continued chronic low-grade inflammation with an increase in pro-inflammatory cytokines and dysfunctional autonomic and haematological systems [57]. Hypocortisolism is now strongly linked to many adverse long-term health outcomes, medically unexplained symptoms, irritable bowel syndrome as well as chronic pain, including back pain and fibromyalgia [39, 58, 59].

Indeed, low cortisol concentrations are associated with an increased mortality risk and a poorer prognosis in cardiovascular disease, sepsis and trauma [60]. In addition, higher levels of macrophage inhibitory factor, a cytokine associated with several inflammatory, oncogenic, cardiovascular and autoimmune diseases [61] has also been linked with reduced cortisol levels. The low cortisol levels highlighted amongst Toombak users could further be related to the psychological reduction of stress [62]. In some studies, low hair cortisol measures are positive predictors of successful stress reduction programs [63, 64]. However, low cortisol levels have also been associated with a greater chance of relapse after quitting tobacco [45], higher withdrawal distress and an increased urge to smoke [64].

Cortisol concentrations are known to reduce with smoking cessation [65], however with the already low cortisol levels found in Toombak users and their chronic nicotine exposure, a challenging difficulty arises with quitting the Toombak habit. Taken together, the high nicotine dependency assessed by the FTND-ST questionnaire and the reduced cortisol concentrations found amongst Toombak users, create a picture of unstable psychological and physiological responses to Toombak habit modulation and are likely a potent combination for the addictive nature of this product and associated quitting failures that Toombak users often relay [66].

Conclusion

In conclusion, we have shown that Toombak users are exposed to significant alterations in long-term body cortisol production. In addition, average cortisol release is inversely related to nicotine dependency scores, whereby those with lower FTND-ST scores of 4, had higher average cortisol levels compared to those with FTND-ST score of 9.

This highlights the long-term damage that smokeless tobacco use can impose on adrenal gland function and cortisol release. Some of the limitations of this study include the first-time analysis of hair cortisol amongst the Sudanese population which prevents the comparison of the study results. Furthermore, while hair cortisol is a stable entity and does not undergo daily diurnal alterations, a larger sample size or triplicate analysis of samples may also be utilised in future research to overcome any sample variations.

It would also be prudent that upcoming research aims to capture the longitudinal cortisol responses to Toombak use and in particular at the time of initiation and quitting and in those with illness and surgery. Such understanding may aid practitioners in the specific treatment of Toombak users and achieve better public health programs to support Toombak users who do express a wish to abort the habit.

Acknowledgements

The participants in the study, the funders Esther and SFI/APC Microbiome Ireland

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

2.6 References

1. Rohleder N, Kirschbaum C. 2006. The hypothalamic–pituitary–adrenal [HPA] axis in habitual smokers. International Journal of Psychophysiology 59:236-243.
2. Lanfear JH, Voegel CD, Binz TM, Paul RA. 2020. Hair cortisol measurement in older adults: Influence of demographic and physiological factors and correlation with perceived stress. Steroids 163:108712.
3. Stalder T, Steudte-Schmiedgen S, Alexander N, Klucken T, Vater A, Wichmann S, Kirschbaum C, Miller R. 2017. Stress-related and basic determinants of hair cortisol in humans: A meta-analysis. Psychoneuroendocrinology 77:261-274.
4. Wester VL, van Rossum EF. 2015. Clinical applications of cortisol measurements in hair. Eur J Endocrinol 173:M1-10.
5. Russell E, Koren G, Rieder M, Van Uum S. 2012. Hair cortisol as a biological marker of chronic stress: current status, future directions and unanswered questions. Psychoneuroendocrinology 37:589-601.
6. Cieszyński Ł, Jendrzejewski J, Wiśniewski P, Owczarzak A, Sworczak K. 2019. Hair cortisol concentration in a population without hypothalamic-pituitary-adrenal axis disorders. Adv Clin Exp Med 28:369-373.
7. Greff MJE, Levine JM, Abuzgaia AM, Elzagallaai AA, Rieder MJ, van Uum SHM. 2019. Hair cortisol analysis: An update on methodological considerations and clinical applications. Clinical Biochemistry 63:1-9.
8. Karlén J, Ludvigsson J, Frostell A, Theodorsson E, Faresjö T. 2011. Cortisol in hair measured in young adults - a biomarker of major life stressors? BMC Clinical Pathology 11:12.
9. Sauvé B, Koren G, Walsh G, Tokmakejian S, van Uum SHM. 2007. Measurement of cortisol in human hair as a biomarker of systemic exposure. Clinical & Investigative Medicine 30:E183-E191.
10. Raul JS, Cirimele V, Ludes B, Kintz P. 2004. Detection of physiological concentrations of cortisol and cortisone in human hair. Clin Biochem 37:1105-11.

11. Gonzalez D, Jacobsen D, Ibar C, Pavan C, Monti J, Fernandez Machulsky N, Balbi A, Fritzler A, Jamardo J, Repetto EM, Berg G, Fabre B. 2019. Hair Cortisol Measurement by an Automated Method. Sci Rep 9:8213.
12. Dajani R, Hadfield K, van Uum S, Greff M, Panter-Brick C. 2018. Hair cortisol concentrations in war-affected adolescents: A prospective intervention trial. Psychoneuroendocrinology 89:138-146.
13. Hodes A, Lodish M, Tirosh A, Meyer J, Belyavskaya E, Lyssikatos C, Rosenberg K, Demidowich A, Swan J, Jonas N, Stratakis C, Zilbermint M. 2017. Hair cortisol in the evaluation of Cushing syndrome. Endocrine 56.
14. Loussouarn G. 2001. African hair growth parameters. British Journal of Dermatology 145:294-297.
15. Moody SN, van Dammen L, Wang W, Greder KA, Neiderhiser JM, Afulani PA, Willette A, Shirtcliff EA. 2022. Impact of hair type, hair sample weight, external hair exposures, and race on cumulative hair cortisol. Psychoneuroendocrinology 142:105805.
16. Wester VL, van der Wulp NRP, Koper JW, de Rijke YB, van Rossum EFC. 2016. Hair cortisol and cortisone are decreased by natural sunlight. Psychoneuroendocrinology 72:94-96.
17. Meyer JS, Novak MA. 2012. Minireview: Hair cortisol: a novel biomarker of hypothalamic-pituitary-adrenocortical activity. Endocrinology 153:4120-7.
18. Binz TM, Gaehler F, Voegel CD, Hofmann M, Baumgartner MR, Kraemer T. 2018. Systematic investigations of endogenous cortisol and cortisone in nails by LC-MS/MS and correlation to hair. Anal Bioanal Chem 410:4895-4903.
19. Izawa S, Miki K, Tsuchiya M, Mitani T, Midorikawa T, Fuchu T, Komatsu T, Togo F. 2015. Cortisol level measurements in fingernails as a retrospective index of hormone production. Psychoneuroendocrinology 54:24-30.
20. Meyer JS, Novak MA. 2021. Assessment of prenatal stress-related cortisol exposure: focus on cortisol accumulation in hair and nails. Dev Psychobiol 63:409-436.
21. Izawa S, Sugaya N, Ogawa N, Shirotsuki K, Nomura S. 2021. A validation study on fingernail cortisol: correlations with one-month cortisol levels estimated by hair and saliva samples. Stress 24:734-741.
22. Sami A, Elimairi I, Patangia D, Watkins C, Ryan CA, Ross RP, Stanton C. 2021. The ultra-structural, metabolomic and metagenomic characterisation of

the sudanese smokeless tobacco 'Toombak'. Toxicology Reports 8:1498-1512.

23. Anderson SM, Brunzell DH. 2015. Anxiolytic-like and anxiogenic-like effects of nicotine are regulated via diverse action at β2*nicotinic acetylcholine receptors. Br J Pharmacol 172:2864-77.
24. Dean R, Duperreault E, Newton D, Krook J, Ingraham E, Gallup J, Franczak BC, Hamilton TJ. 2020. Opposing effects of acute and repeated nicotine exposure on boldness in zebrafish. Sci Rep 10:8570.
25. Ebbert JO, Patten CA, Schroeder DR. 2006. The Fagerström Test for Nicotine Dependence-Smokeless Tobacco [FTND-ST]. Addictive behaviors 31:1716-1721.
26. Mushtaq N, Huque R, Beebe LA, Shah S, Siddiqi K. 2019. Evaluation of tobacco dependence measures in South Asian smokeless tobacco users. Drug Alcohol Depend 203:66-71.
27. Parashar M, Agarwalla R, Mallik P, Dwivedi S, Patvagekar B, Pathak R. 2016. Prevalence and correlates of nicotine dependence among construction site workers: A cross-sectional study in Delhi. Lung India 33:496-501.
28. Sushanthi S, Doraikannan S, Indiran MA. 2022. Assessment of anxiety, depression and nicotine dependence among construction workers in Chennai - A cross sectional study. J Oral Biol Craniofac Res 12:263-267.
29. Rodu B, Plurphanswat N, Fagerström K. 2015. Time to First Use Among Daily Smokers and Smokeless Tobacco Users. Nicotine Tob Res 17:882-5.
30. Mushtaq N, Beebe LA. 2017. Psychometric Properties of Fagerström Test for Nicotine Dependence for Smokeless Tobacco Users [FTND-ST]. Nicotine & Tobacco Research 19:1095-1101.
31. Subedi K, Shrestha A, Bhagat T. 2021. Assessment of nicotine dependence among tobacco users visiting outreach programs in Dharan, Nepal: a cross-sectional study. BMC Public Health 21:1515.
32. Shrestha A, Marla V, Lamsal M, Sarraf DP, Rao N. 2021. Evaluation of Nicotine Dependence in Adult Population of Eastern Nepal. Kathmandu Univ Med J [KUMJ] 19:210-215.
33. Cui BR, Zhang BB, Chu CP, Cui X, Qiu DL. 2016. Nicotine excites corticotropin-releasing hormone mRNA-expressing neuron in the hypothalamic paraventricular nucleus in vitro in rats. Neuroreport 27:580-6.

34. Rovaris DL, Mota NR, de Azeredo LA, Cupertino RB, Bertuzzi GP, Polina ER, Contini V, Kortmann GL, Vitola ES, Grevet EH. 2013. MR and GR functional SNPs may modulate tobacco smoking susceptibility. Journal of Neural Transmission 120:1499-1505.
35. Lovallo WR. 2006. Cortisol secretion patterns in addiction and addiction risk. Int J Psychophysiol 59:195-202.
36. Miller GE, Chen E, Zhou ES. 2007. If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. Psychol Bull 133:25-45.
37. Raffetti E, Landgren AJ, Andersson F, Donato F, Lavebratt C, Forsell Y, Galanti MR. 2021. Cortisol Concentration as Predictor of Tobacco Initiation in Adolescents: Results From a Population-Based Swedish Cohort. Journal of Adolescent Health 68:758-764.
38. Ginty AT, Jones A, Carroll D, Roseboom TJ, Phillips AC, Painter R, de Rooij SR. 2014. Neuroendocrine and cardiovascular reactions to acute psychological stress are attenuated in smokers. Psychoneuroendocrinology 48:87-97.
39. Sharpley CF, McFarlane JR, Slominski A. 2011. Stress-linked cortisol concentrations in hair: what we know and what we need to know. Rev Neurosci 23:111-21.
40. Reuter M, Hennig J. 2003. Cortisol as an indicator of dopaminergic effects on nicotine craving. Hum Psychopharmacol 18:437-46.
41. Netter P, Toll C, Lujic C, Reuter M, Hennig J. 2002. Addictive and nonaddictive smoking as related to responsivity to neurotransmitter systems. Behav Pharmacol 13:441-9.
42. Mello NK. 2010. Hormones, nicotine, and cocaine: clinical studies. Horm Behav 58:57-71.
43. Banoo SN. 2019. Culturally Accepted Smokeless and Chewing Tobacco Use among South Asian Immigrants in the USAThe Chicago School of Professional Psychology.
44. Chaturvedi S, Somashekar B, Manjunatha N, Devendran Y, Naik S. 2021. Stress and Struggles: The Comprehensive book of Stress, Mental health and Mental Illness.
45. Newton DF. 2019. Peripheral Biomarkers of Tobacco-Use Disorder: A Systematic Review. medRxiv:19004150.

46. Biswas S, Manna K, Das U, Khan A, Pradhan A, Sengupta A, Bose S, Ghosh S, Dey S. 2015. Smokeless tobacco consumption impedes metabolic, cellular, apoptotic and systemic stress pattern: A study on Government employees in Kolkata, India. Scientific Reports 5:18284.
47. Liu N-A, Araki T, Cuevas-Ramos D, Hong J, Ben-Shlomo A, Tone Y, Tone M, Melmed S. 2015. Cyclin E-Mediated Human Proopiomelanocortin Regulation as a Therapeutic Target for Cushing Disease. The Journal of Clinical Endocrinology & Metabolism 100:2557-2564.
48. Bagchi M, Bagchi D, Hassoun EA, Stohs SJ. 1998. Subchronic effects of smokeless tobacco extract [STE] on hepatic lipid peroxidation, DNA damage and excretion of urinary metabolites in rats. Toxicology 127:29-38.
49. Fries E, Hesse J, Hellhammer J, Hellhammer DH. 2005. A new view on hypocortisolism. Psychoneuroendocrinology 30:1010-6.
50. Koumantarou Malisiova E, Mourikis I, Darviri C, Nicolaides NC, Zervas IM, Papageorgiou C, Chrousos GP. 2021. Hair cortisol concentrations in mental disorders: A systematic review. Physiology & Behavior 229:113244.
51. Roerink ME, Roerink S, Skoluda N, van der Schaaf ME, Hermus A, van der Meer JWM, Knoop H, Nater UM. 2018. Hair and salivary cortisol in a cohort of women with chronic fatigue syndrome. Horm Behav 103:1-6.
52. Rohleder N. 2018. Burnout, hair cortisol, and timing: Hyper- or hypocortisolism? Psychoneuroendocrinology 87:215-217.
53. Lovallo WR, Farag NH, Sorocco KH, Cohoon AJ, Vincent AS. 2012. Lifetime Adversity Leads to Blunted Stress Axis Reactivity: Studies from the Oklahoma Family Health Patterns Project. Biological Psychiatry 71:344-349.
54. Lovallo WR, Enoch M-A, Sorocco KH, Vincent AS, Acheson A, Cohoon AJ, Hodgkinson CA, Goldman D. 2017. Joint impact of early life adversity and COMT Val158Met [rs4680] genotypes on the adult cortisol response to psychological stress. Psychosomatic medicine 79:631.
55. Steudte S, Stalder T, Dettenborn L, Klumbies E, Foley P, Beesdo-Baum K, Kirschbaum C. 2011. Decreased hair cortisol concentrations in generalised anxiety disorder. Psychiatry Res 186:310-4.
56. Staufenbiel SM, Penninx BWJH, Spijker AT, Elzinga BM, van Rossum EFC. 2013. Hair cortisol, stress exposure, and mental health in humans: A systematic review. Psychoneuroendocrinology 38:1220-1235.

57. Stea TH, Shatri H, Haugland SH, Kleppang AL. 2022. Association between Self-Reported Childhood Difficulties and Obesity and Health-Related Behaviors in Adulthood—A Cross-Sectional Study among 28,047 Adults from the General Population. International Journal of Environmental Research and Public Health 19:1395.
58. Lovallo WR, Cohoon AJ, Sorocco KH, Vincent AS, Acheson A, Hodgkinson CA, Goldman D. 2019. Early-Life Adversity and Blunted Stress Reactivity as Predictors of Alcohol and Drug use in Persons With COMT [rs4680] Val158Met Genotypes. Alcoholism: Clinical and Experimental Research 43:1519-1527.
59. Fischer S, Skoluda N, Ali N, Nater UM, Mewes R. 2022. Hair cortisol levels in women with medically unexplained symptoms. J Psychiatr Res 146:77-82.
60. Reynolds R, Walker B, Haw S, Newby D, Mackay D, Cobbe S, Pell A, Fischbacher C, Pringle S, Murdoch D, Dunn F, Oldroyd K, Macintyre P, O'Rourke B, Pell J. 2010. Low serum cortisol predicts early death after acute myocardial infarction. Critical care medicine 38:973-5.
61. Sharma P, Sandhu SV, Bhandari R, Verma I, Bhullar RK, Khangura RK. 2018. Estimation of cortisol levels in patients with premalignant disorders and oral squamous cell carcinoma. J Oral Maxillofac Pathol 22:27-34.
62. Choi D, Ota S, Watanuki S. 2015. Does cigarette smoking relieve stress? Evidence from the event-related potential [ERP]. Int J Psychophysiol 98:470-6.
63. Iglesias S, Jacobsen D, Gonzalez D, Azzara S, Repetto EM, Jamardo J, Gómez SG, Mesch V, Berg G, Fabre B. 2015. Hair cortisol: A new tool for evaluating stress in programs of stress management. Life Sci 141:188-92.
64. Goldberg SB, Manley AR, Smith SS, Greeson JM, Russell E, Van Uum S, Koren G, Davis JM. 2014. Hair cortisol as a biomarker of stress in mindfulness training for smokers. J Altern Complement Med 20:630-4.
65. Ussher M, West R, Evans P, Steptoe A, McEwen A, Clow A, Hucklebridge F. 2006. Reduction in cortisol after smoking cessation among users of nicotine patches. Psychosom Med 68:299-306.
66. Benowitz NL. 2008. Neurobiology of nicotine addiction: implications for smoking cessation treatment. Am J Med 121:S3-10.

3 The fermented foods of Sudan through a metagenomic lens

Amel Sami 1,2, Imad Elimairi 2, *, Catherine Stanton 1,3, Dhrati Patangia 1, R. Yasmeen ElYass [4], Marmar Elsiddig [4], Heyam Salih [4] and R Paul Ross 1,

1 APC Microbiome Ireland, School of Microbiology, University College Cork, Cork T12 YN60, Ireland.
2 Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, National Ribat University, Nile Street, Khartoum 1111, Sudan
3 Teagasc Food Research Centre, Moorepark, Fermoy, Cork P61 C996, Ireland
4 Department of Microbiology, Khartoum

This chapter is published in the Boolean Journal, University College Cork, Boolean Journal, December 2022 [see Appendix 5].

3.1Abstract

The fermented foods of Sudan are uniquely preserved in their methodologies and offer an invaluable source of affordable nutrition to the Sudanese population. These foods originate from a vast range of raw ingredients. In this study, we sought to analyse the most extensive collection of Sudanese fermented foods from the following categories: crop, plant, animal, fish, and dairy produced foods.

Forty-six samples of Sudanese fermented foods were collected from various locations that included markets in Khartoum and Western Sudan as well as being sourced from local homes. The un-fermented Combu or potash that is often added to many Sudanese fermented foods was also collected. The microbiome and mycobiome were analysed through 16S rRNA and internal transcriber [ITS] sequencing, respectively at different stages of these foods production. DNA extraction was achieved, and polymerase chain reaction followed by the Illumina sequencing protocol. Data were processed by Microbiome analyst, Calypso version 8.84 and BLASTn for species recognition.

We found a vast array of bacterial and fungal composition amongst Sudanese fermented foods that have the potential for numerous health benefits. Beta diversity [p=0.001] was significantly varied between the categories of Sudanese fermented foods. Animal, fish, and plant-based foods exhibited a high alpha diversity richness [p=8.7229e-07] while fish and plant-based foods had an elevated Shannon index [p=0.001]. Crop-based foods [sorghum and millet] were found to be enriched in the genus *Acetobacter* [96.89%], plant-based foods in *Aeriscardovia* [99.14%], dairy-based foods in *Enhydrobacter* [85.03%], animal-based foods in *Bacteroide*s [98.22%] and fish-based foods in *Sporosarcina* [99%]. Although un-fermented, the Combu was found to contain an enrichment in *Lactobacillus* [23%]. Evaluation of the preparatory and final stages of sorghum and millet fermented foods showed a loss of *Lactobacillus* abundance however from 26.91% to 4.38% when comparing the former and latter stages. The Kawal food, obtained from the fermentation of the Cassia obtusifolia plant was found to harbour an enriched microbiome [*Bacillus, Lactobacillus, Pediococcus, Citrobacter* and *Bifidobacterium*] and mycobiome [> 90 fungal species] that included but was not limited to *Candida, Aspergillus, Cladosporium* and *Malassezia* species which could potentially harbour unknown but novel probiotic benefits to consumers. Sudanese fermented foods were however also found to be a

source for pathogenic bacteria that included *Escherichia_Shigella* which was found abundantly in the plant-based foods [46.8%] and *Wohlfahrtiimonas* which were found to be abundant in animal-based foods [79.07%].

Sudanese fermented foods are rich in beneficial microorganisms but also carry a sinister microbial picture due to external contamination. However, the methodological preservation of these foods should be strongly sought with insight into more modern techniques that improve their biological standards.

3.2 Introduction

Fermented foods of Sudan are numerous in variety, carrying indigenous origins dating back centuries. These foods often have preserved forms of methodologies and in some societies within Sudan they act as a vital source of survival offering nutrition in times of food shortages and famine [1]. Indeed fermentation plays an essential role in stabilising food security in Sudan by preserving key raw materials and modifying foods so that they are fit for consumption [2].

Sudanese fermented food production is largely influenced by the rich tribal backgrounds of the Sudanese people, climate variations such as drought, economic instability, and the migration or movement of people. While some fermented foods are consumed all over Sudan, such as the sorghum pancake or Kisra, others are unique to specific geographic locations, particularly the Western part of the country, that include the fermented foods, Kawal and Sigda, produced from the Cassia obtusifolia and sesame plants respectively. Many of these foods vary in taste, palatability, and form. Some have a soft texture such as the fat sheath product Miriss while others are hard such as the Musran or dried intestine.

Sudanese fermented foods originate from a vast source of starting ingredients that can be classified into crop-based fermented foods that include sorghum or millet and plant-based fermented foods such as the Cassia obtusifolia, sesame and hibiscus plants. Fruit and vegetable-based fermented foods include sources such as watermelon, cucumbers, dates, boaob, pumpkin, onions, and mushroom while animal-based fermented foods are sourced from the intestines, fat sheaths, bones or meat of the animal. Fish-based fermented foods are produced from variously sized fish while dairy-based Sudanese fermented foods are rich in heritage and include yoghurts, cheeses, thickened milk and biscuit forms [3].

While several microbiological studies have been achieved to date on Sudanese fermented foods [4-6], there are limited reports using next generation sequencing techniques [7] with many

Sudanese fermented foods such as the Musran, Miriss and Shermout never previously explored through metagenomics. In this study, we performed metagenomic sequencing on the most extensive collection of Sudanese fermented food samples to date, using 16SrRNA and ITS sequencing for microbiome and mycobiome analysis, respectively. Table 1 highlights a summary of the Sudanese fermented foods in this study outlining each of the food's production steps.

Table1: Summary of Sudanese fermented foods

Fermentation source	**Food name and type**	**Summary of methodological production**
Sorghum	Kisra - sorghum pancakes	Water to sorghum flour ratio is often 1:1. The mixture is covered and left for 12-19 hours in warm temperatures. Kisra is baked on hot flat plates for no longer than 15-20 seconds for each pancake. Final product is a large, smooth, coherent pliable paper-thin sheet that is sour in flavour.
	Abreh -a flavourful sweet-bitter drink	Two types of sorghum are used [cooked sorghum grain as a porridge and ground germinated grain]. The ground germinated grain is added to the hot porridge and left to ferment for around 12 hours. Liquefaction occurs. Spices are added such as cinnamon, ginger, cloves and cardamon. Dates are also added. The mixture is left to ferment for around another 12 hours. Abreh is baked on flat plates as thin red sheets, further air-dried, and stored for up to one year but can be more. When required for use, a sheet is resuspended in water for 5-6 hours, sieved and a clear juice is prepared. Sugar is often added.
Millet	Damirga – can be made into a pancake, porridge, or drink	Pancakes - millet flour is mixed with water and allowed to ferment with similar preparation methods to sorghum-based pancakes Porridge – millet flour is also allowed to ferment after which it is heated into a thick porridge Damirga drink – this is obtained from the fermented water of pancake or porridge preparation
Plant	Kawal – fermented leaf protein source	Only unwashed Cassia Obtusifolia leaves are utilised. No starter is utilised.

	eaten in solitude or as a stew [predominantly consumed by the Western Sudanese population]	An earthenware pot is placed underground. A mucilaginous substance [okra] is slathered on the surface to prevent mixing between Kawal and soil. Fresh leaves are crushed into a light paste and placed into the pot away from direct sunlight. Sorghum leaves may be placed as a further covering. The pot is sealed and left for 3-4 days. The paste begins to turn dark brown/black. It is re-mixed and covered for another 15 days. The final product is a thick paste that is shaped into even balls and air-dried [often another 3-4 days] before storing for up to year. A yellow 'pickle sauce' also develops like soy sauce and can be utilised in food preparation which is said to have as high amino acid content.
	Sigda – fermented sesame seed balls eaten as a stew or whole	Sesame seeds are crushed with water and left to ferment for 3-7 days in a tightly sealed container. Combu* is often added. Occasional mixing and resealing occur during this time. The final product is moulded into balls and sundried.
Dairy	Mish - deep yoghurt	Milk to be fermented is added to a clean container and a starter of 'Mish' is added. Starters are often many years old and secluded to families. Black cumin and fenugreek seeds are added. The Mish is left at room temperatures for two days to complete fermentation. The final product is then placed under cool storing to stop fermentation and is often consumed for up to one month.
	Garis – fermented camel milk	Fermented camel milk. Previous starter of Garis is often added to new camel milk. Shaking is then introduced

	Gergosh – dairy/legume biscuit	Milk is boiled to which lentils are added. The mixture is left for 15-20 hours. Gaseous fermentation occurs. The now fermented milk is then mixed with wheat flour. Water and nigella seeds are also incorporated. The mixture is made into a dough and baked. Gergosh is used as a morning snack and can keep for years without spoiling.
Fish	Faseekh – moist salted fish	Production is initiated in October/November until May-June Fish are cleaned and salted individually. They are stacked in layers in wooden racks and protected from air by wrapping after which they are left for 3-7 days depending on ambient temperatures [shorter time in summer and longer in winter] The fish now become semi-dry with a pungent aroma. They are packed into tighter airtight containers for another 3-7 days to produce Faseekh.
	Mendeshi – pasted fish of medium sized origin	This product is usually produced from the Mormyrus genus of fish that are small-medium sized fish [7-12cm] often chosen around December and January. Fish are left to dry then either left whole or made into a paste.
	Kejeik – sun-dried fermented fish often of a large size that maybe broken into pieces before selling	Fish are split longitudinally before drying for days to weeks.
Meat	Shermout – air-dried meat [beef] strips	Meat containing fat is often chosen. Meat is cut into strips and hung on ropes left to sun dry at hot temperatures.

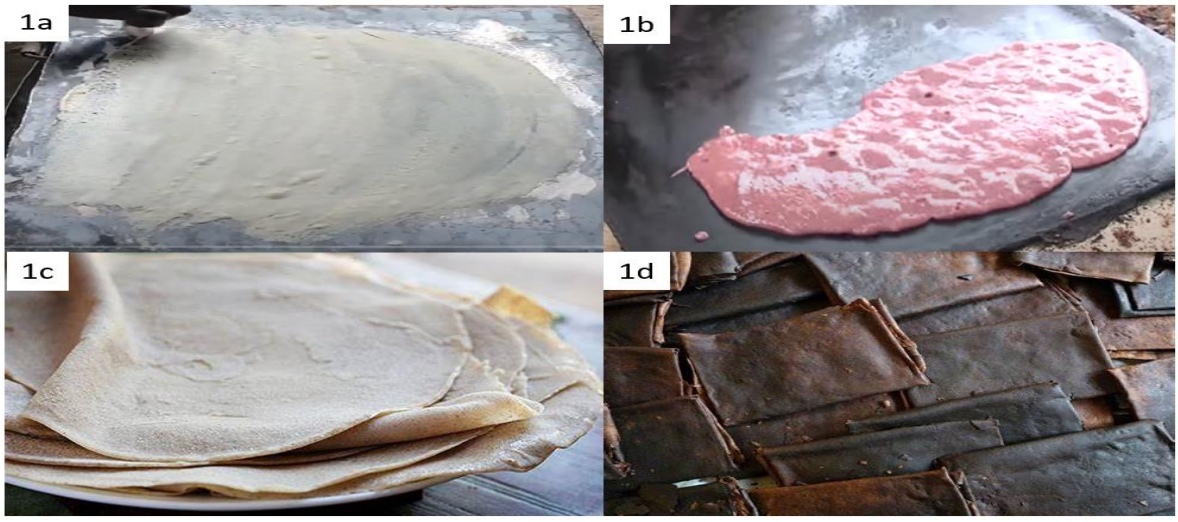
	Miriss – fat sheath of sheep	Made from animal fat [sheep]. The fat is cleaned and made into a paste. It is then placed in a tightly closed bucket for 2-6 days. Sometimes various pieces of intestine may be added. The final product is a white pungent-smelling paste. Can be stored at room temperature for up to a year.
	Musran – dried intestine usually sheep	Small intestine contents are emptied and placed in closed container to undergo fermentation for 3-4 days. After this time, they are spread out in the sun to dry.
*	Combu	This is produced by the burning of sorghum seed heads with mixing of ash and water. The filtrate is collected and left to dry. The end product is a grey crystalline mass [82].

3.3 Materials and methods

Sample summary

Forty-six Sudanese fermented food samples were collected from various markets across Sudan [Omdurman and Khartoum] while some samples were collected from local homes and the west of the country, Darfur. Thirteen samples were obtained from two sorghum products, Kisra and Abreh during their preparation and final stage [Figures 1a-1d]. Two samples were of a millet product, Damirga, nine samples were from the Cassia obtusifolia plant fermented product, Kawal [Figure 1e] and one sample was of the fermented sesame seed block, Sigda [Figure 1f]. A further eight samples were the fish fermented products, Kajeik [Figure 1g], Mendeshi [Figure 1h] and Faseekh while six samples were of the animal fermented foods, Shermout [Figure 1i], Musran [Figure 1j] and Miriss [Figure 1k]. Seven dairy fermented food samples were included, one of which was from the deep yoghurt Mish [fl], two samples were from the fermented camel's milk, Garis, and four samples were from the preparatory and final stages of the dairy/legume-based biscuit, Gergosh.

Figures 1a-1l: Types of Sudanese fermented foods



Dry fermented products of Sudan. 1a: preparation stage of sorghum pancake. 1b: preparation stage of sorghum juice. 1c: Final sorghum pancake that is ready to eat. 1d: sorghum juice sheets which require soaking in water for several hours and sieving after which the juice is consumed by many Sudanese. 1e. Kawal produced from the Cassia obtusifolia in Western Sudan. 1f: Sigda, a sesame seed fermented product 1g: Kajeik fish fermented samples and 1h: Mendeshi that are medium sized fermented fish. 1i: Shermout or air-dried fermented shredded beef. 1j: Musran or fermented intestine. Moist fermented foods of Sudan. 1k: Miriss produced from the fat sheath of sheep. 1l: Mish or deep yoghurt. Black nigella seeds are often added for taste.

Finally, one sample of Combu or potash, although not a fermented product, was collected. Combu is often added during the preparation of many Sudanese fermented foods. Some samples were stored at room temperature such as the Kawal, Sigda, Abreh, Shermout and

Musran while others were placed in a refrigerated setting such as the Miriss, Mish and fish samples.

Sorghum and millet fermented foods

Sorghum foods included the sorghum pancake, Kisra and sorghum juice, Abreh. Five samples were obtained during the pre and post spice preparatory stages of sorghum juice while one sample was of the final stage. Six sorghum pancake preparations were included for sequencing while one sample was analysed at the final stage of sorghum pancake. Two millet products, the millet flour preparation, Damirga and its fermentation water that is a common addition to the diet of the people of Western Sudan, were also included.

Plant fermented foods

Nine samples of the fermented Cassia obtusifolia product, Kawal, were collected. Two of these samples were retrieved from Western Sudan while another two and five samples were obtained from Khartoum and Omdurman markets respectively. One sample of the sesame seed fermentation block, Sigda was obtained from a Khartoum market.

Animal fermented foods

Six samples were analysed in this category. Four samples were of the fat sheath fermented product; Miriss, one sample was of the fermented air–dried meat product, Shermout and one sample was obtained of the fermented hardened intestine, Musran.

Fish fermented foods

Eight samples were analysed in this subset. These included one sample from the moist fish fermentation product, Faseekh, five samples from the sun-dried fish product Kajeik and two samples from the medium sized pasted fish product, Mendeshi.

Dairy fermented foods

Seven samples were included of which four samples were of the preparatory and final stages of the dairy/legume biscuit, Gergosh, two samples were of the fermented camel's milk, Garis and one sample was that of deep yoghurt, Mish.

Metagenomic analysis

DNA extraction

DNA extraction of the fermented food samples was achieved with the WizPrep™ Plant DNA Mini kit [WizBiosolutions inc] that allows for the isolation of DNA from plant material including leaves, stems and seeds as well as other tissue. This was carried out in Khartoum, Sudan. Briefly, 100mg of each sample was weighed and vortexed with 400 µl lysis buffer and 5µl RNase A. 500 µl of Cetyl Trimethyl Ammonium Bromide [CTAB] buffer was added to samples with rigorous starting material and included the Kawal, fish and Shermout samples. All samples were then incubated at 65°C for 10 minutes, a further 100µl of buffer was vortexed with all samples which were then incubated on ice for 3 minutes. The lysate was transferred to filter columns and centrifuged for 1 minute at 18,000 X g at 12°C. The supernatant was transferred to a collection tube and each sample underwent a further binding step with 600µL buffer after which they were placed in spin columns and centrifuged for a further 2 minutes. After discarding the flow through, samples were washed with 400µL and 600µL buffers respectively and centrifuged during each step for 30 seconds at 18,000 X g at 12°C. Finally, samples were eluted with a 50µL elution buffer and the purified DNA samples were stored at -80 °C in Eppendorf tubes for further analysis.

Bacterial amplicon sequencing

To target the 16S V3–V4 hypervariable regions, standard IUPAC nucleotide nomenclature and universal forward and reverse primers with overhang adapters were utilised on DNA extracted samples. Amplicon polymerase chain reaction [PCR] was achieved using AMPure XP beads [Beckman Coulter]. PCR was then performed in a 50 µl reaction mixture containing 10ng of template DNA and 2x KAPA HiFi HotStart Ready Mix. The following thermal cycling conditions were used: 3 min of initial denaturation at 95°C; 30 cycles each of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 30 s; and a last step at 72°C for 5 min. PCR clean up using AMPure XP beads [Beckman Coulter] to purify the 16S V3 and V4 amplicon from free primers and primer dimer species was then achieved.

Electrophoresis was performed on the PCR amplified products using 1.5% agarose gel with Invitrogen DNA loading dye [ThermoFisher scientific] for size verification [~550 base pairs]. The gel was then visualised under 300 nm ultraviolet light. Index PCR was then continued by attaching dual indices and Illumina sequencing adapters using Nextera XT index kits and the following PCR thermal cycling conditions were used: 3 min of initial denaturation at 95°C; 8

cycles each of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 30 s; and a last step at 72°C for 5 min. AMPure XP beads [Beckman Coulter] were utilised for final clean-up and the final library quantified using the Qubit® 2.0 Fluorometer and the Qubit dsDNA HS Assay kit [ThermoFisher Scientific]. Samples were then normalised, and the final library run on an Agilent high sensitivity chip and quantified by qPCR using the KAPA Illumina Library quantification kit.

Fungal amplicon sequencing

DNA amplification was performed on the ITS with primers that were ITS1-ITS2 region specific and followed a similar protocol to Walsh et al. [2016] [8]. Primers were modified to incorporate the Illumina overhang adaptor [ITS1F primer 5′TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTGGTCATTTAGAGGAAGTAA 3′ and ITS2 primer 5′GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTGCGTTCTTCATCGATGC 3′]. Polymerase chain reaction products were then treated as described in the Illumina protocol. All samples were sequenced on the Illumina MiSeq with a 2 × 250 cycle in accordance with Illumina sequencing instructions.

Microbiome bioinformatics

Samples were sequenced using the V3-V4 regions of the 16S rRNA gene. The 300 bp paired end FASTQ product generated from 16S sequencing were merged using FLASH [fast length adjustment of short reads] using default parameters [9]. QIIME's split_libraries_fastq.py script was used for demultiplexing and filtering of the FASTQ sequence data. Further quality filtering was performed using the USEARCH analysis tool. Briefly, single unique reads were removed; following denoising, chimera removal and grouping into operational taxonomic units [OTUs] at 97% similarity was performed using USEARCH v7 [64 bit] [10]. Alignment of OTUs was performed using PyNAST [python nearest alignment space termination], a flexible tool for aligning sequences to a template alignment [11]. Further, assigning of taxonomic ranks was performed by using BLAST against the SILVA SSURef database release 132 [12].

Mycobiome bioinformatics

The primer pair internal transcribers, ITS1F and ITS2 were used for performing fungal sequencing. The resulting sequencing reads were highly variable in length and were processed using DADA2 [v1.18.0] in R, following the DADA2 pipeline workflow for ITS sequences

[13]. Briefly, the reads were filtered to remove sequences with ambiguous 'N' bases. This was followed by primer removal from the reads using Cutadapt [v3.5] to remove primers as described in the dada2 ITS pipeline workflow [14]. The quality of the reads was inspected, and the reads were further filtered and trimmed to retain reads that were minimum 50 base pair long with default parameters. Dereplicating and merging of paired end reads was done following the workflow, following which chimeras were removed and taxonomy was assigned to the resulting ASVs [Amplicon Sequence Variants] using the UNITE ITS database.

Statistical analysis

The resulting BIOM and ASV table was used for further analysis with microbiome analyst CA. Low count filtering was performed to remove features with very small counts due to sequencing errors and was achieved at 10% prevalence filter [at least 10% of its values should contain at least 4 counts]. Total sum normalisation aimed to address variability in sampling depth and was carried out on samples. Data were not rarefied. Alpha diversity was assessed using unfiltered data.

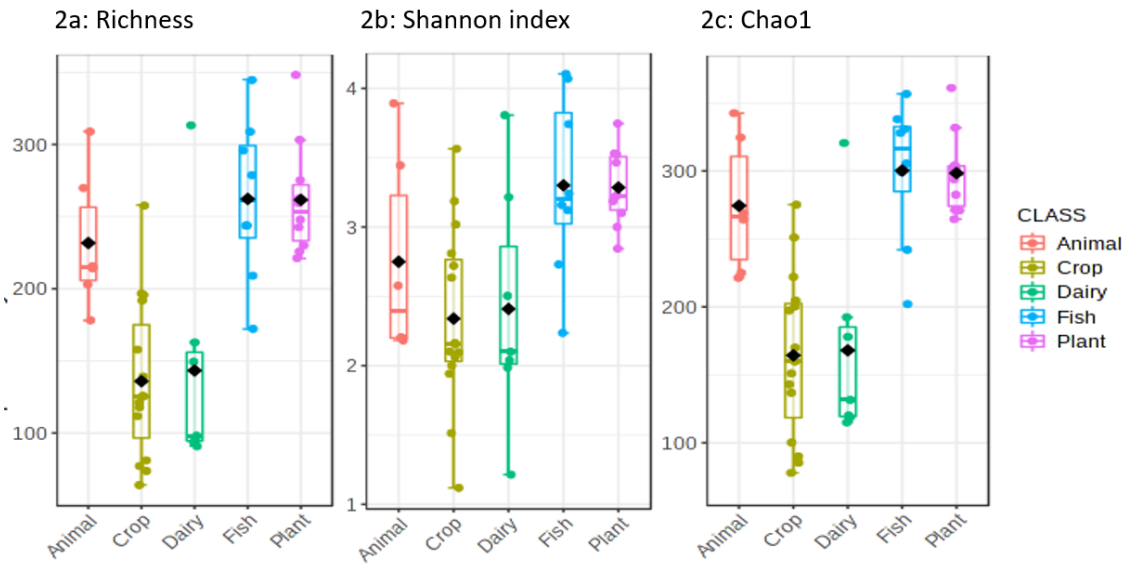
3.4 Results

Microbial variations of Sudanese fermented foods by category composition [animal, crop, dairy, fish, and plant fermented foods]

Alpha and Beta diversity of Sudanese fermented foods

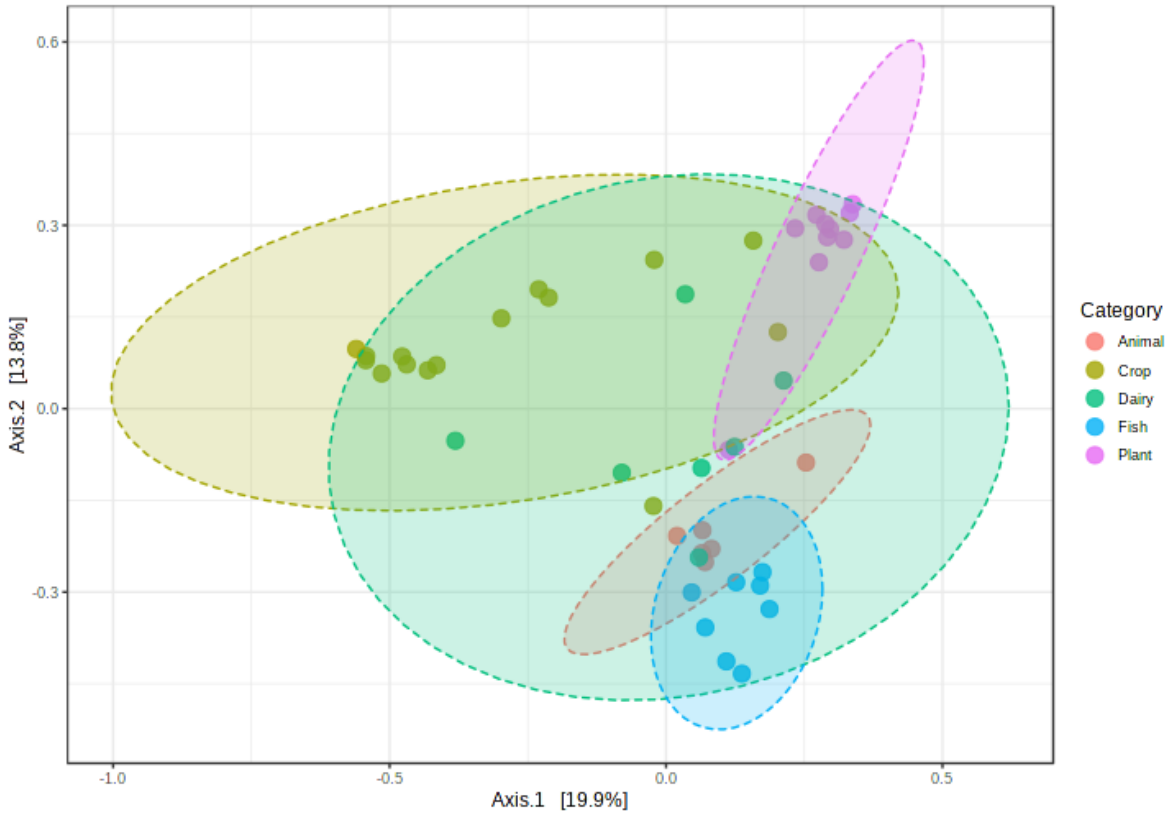
Significant alpha diversity variations could be observed between the food categories. Richness was found to be highest in the animal, fish, and plant-based foods and lowest in the crop and dairy-based foods [p=8.7229e-07] [Figure 2a], while Shannon index was elevated in the fish and plant-based foods [p=0.001] [Figure 2b]. Chao1, a measure of rare genera findings was also high in animal, fish, and plant-based foods compared to crop and dairy-based foods [p=7.9126e-08] [Figure 2c]. Meanwhile, significant Beta diversity [p=0.001] was found between the Sudanese fermented food categories, albeit crop and dairy categories could be seen to share some elliptical similarities [Figure 2d].

Figures 2a-2c:



2a-2c: α diversity measures. 2a: Richness was highest in the animal, fish, and plant-based foods and lowest in the crop and dairy-based foods [p=8.7229e-07]. 2b: Shannon index was highest in the fish and plant-based foods [p=0.001]. 2c: Chao1, a measure of rare genera findings indicated animal, fish, and plant-based foods to have the most increased α measure compared to crop and dairy based foods [p=7.9126e-08].

Figure 2d:



2d: β diversity of Sudanese fermented foods highlights significant diversity between food types [p=0.001] albeit, crop and dairy based foods shared some elliptical resemblance.

There were 1393 OTUs in the library overview with 3472282 reads. A total of 654 low abundance features were removed based on prevalence. The number of features remaining after data filtering were 461. Relative abundance of top genera was expressed in percentages and evaluated under the categories: plant, crop, dairy, fish, and animal- based foods [Table 2].

In plant fermented foods [Kawal and Sigda], abundant genera included *Aeriscardovia* [99.14%], *Brevibacillus* [96.98%], *Bifidobacterium* [94.42%], *Pediococcus* [91.27%], *Bacillus* [89.11%], *Enterococcus* [87.41%], *Corynebacterium_1* [72.22%], *Enterobacter* [61.36%], *Pantoea* [52.79%], *Citrobacter* [51.07%], and *Acinetobacter* [47%].

In crop fermented foods [sorghum and millet], the most abundant genera were found to be *Acetobacter* [96.89%], *Lactobacillus* [63.87%], *Kosakonia* [51.10%] and *Lachnoclostridium_5* [43.67%]. *Actinomyces* was similar in abundance in both the crop [37.72%] and dairy [34.06%] fermented foods.

Genera in most abundance in dairy fermented foods compared to the other categories were *Enhydrobacter* [85.03%], *Pseudomonas* [68%], *Streptococcus* [50.15%], and *Prevotella_7* [50.33%].

In animal fermented foods, the genera in abundance included *Bacteroides [98.22%], Oblitimonas* [98.86%], *Kerstersia* [97.68%], *Peptoniphilus* [87.66%], *Wohlfahrtiimonas* [79.07%], *Lysinibacillus* [72.99%], *Vagococcus* [69.13%], *Ignatzschineria* [68.61%] and *Providencia* [61.76%]. The genera *Sporosarcina, Cetobacterium, Sarcina, Savagea* and *Psychrobacter* were found to be highly abundant in fish fermented foods at 99%, 97.23%, 91.13%, 97.11% and 62.61% abundances respectively. *Aeromonas* was highest in abundance in the fish samples [56.16%] followed by dairy [24.40%] and animal fermented foods [17.89%].

We further found species variations between the categories of Sudanese fermented foods. *Lactobacillus ginsenosidimutans* [q=0.018], was found to be highest in the animal fermented foods, *Lactobacillus acidipiscus* [q=0.0099] and *dextrinicus* [q=0.01] were abundant in plant-based foods, while *Lactobacillus pontis* were enriched in crop-based foods [q=0.008] [Figure 2e].

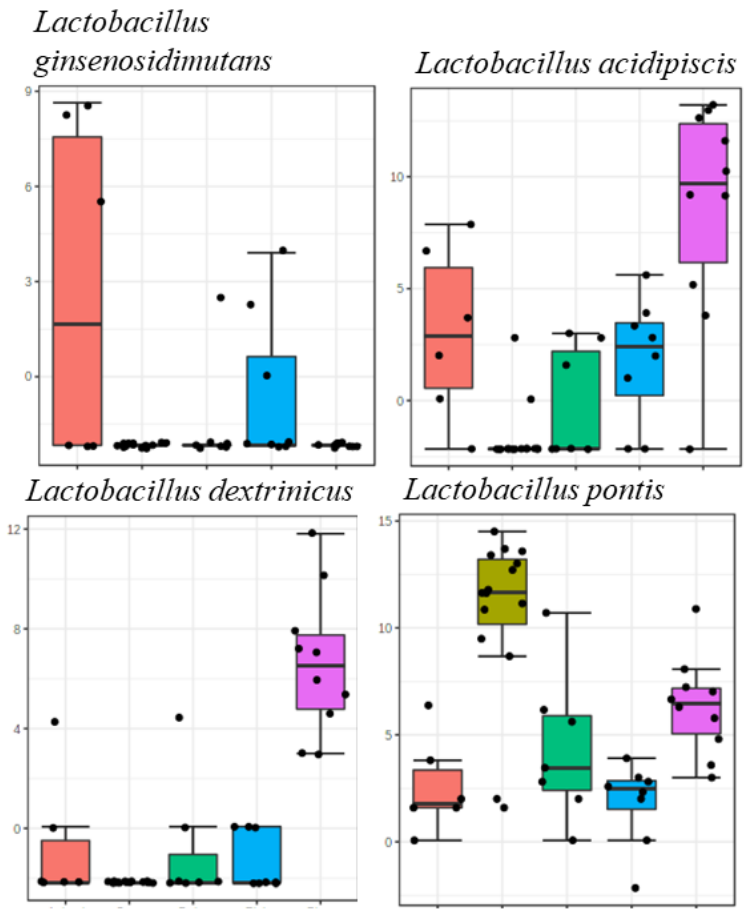
Table 2: Top genera associated with Sudanese food fermentation [relative abundance in %]

Top bacterial genera [relative abundance in %]	**Category of Sudanese food fermentation**				
	Animal	Crop	Dairy	Fish	Plant
Acetobacter	0.03%	96.89%	2.43%	0.46%	0.18%
Acinetobacter	12.67%	8.56%	5.91%	25.86%	47.00%
Actinomyces	9.02%	37.72%	34.06%	7.21%	11.99%
Aeriscardovia	0.14%	0.02%	0.14%	0.56%	99.14%
Aeromonas	17.89%	1.16%	24.40%	56.16%	0.39%
Bacillus	1.40%	7.38%	1.45%	0.65%	89.11%
Bacteroides	98.22%	0.01%	0.22%	0.27%	1.28%
Bifidobacterium	1.17%	0.14%	3.99%	0.27%	94.42%
Brevibacillus	0.28%	1.59%	1.08%	0.07%	96.98%
Cetobacterium	2.52%	0.03%	0.17%	97.23%	0.05%
Citrobacter	9.87%	8.36%	27.74%	2.97%	51.07%
Corynebacterium_1	17.51%	0.39%	1.33%	8.56%	72.22%
Enhydrobacter	0.65%	0.65%	85.03%	13.52%	0.15%
Enterobacter	5.57%	11.85%	19.55%	1.67%	61.36%
Enterococcus	1.43%	5.92%	4.20%	1.04%	87.41%
Escherichia_Shigella	27.37%	2.29%	22.70%	0.83%	46.80%

Ignatzschineria	68.61%	0.00%	0.21%	30.34%	0.84%
Kerstersia	97.68%	0.01%	0.03%	0.06%	2.22%
Kosakonia	2.73%	51.10%	2.81%	0.58%	42.79%
Lactobacillus	0.53%	63.87%	15.49%	0.07%	20.03%
Lactococcus	12.67%	10.25%	65.98%	2.83%	8.27%
Lysinibacillus	72.99%	0.06%	0.44%	11.62%	14.90%
Oblitimonas	98.86%	0.01%	0.06%	0.44%	0.63%
Paeniclostridium	59.36%	0.59%	23.91%	15.58%	0.57%
Pantoea	9.46%	31.02%	4.90%	1.83%	52.79%
Pediococcus	0.52%	8.02%	0.12%	0.06%	91.27%
Peptoniphilus	87.66%	0.02%	0.42%	11.72%	0.18%
Peptostreptococcus	43.64%	1.11%	1.51%	50.25%	3.49%
Proteus	40.44%	0.01%	1.50%	42.70%	15.35%
Providencia	61.76%	0.02%	0.28%	34.75%	3.19%
Pseudomonas	1.72%	10.54%	68.00%	2.48%	17.26%
Psychrobacter	32.84%	0.10%	3.93%	62.61%	0.52%
Sarcina	0.31%	6.90%	1.59%	91.13%	0.07%
Savagea	2.20%	0.01%	0.53%	97.11%	0.15%
Sporosarcina	0.21%	0.01%	0.67%	99.00%	0.12%
Streptococcus	13.93%	25.84%	50.15%	5.90%	4.19%
Vagococcus	69.13%	0.45%	1.03%	22.42%	6.97%

Weissella	0.69%	30.38%	6.54%	0.56%	61.83%
Wohlfahrtiimonas	79.07%	0.01%	0.07%	19.61%	1.25%

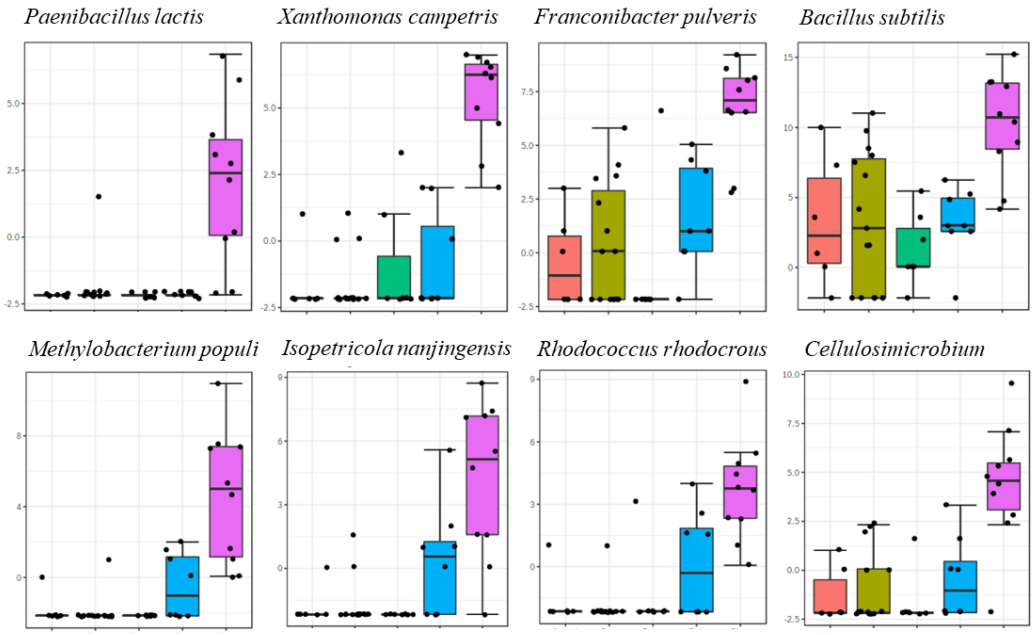
Figure 2e:



2e: Animal = pink, Crop = mustard, Dairy = teal, Fish=blue, Plant =purple. Lactobacillus species significantly varied between Sudanese fermented foods categories. Lactobacillus ginsenosidimutans [q=0.018], was found to be highest in the animal-based foods, Lactobacillus acidipiscus [q=0.0099] and dextrinicus [q=0.01] were abundant in plant-based foods while Lactobacillus pontis were enriched in crop-based foods [q=0.008].

In plant-based foods, significantly increased species included *Paenibacillus lactis* [q=0.025], *Franconibacter pulveris* [q=0.0011], *Bacillus subtilis* [q=0.009], *Methylobacterium populi* [q=0.009], *Isopetricola nanjingensis* [q=0.009] and *Cellulosimicrobium* [q=0.004]. Plant associated microbial species included *Xanthomonas campetris* [q=7.8136e-7] and *Rhodococcus rhodochrous* [q=0.004] [Figure 2f].

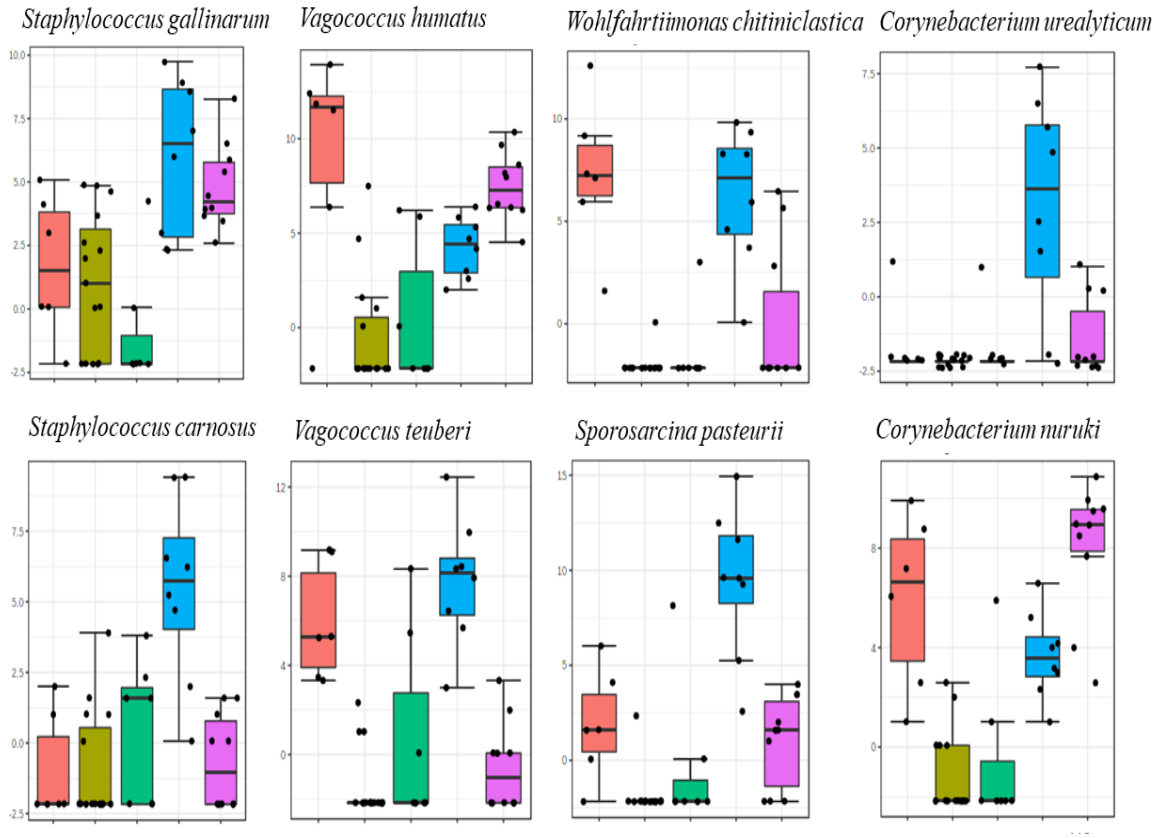
Figure 2f:



2f: Animal = pink, Crop = mustard, Dairy = teal, Fish=blue, Plant =purple. Abundant genera and species in plant-based fermented foods. Paenibacillus lactis [q=0.025], Franconibacter pulveris [q=0.0011], Bacillus subtilis [q=0.009], Methylobacterium populi [q=0.009], Isopetricola nanjingensis [q=0.009] and Cellulosimicrobium [q=0.004]. Plant associated microbial features included Xanthomonas campetris [q=7.8136e-7] and Rhodococcus rhodochrous [q=0.004].

Staphylococcus gallinarum [q=0.008] was found to be abundant in both fish and plant-based foods while *Staphylococcus carnosus* was abundant in fish-based foods [q=0.008]. *Vagococcus humatus [q=0.01]* was abundant in several categories but most enriched in animal-based foods while *Vagococcus teuberi [q=0.009]* was abundant in both fish and animal-based foods. *Corynebacterium urealyticum* [q=0.004] and *Sporosarcina pasteurii* [q=0.009] were abundant in fish-based products while *Corynebacterium nuruki* [q=0.005] were abundant in several categories that included plant, animal, and fish-based foods. Finally, *Wohlfahrtiimonas chitiniclastica* was abundant in animal and fish-based foods [q=0.03] [Figure 2g].

Figure 2g:



2g: Animal = pink, Crop = mustard, Dairy = teal, Fish=blue, Plant =purple. Log fold abundance variations between Sudanese fermented foods categories highlighting significant species variations. Staphylococcus species included, gallinarum [q=0.008], abundant in fish and plant-based foods and carnosus, abundant in fish-based foods only [q=0.008]. Vagococcus species included, humatus [q=0.01], found most enriched in animal-based foods, while teuberi [q=0.009], was enriched in fish and animal-based foods. Corynebacterium species included urealyticum [q=0.004], in fish-based foods and nuruki [q=0.005] which was abundant in several food categories that included plant, animal, and

fish-based foods. Sporosarcina pasteurii [q=0.009] was abundant in fish-based products while Wohlfahrtiimonas chitiniclastica was abundant in animal and fish-based products [q=0.03].

Metagenomic [microbiome and mycobiome] findings in relation to intra-food categorical analysis

Sorghum and millet fermented foods

The most predominant phyla in sorghum and millet fermented foods were *Firmicutes, Proteobacteria* and *Cyanobacteria*. The core microbiome of sorghum and millet foods consisted of 60 genera. Twenty-seven of these genera were shared between the sorghum juice, pancake, and the millet-based food, Damirga. Eleven unique genera were found amongst these foods, in which two were unique to the Damirga [*Lachnoclostridium_5* and *Clostridium sensu stricto 12*], one to the sorghum pancake [*Komagataeibacter*] and eight to the sorghum juice [*Vagococcus, Sphingobacterium, Unclassified, Paracoccus, Diaminobutyricimonas, Chryseobacterium, Paraburkholderia* and *Bacillus*].

64% of *Lactobacillus* abundance was found in sorghum pancakes, 54% in fermented millet Damirga products and 41% in sorghum juice. *Acetobacter* was abundant in the millet fermented food, Damirga [17%] and sorghum pancakes [14%].

When comparing preparations and final stages of crop-based fermented foods, Acetobacter were found to be most abundant in the preparation stages of sorghum pancakes [56.14%] while only [1.72% and 0.53%] of *Acetobacter* could be found in the final sorghum pancake and juice respectively. Sorghum and millet pancake preparations were enriched with *Lactobacillus* [26.91%] while the final sorghum pancake consisted of [4.38%] *Lactobacillus*.

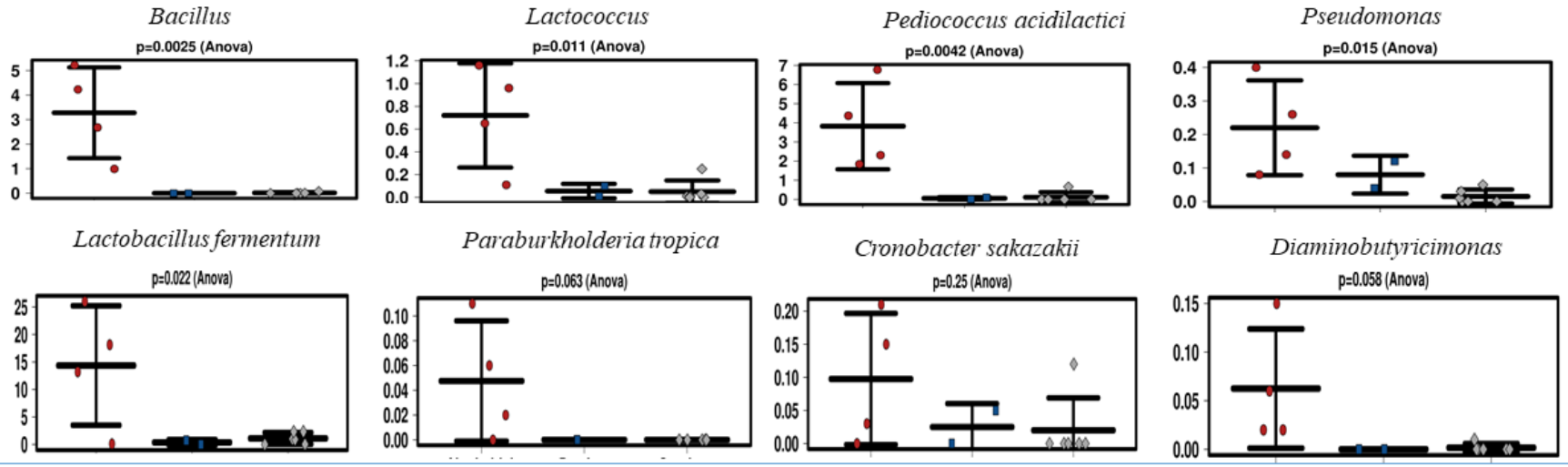
The final sorghum juice was significantly enriched with the genera, *Bacillus* [p=0.025], *Lactococcus* [p=0.011] and *Pseudomonas* [p=0.015] and the species, *Pediococcus acidilactici* [p=0.0042] and *Lactobacillus fermentum* [p=0.022]. *Paraburkholderia tropica* and *Cronobacter sakazakii* were higher in abundance in the sorghum juice but in non-significant variation compared to other species such as *Lactobacillus fermentum* [Figure 3a]. Many other *Lactobacillus* species could be identified in the sorghum and millet fermented samples and included the species *ingluviei, brevis, hamsteri, amylolyticus, kefiri,* and *nagelii*.
The millet fermented food, Damirga was found to have increased abundance of the *Clostridium* species *nitrophenolicum* and *tarantellae*, and the species, *Streptococcus equinus, Comamonas*

aquatica and *Kosakonia radicincitans*. Damirga and the final sorghum pancake shared commonalities in harbouring the *Acetobacter* species; *pasteurianus* and *tropicalis*, *Lactobacillus pontis* and *Citrobacter sedlakii* [Figure 3b].

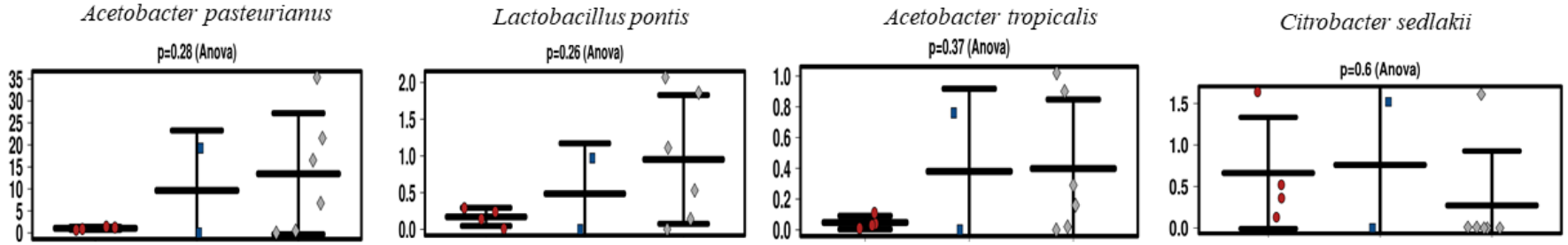
LEfSe analysis highlighted the order *Enterobacteriales* and species *Clostridium tarantellae* as the most distinguishing genera for Damirga fermentation while sorghum juice fermentation could be distinguished by the genus *Chryseobacterium and* the *Bacillus* species; *subtilis* and *coagulans* as well as the species, *Pediococcus acidilactici, Lactococcus lactis,* and *Stenotrophomonas maltophilia*.

The most abundant fungal genera amongst the sorghum and millet foods sampled, were *Issatchenkia* [45.52%] and *Cyberlindnera* [26.25%]. There were many unassigned fungal genera [16.61%] while *Fusarium* and *Blumeria* were abundant at 4.51 and 3.2% respectively. A heatmap of the most abundant fungi in fermented sorghum and millet samples is presented in [Figure 3c]. Samples exhibited a maximum of 2-4 shared fungal genera while one sample [F35], a preparatory sorghum juice sample had the highest number of shared fungal genera.

Figure 3a:

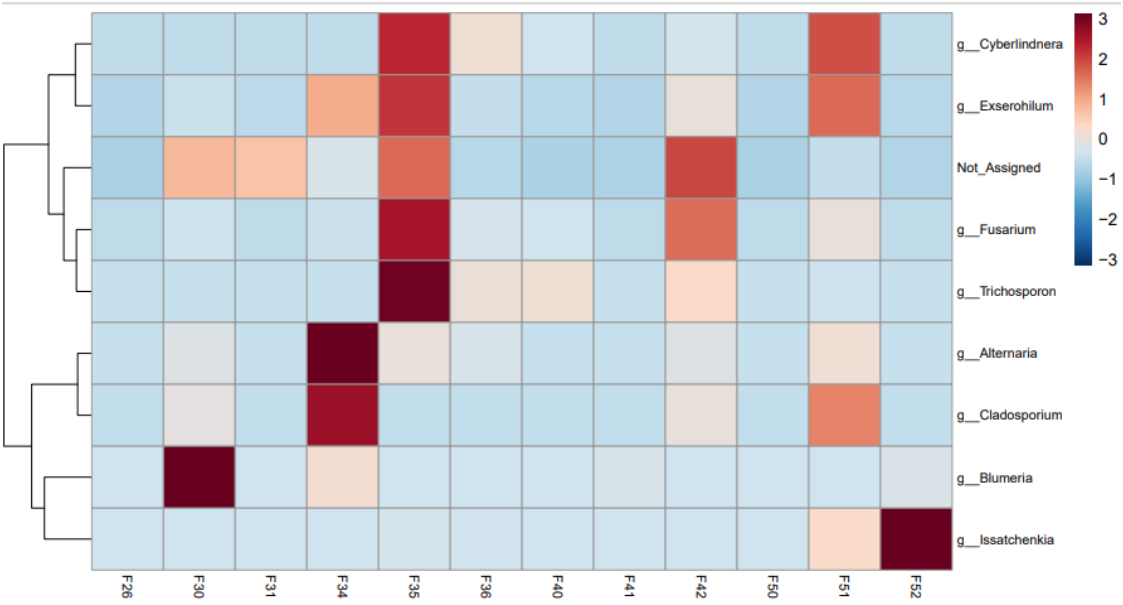


Figures 3b



One-way Anova variations amongst sorghum and millet fermented products: Y axis = abundance total sum scaling. X axis = sorghum and millet-based products. Red= sorghum juice, Abreh, Blue = millet [Damirga] and Grey = sorghum pancakes, Kisra. 3a: Raised abundances of specific genera and species amongst sorghum juice that includes Bacillus, Lactococcus, Pediococcus acidilactici, Pseudomonas, Lactobacillus fermentum Paraburkholderia tropica and Cronobacter sakazakii. 3b: Abundances were shared between Damirga and sorghum preparations in Acetobacter, Lactobacillus and Citrobacter species.

Figure 3c:

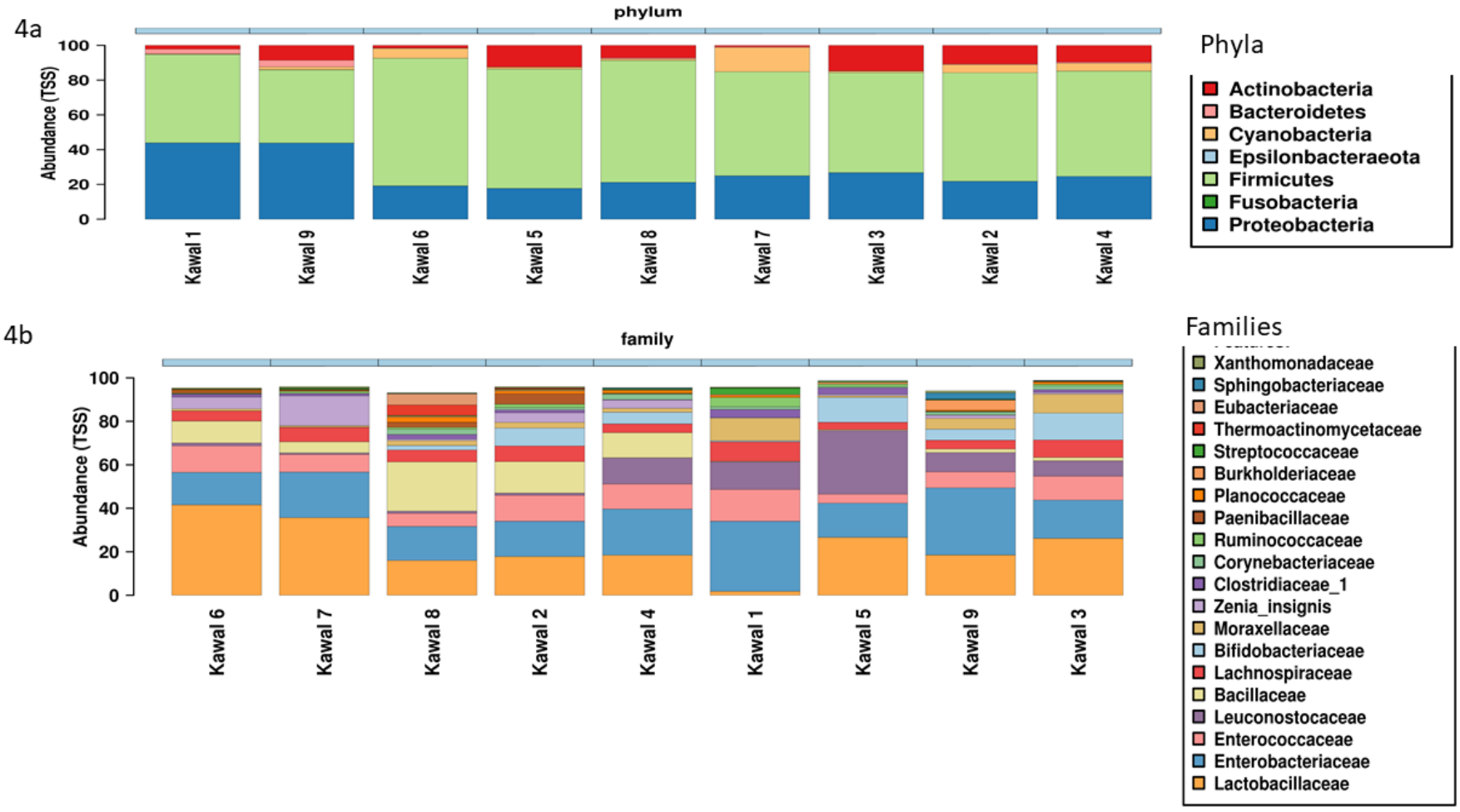


3c: Heatmap of mycobiome genera amongst fermented sorghum samples. F26, F30, F34, F,40, F41, F52 = sorghum pancake samples. F35, F36, F42, F50, F51 = sorghum juice samples. F31= Damirga. One sample of sorghum juice [F35] exhibited the highest abundances of five fungal genera compared to other samples that consisted of a maximum of two to four genera.

Kawal [Cassia obtusifolia]

Predominant phyla in Kawal samples were found to be *Firmicutes, Proteobacteria, Actinobacteria* and *Cyanobacteria* [Figure 4a]. Predominant families included *Lactobacillaceae, Clostridiaceae_1, Enterobacteriaceae, Enterococcaceae, Leuconostoceae, Bacillaceae,* and *Lachnospiraceae* [Figure 4b].

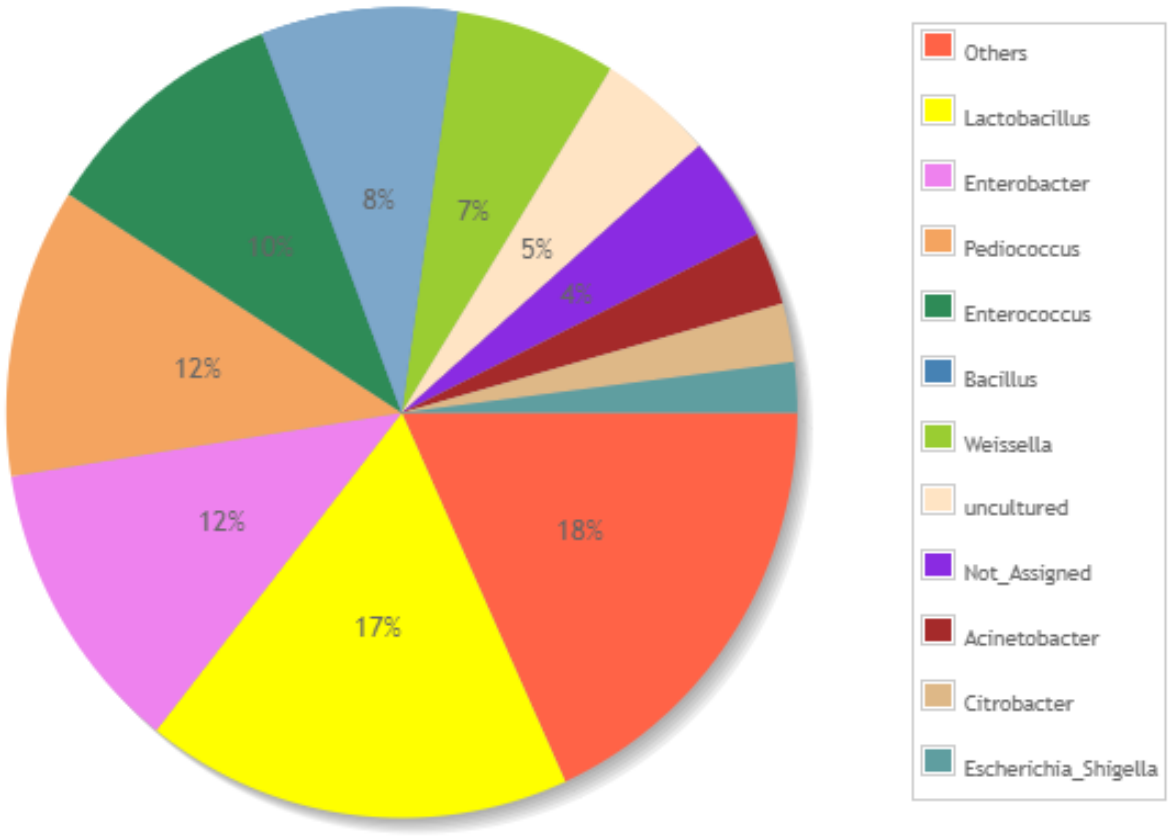
Figure 4a and 4b:



4a: Phyla composition of nine Kawal fermented food samples made from the Cassia obtusifolia plant predominantly grown and produced in Western Sudan. Firmicutes, Proteobacteria and Cyanobacteria are the most abundant phyla. 4b: Microbial family composition amongst Kawal samples. Lactobacillaceae [43.85%] were the most abundant family in Kawal after which Clostridiaceae_1 [7.36%], Streptococcaceae [6.71%] and Enterobacteriaceae [5.35%] were found.

The genera *Lactobacillus [17%], Enterobacter [12%], Pediococcus [12%], Enterococcus [10%], Bacillus [8%] and Weissella [7%]* were found to be the most abundant genera in Kawal [Figure 4c].

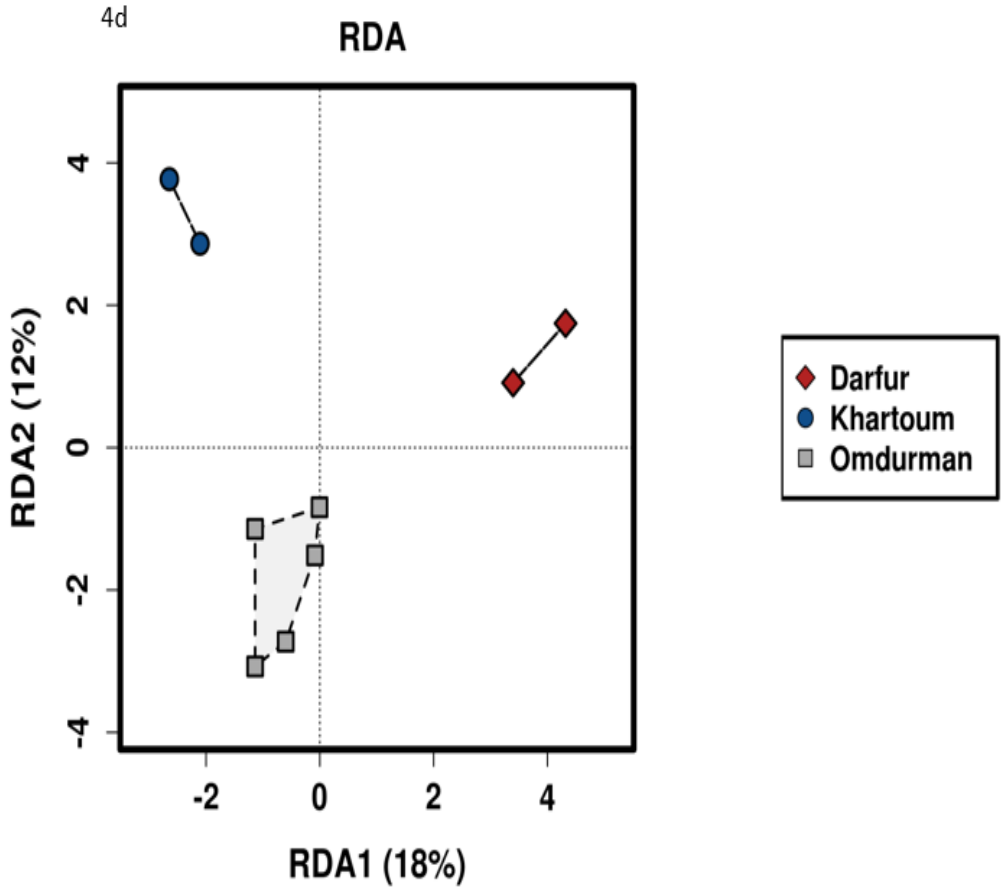
Figure 4c:



4c: Relative abundance of genera found in Kawal food samples expressed in percentages. Lactobacillus [17%], Enterobacter [12%], Pediococcus [12%], Enterococcus [10%], Bacillus [8%] and Weissella [7%] are the most abundant genera in Kawal food type.

Redundancy analysis revealed location of Kawal obtainment had a prominent influence on the microbiome composition [Figure 4d].

Figure 4d:



4d: RDA plotting of Kawal obtainment sites highlights microbiome variation is affected by Kawal origin from Sudan.

Those samples received from Western Sudan were significantly enriched in the genera *Pseudomonas* [p=0.0071], *Aureimonas* [p=0.0081], *Pediococcus* [p=0.00029], *Citricoccus* [p=0.00012] and *Methylobacterium* [p=0.03], while those samples obtained from Khartoum city were found to be significantly enriched in the genera, *Isoptericola* [p=0.043], *Pseudochrobactrum* [p=0.011] and *Pseudoxanthomonas* [p=4e-04] [Figure 4e]. Many bacterial species could be identified in Kawal and included the *Lactobacillus* species; *acidipiscus,*

farciminis and *plantarum, and the species*, *Bacillus subtilis*, *Weissella confusa*, *Pediococcus acidilactici*, *Enterococcus faecalis* and *Citrobacter sedlakii*.

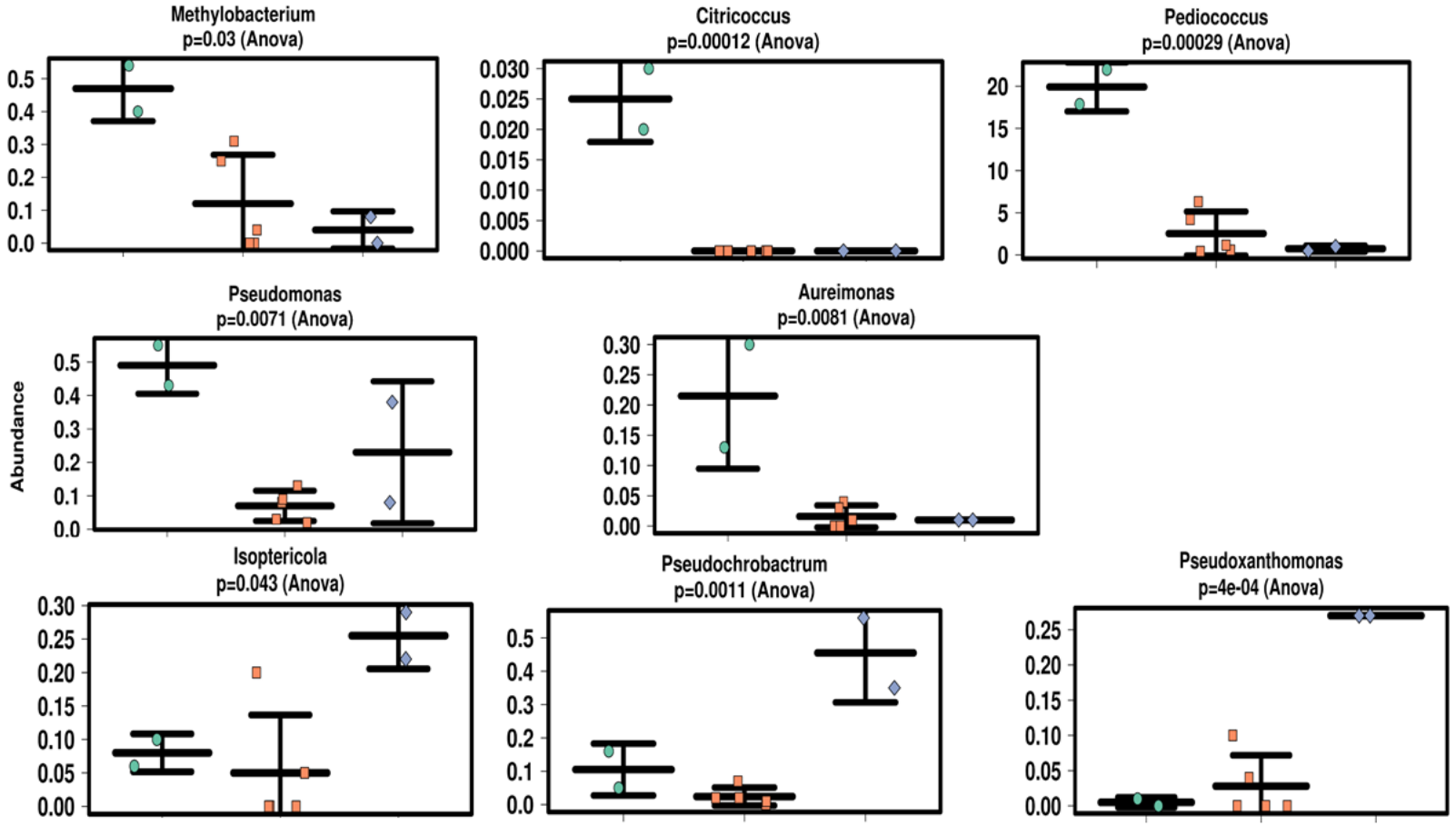
Seven Kawal samples were then analysed for mycobiome composition. The most predominant families were found to be *Saccharomycetales* [12%], *Pleosporaceae* [12%], *Saccharomycodaceae* [12%], *Aspergillaceae* [9%], *Cladosporiaceae* [8%] and *Dipodascaceae* [7%]. Over 60 fungal genera and over 90 fungal species were recognised in the Kawal samples. The most predominant fungal genera were *Issatchenkia* [25%], *Dipodascus* [12%], *Trichosporon* [9%], *Saccharomyces* [5%], *Cladosporium* [3%], *Candida* [2%], *Aspergillus* [2%] and below 1% were the genera *Alternaria* and *Hanseniaspora*. [33%] were of unassigned fungal genera. Fungal species found in the Kawal samples are presented in Table 3. The product Sigda was found to be enriched in *Pediococcus* [44%].

Table 3: Fungal species found in Kawal samples in association with sample abundance

One sample abundance	Two or more sample abundance
Genera with more than one species: *Candida mengyuniae, pseudorugosa* *Aspergillus penicilliodes, aculeatus, sydowii, latus* **Others:** *Fellomyces ogasawarensis, Macrophomina phaseolina, Ceraceosorus guamensis, Podospora prethopodalis, Alternaria angustiovoidea, Rhizopus arrhizus, Rhizomucor miehei, Emergomyces orientalis, Trichosporon insectorum, Peronospora sparsa and Pityriasis rosea*	**Genera with more than one species:** *Cladosporium exasperatum, halotolerans, sphaerospermum,* *Aspergillus caespitosus, quadrilineatus,* *Malassezia restricta, globose* **Others:** *Gibberella intricans, Saprochaete captata, Hypopichia burtonii, Apiosordaria verruculosa, Kluyveromyces marxianus, Sporidiobolus ruineniae, Rhodosporidium toruioides, Neocosmospora falciformis, Neotestudina rosati, Cyberlindnera fabianii, Kodamaea ohmeri, Pichia membranifaciens, Davidiella tassiana, Rhizomucor pusillus, Cryptococcus laurentii, Rhodotorula mucilaginosa, Brettanomyces bruxelensis, Candida orthopsilosis, Ramaria flava,*

	Meyerozyma caribbica and Saitozyma paraflava

Figure 4e:



4e: Green=Darfur, orange = Omdurman and blue= Khartoum. Samples from Western Sudan were enriched in Methylobacterium, Citricoccus, Pediococcus, Pseudomonas and Aureimonas. Samples from Khartoum city highlighted significant elevations in Isoptericola, Pseudochrobactrum and Pseudoxanthomonas

Animal-based products

Miriss [fat sheath fermentation]

Bacteroidaceae, Planococcaceae, Halomonadaceae and *Enterobacteriaceae* were the most abundant families in the fat sheath Miriss samples. After others [26%], predominant genera in Miriss included *Bacteroides* [29%], *Vagococcus* [8%], *Lysinibacillus* [7%], *Halomonas* [7%], *Ignatzschineria* [7%], *Oblitimonas [5%]* and *Peptoniphilus* [3%]. The genera with lesser abundance throughout the Miriss samples included *Proteus, Enterobacter, Streptococcus, Vibrio, Providencia, Kerstersia, Wohlfahrtiimonas* and *Lactobacillus*. We further assessed the Miriss samples with regards microbiome locational variation comparing samples produced in two cities: Omdurman and Khartoum. Significant Shannon index [>2.8, p=0.0039], richness [>170, p=3.3e-05] and evenness [>0.55, p= 0.043] were found between samples with the Miriss obtained from Omdurman having an increased alpha diversity compared to those samples from Khartoum. Species detected in the Miriss included *Bacteroides coprosuis, Megasphaera cerevisiae, Vagococcus humatus* and *Halomonas hamiltonii.*

Fungal analysis of three Miriss samples revealed an enrichment in *Dipodascus* [89.4%] while genera of lower abundance included *Issatchenkia* [4.87%] and *Trichosporon* [3.59%]. Other fungal genera in Miriss included *Exserohilum, Rhodotorula, Blumeria, Cyberlindnera, Alternaria* and *Fusarium*.

Shermout [air-dried meat fermentation]

The Shermout was found to contain *Acinetobacter* [13%], unassigned taxa [12%], *Psychrobacter* [7%], *Myroides* [6%], Aeromonas [6%] and Streptococcus [5%]. *Vagococcus, Proteus* and *Vibrio* were present in 4-5% abundance. Triticum aestivum or bread wheat was also found at 8% relative abundance.

Musran [fermented intestine]

The Musran had a microbiota composition predominantly consisting of *Escherichia_Shigella* [29%], *Paeniclostridium* [24%] and *Clostridium sensu stricto_1* [17%]. Musran also had high abundances in uncultured bacterium [12%] and *Streptococcus* [9%]. Comparing all samples, more than half [57.24%] of *Paeniclostridium* abundance in Sudanese fermented foods was found in the Musran or fermented intestine

Fish-based products

Kajeik

Firmicutes and *Proteobacteria* were the most abundant phyla in Kajeik with one sample harbouring increased *Fusobacteria.* Kajeik microbiome was composed of others [29%], *Sporosarcina [17%], Savagea [13%], Clostridium sensu stricto [7%], Sarcina [6%], Clostridium sensu stricto 4, Proteus [6%], Ignatzschineria [5%], Vagococcus [4%] and Cetobacterium [4%]*. The genera, *Acinetobacter, Kurthia and Aeromonas* could also be identified in the Kajeik samples.

While only one Kajeik sample could be analysed for fungal composition, this highlighted a unique mycobiome composition with abundances of *Aspergillus* [35%], *Rhodotorula* [11%], *Trichosporon* [11%], *Candida* [7%], *Cyberlindnera* [4%], *Lichtheimia* [4%] and *Dipodascus, Cladosporium* and *Rhizopus* at 3% abundance each. Fungal genera in lower abundance included *Diutina, Lasiobolidium, Podospora, Bannoa, Naganishia* and *Preussia*.

Mendeshi

The Mendeshi samples were found to have high abundances of *Acinetobacter [16%], Cetobacterium [10%], Psychrobacter [8%], Aeromonas [7%], Streptococcus [6%], Lysinibacillus [5%], Macrococcus [3%]*. *Savagea* were also identified.

Faseekh

The Faseekh fish product contained significant *Tetragenococcus* [p=7.6e-17] which was found in 32% abundance and *Peptoniphilus* [p=1.7e-07], found to be 15% abundant. *Streptococcus* and *Peptostreptococcus* were found at 4% abundance each, while *Citrobacter, Fusobacterium, Enterobacter and Staphylococcus* were at 3% abundance each in this food type. *Lactobacillus* were significantly increased in this type of fish food product compared to the other forms of fish-based fermented foods [p=0.00079].

Dairy-based products

Mish, Garis and Gergosh

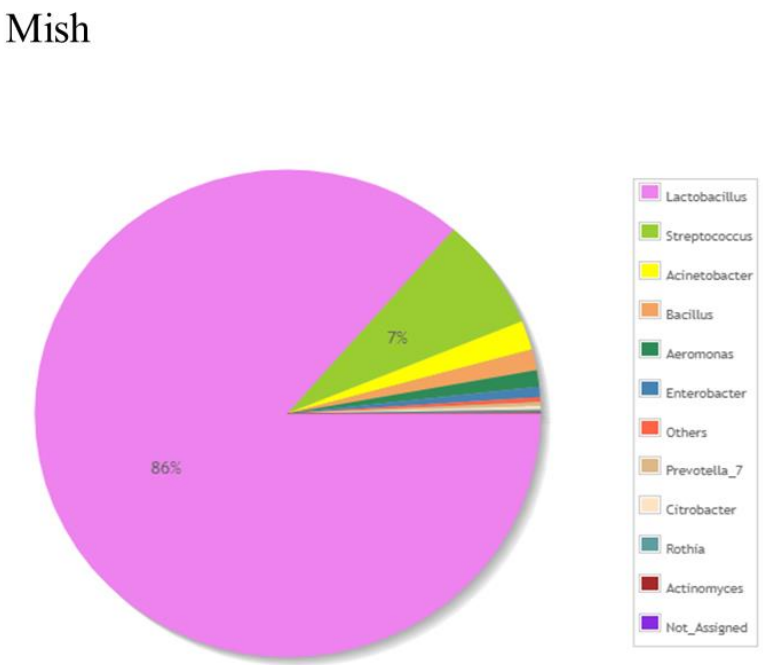
There was enrichment with *Lactobacillus* in the Mish or deep yoghurt [86%] compared to the Gergosh [24%] and Garis samples [10%]. *Streptococcus* abundance was increased in the Gergosh [15%] and Garis samples [12%] compared to Mish [7%]. The Garis samples were most abundant in *Enterobacter* [31%] as well as *Enhydrobacter* [11%] and *Lactococcus* [9%] [Figure 5a]. In Mish, *Bacillus, Aeromonas, Enterobacter* and *Citrobacter* were found in low

abundance. The preparatory stages of Gergosh were enriched in *Lactobacillus* with this trend being lost in the final stage of Gergosh biscuit most likely due to the heat of baking.

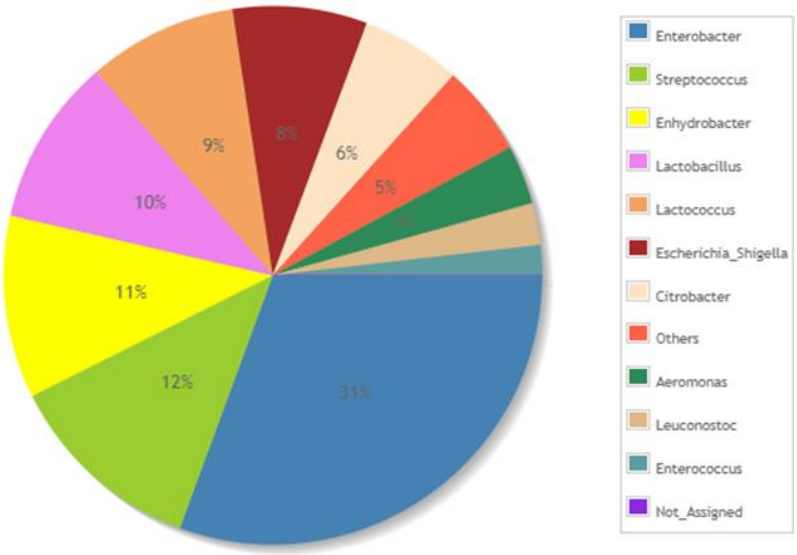
With regards species, the Mish deep yoghurt was significantly enriched with *Bacillus thuringiensis* [p=6.8e-05], *Lactobacillus delbrueckii* [p=0.0069] and *Acinetobacter johnsonii* [p=0.00011] [Figure 5b]. Garis samples were found to be enriched in *Enterobacter hormaechei* [p=0.00075], *Cronobacter sakazakii* [p=0.0097], *Citrobacter sedlakii* [p=0.05] and *Lactococcus* [p=0.0029] with the species, *Lactococcus lactis* [p=0.05] identified [Figure 5c]. Although non-significant, samples of Gergosh dairy/legume biscuit were found to contain *Peptostreptococcus anaerobius, Enterococcus faecalis* and *Lactobacillus plantarum* [Figure 5d].

The most abundant fungal genera in Garis samples were found to be *Kluyveromyces* [79.65%], *Wickerhamiella* [13.09%], and *Cladosporium* [2.14%] while *Candida* comprised of only [0.84%] abundance of the total mycobiome composition of Garis. Other fungal genera detected in Garis included *Blumeria, Saccharomyces, Issatchenkia, Papiliotrema, Tausonia, Kazachstania, Cystobasidium, Colletotrichum* and *Alternaria*.

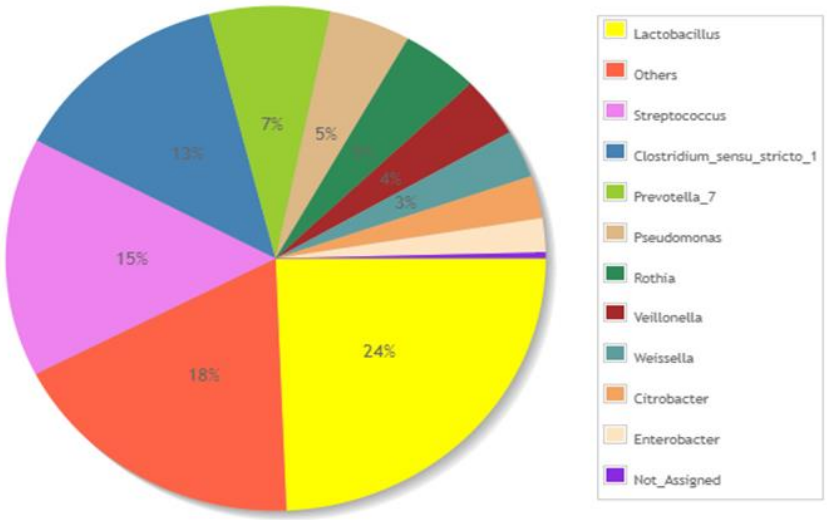
Figure 5a:



Garis

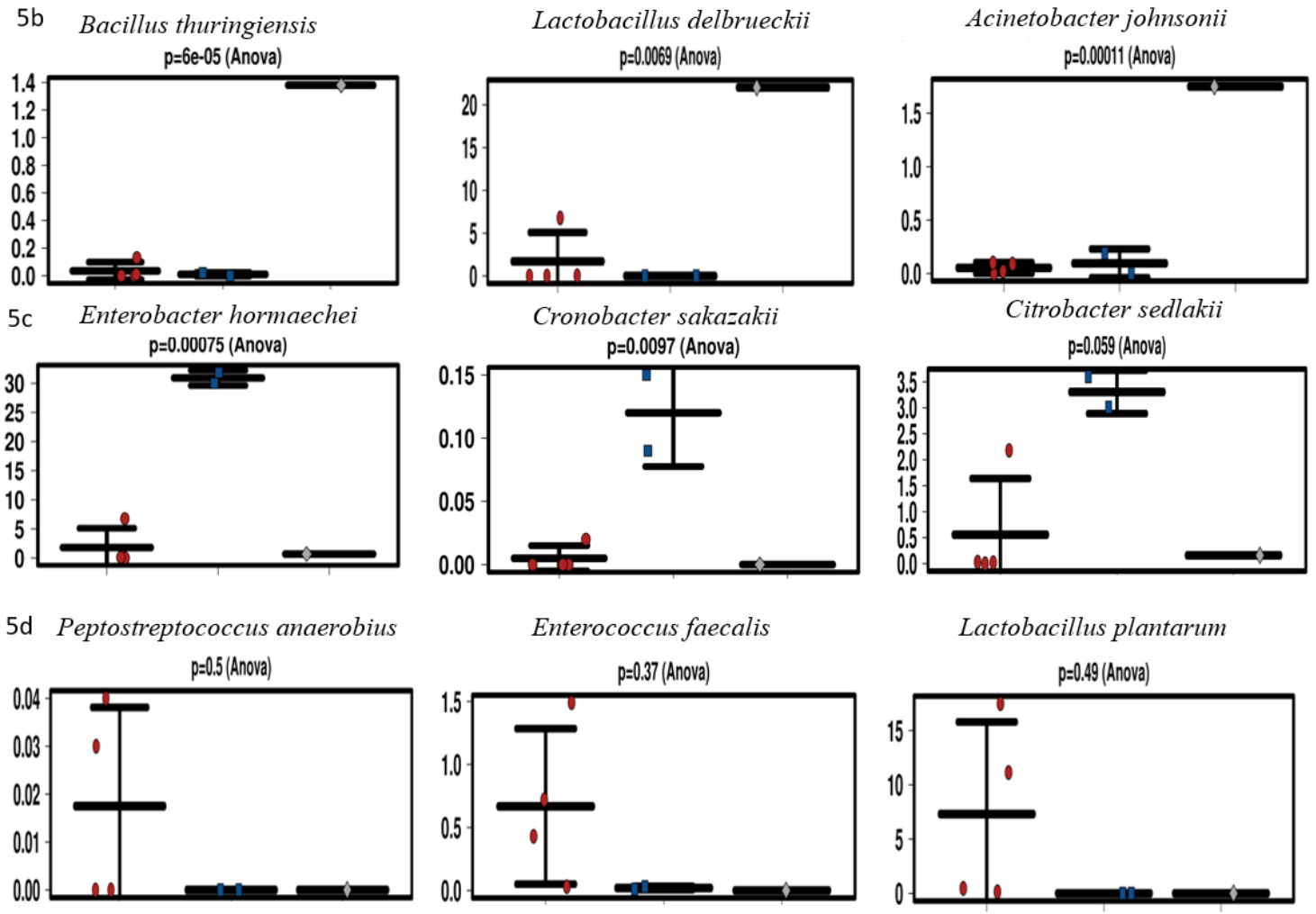


Gergosh



5a: Pie chart relative abundances in % of Sudanese dairy fermented samples. High enrichment with Lactobacillus in the Mish or deep yoghurt [86%] compared to the Gergosh samples [24%] and Garis samples [10%]. Streptococcus abundance was increased in the Gergosh samples [15%] and Garis [12%] compared to Mish [7%]. The Garis samples were most abundant in Enterobacter [31%], Enhydrobacter [11%] and Lactococcus [9].

Figure 5b-5d:


5b
Bacillus thuringiensis
p=6e-05 (Anova)
Lactobacillus delbrueckii
p=0.0069 (Anova)
Acinetobacter johnsonii
p=0.00011 (Anova)
5c
Enterobacter hormaechei
p=0.00075 (Anova)
Cronobacter sakazakii
p=0.0097 (Anova)
Citrobacter sedlakii
p=0.059 (Anova)
5d
Peptostreptococcus anaerobius
p=0.5 (Anova)
Enterococcus faecalis
p=0.37 (Anova)
Lactobacillus plantarum
p=0.49 (Anova)

5b and 5c: Species analysis amongst dairy based fermented products: Y axis = abundance total sum scaling. X axis = three dairy based products. red= dairy biscuit; Gergosh, blue = camel fermented milk; Garis and grey = deep yoghurt; Mish. High abundances of Bacillus, Lactobacillus and Acinetobacter species in deep yoghurt while Garis camel milk samples were enriched with other microbial environments that included Enterobacter, Cronobacter and Citrobacter species. 5d: Although significance was not detected, dairy biscuit samples had increased abundances of Peptostreptococcus, Enterococcus and Lactobacillus species.

Combu

While the Combu is un-fermented, it is utilised commonly during Sudanese fermented food preparation to improve taste and further prevent food spoilage. One sample of Combu was analysed in this study which was found to be enriched in *Lactobacillus* [23%], *Streptococcus* [14%], *Enterobacter* [10%] and *Citrobacter* [10%].

3.5 Discussion

In this study, we have characterised the bacterial and fungal composition in the largest group of Sudanese fermented foods analysed to date through next generation Illumina sequencing. We have shown a diverse range of microbial composition associated with these foods with significant alpha and beta diversity that allows them to predominantly benefit consumers but may also harbour pathogenic viability. Around 60% of Sudanese fermented foods are described as famine or survival foods however from our findings, these foods have many other consumption benefits [2].

Crop-based foods were found to be enriched in the genus *Acetobacter* [96.89%] which oxidise ethanol to acetic acid and are commonly associated with back-slopped fermented foods such as sorghum and millet [15, 16]. Plant-based foods were found to be enriched in *Aeriscardovia* [99.14%], a butyrate producing genus associated with fermentation of food waste [17] while dairy-based foods were abundant in *Enhydrobacter* [85.03%] a genus that can tolerate extremely low Ph [18]. Animal-based foods were abundant in *Bacteroides* [98.22%], a genus associated with late stages of fermentation [19] but are not commonly reported in fermented foods while *Sporosarcina* [99%] were enriched in fish-based foods, a genus previously isolated from other fish fermented foods from other areas of the world [20].

Sorghum and millet are the most abundantly used crops in Sudan due to their ability to survive drought climates [21], [22] but they continue to gain worldwide use due to their gluten-free

properties [23]. Sorghum is low in fat and high in fibre, protein [essential and non-essential amino acids], minerals such as potassium and phosphorous and vitamins such as thiamine and riboflavin. Sorghum pancakes, Kisra have formed the staple diet of Sudanese families for centuries and contribute up to 85% of total energy consumption [24] but only 15% of protein intake [25].

Their production is often initialised by a 10% starter culture that is added to new dough and fermentation is allowed to take place often for a period of 12-19 hours [26]. In a study by Mustafa et al. [2009], the *Lactobacillus* species; *brevis, amylovorus* and *fermentum* and *Saccharomyces cerevisiae* were dominant genera in starter cultures in sorghum pancake [27]. In other studies, the fungi, *Candida intermedia* and *Debaryomyces hansenii* were found to be part of the preparatory stages of sorghum pancake [28] while *Hansenula, Pichia, Trichosporon* and *Yarrowia* have also been reported at this stage [26].

During the initial stages of sorghum fermentation, lactic acid bacteria and yeasts work synergistically, the latter providing vitamins and other growth factors required by bacteria to grow but towards the end of fermentation, acetic acid bacteria predominate [28]. In Damirga fermentation, lactic acid bacteria are also abundant [28] but other genera reported have included *Corynebacterium* and *Streptococcus* as well as the fungi *Fusarium* and *Cladosporium* [29, 30].

Through fermentation, sorghum related foods obtain antimicrobial and antioxidant activity [31] while their digestibility is considerably enhanced in the gastrointestinal tract [21]. Metabolites produced during sorghum fermentation inhibit and suppress the growth of pathogenic bacteria [32]. Polyphenolic compounds such as tannins in sorghum and phytic acid found in millet are reduced during fermentation by *Lactobacillus* which further increases the bioavailability of nutrients in these products [33] [22]. The fermentation of sorghum decreases fat and carbohydrate which are utilised by the microbiota [34] but increases in starch were found after the fermentation of Damirga [22]. Millet flour is a rich source of niacin, riboflavin, and thiamine as well as calcium. Fermentation reduces potassium, sodium, and iron levels in millet [22] but has been shown to raise magnesium and manganese levels [35].

The high acidity achieved by fermentation also provides the sour taste that is desired by the many consumers of these foods. Similar to our findings, *Firmicutes* and *Proteobacteria* were the most abundant phyla in a pyrosequencing based study of sorghum fermented foods of Sudan [7]. Mohamed et al. [1991], isolated *Lactobacillus brevis*, *Klebsiella pneumoniae* and

the fungi *Candida intermedia* and *Aspergillus* from fermented sorghum foods in Sudan [36] while Hamad et al. [1992], detected the *Lactobacillus* species; *reuteri, fermentum* and *amylovorus* and the fungus, *Candida krusei* [25]. In another study, *Pediococcus* were the most dominant lactic acid bacteria in Kisra [37] while *Lactococcus*, *Enterococcus faecalis*, and *Lactobacillus helveticus* were abundant in another study [38]. Other *Lactobacillus* species such as *zeae* and *coleohominis* have also been detected in sorghum fermented preparations [7].

We also found an extensive array of *Lactobacillus* species in both sorghum and millet fermented products that included the species; *lactis, fermentum, ingluviei, brevis, hamsteri, amylolyticus, kefiri,* and *nagelii,* while the abundant fungi in sorghum foods in this study were *Issatchenkia* and *Cyberlindnera*. Fungi detected in other studies from sorghum fermented foods included *Rhizopus, Alternaria, Penicillium, Gibberella, Lasiodiplodia* and *Aspergillus* [7] but similar to our findings, great variations generally exist in the sorghum fungal microbiome [7]. *Issatchenkia*, found to be the most abundant [45.52%] in our samples has also been detected from other African fermented foods as well as in starter cultures [39] while in Sudan, they comprise an important part of the fermented camel's milk, Garis microbiome [40]. Presence of *Issatchenkia* in Kisra has also been reported to improve the growth of *Lactobacillus* species and they contribute to its taste and nutritional value through the production of amino acids [41].

The baking of sorghum pancakes is short, not exceeding 15-20 seconds, allowing them to keep their moisture but also retain many bacterial spores. Their acidic environment further encourages the activation of bacterial endospore germination. However, *Lactobacillus* abundance was found to be lost in the final sorghum pancake in this study, largely due to baking with only 4.38% of *Lactobacillus* abundance appreciated at the final stage compared to 26.91% during the preparatory stages. The baking of sorghum pancakes increases the amino acids asparagine, threonine, valine, isoleucine, leucine and serine but decreases cysteine, lysine, tryptophan and phenylalanine [24].

We further found microbial variation between the sorghum pancakes, Kisra, sorghum juice, Abreh and millet flour and drink, Damirga. We found the Damirga to have increased abundances of the Clostridium species *nitrophenolicum* and *tarantellae, Streptococcus equinus*, *Comamonas aquatica, Kosakonia radicincitans* and *Enterobacteriales.* In other studies, the *Lactobacillus* species; *plantarum* and *fermentum*, *Acetobacter* and the fungus *Kluyveromyces*, have been isolated from Damirga [28].

Clostridium species have been isolated from sorghum fermented foods in one study, while in our findings, *Clostridium* were appreciated in millet fermented foods [7]. *Clostridium* species found in our samples have not been previously detected in Sudanese fermented foods and are associated with enhanced ethanol production [42]. *Comamonas* and *Kosakonia* both detected in the Damirga samples may be regarded as pathogenic additions to the microbiome of Damirga where *Comamonas aquatica* has recently been identified as a potential contributor to gastrointestinal pathology [43] and *Kosakonia radicincitans* have been associated with bloodstream infections [44].

The sorghum juice, a singular production to Sudan was most unique with eight genera distinguishing sorghum juice. Spices such as nutmeg, cinnamon and ginseng are added mid-way in the fermentation process affecting the microbiome. After the addition of spices, lactic and acetic acid bacteria as well as yeasts reduce in counts, genera such as *Corynebacterium* and *Actinobacillus* are stimulated to grow while Staphylococcal growth is inhibited [45]. In our findings, sorghum juice could be distinguished by the genus, *Chryseobacterium,* the *Bacillus* species; *subtilis* and *coagulans* and the species *Pediococcus acidilactici, Lactococcus lactis* and *Stenotrophomonas maltophilia*. *Actinomyces* were also found to be abundant in the final sorghum juice compared to the other samples in this study. In other studies, *Streptococcus*, *Leuconostoc* and *Pediococcus* [45] have been found to be enriched in sorghum juice while Yousif et al. [2010], found the species, *Pediococcus acidilactici*, to distinguish another similar product to sorghum juice 'Hussuwa' also produced in Sudan [4]. The presence of this species has been found in various types of sorghum flour [46] while *Chryseobacterium* is a genus found to contaminate several flours including wheat and rye [47]. While *Stenotrophomonas maltophilia* have been shown to produce the antifungal compound, maltophilin, they are also regarded as newly emerging pathogens with intrinsic resistance to many antibiotic families [48, 49].

We analysed the fermented camels milk, Garis and the deep yoghurt, Mish which are said to be vital indigenous foods of Sudan [50]. Garis is commonly consumed amongst camel herders but is increasingly being sold in city markets as an alleviator of the symptoms of autism, leishmaniasis, hepatitis C, tuberculosis, diabetes, anaemia, and even cancer [50]. Garis also features healthy fatty acids and has been shown to reduce cholesterol levels [28, 51]. It contains a higher ethanol content compared to the other fermented dairy products of Sudan [52]. Its low casein content allows it to neither set nor undergo curdling easily.

Previous reports have found several *Lactobacillus* species in Garis that include *acidipiscus, bulgaricus, delbrueckii, fermentum, helveticus, hordinae, paracasei, plantarum, rhamnosus and raffinolactis* as well as *Lactococcus lactis* and *Streptococcus infantarius.* In this study we found the Garis product to be the least enriched Sudanese dairy-based fermented food with *Lactobacillus [10%]* compared to the Mish or deep yoghurt [86%]. Other genera in Garis in this study were *Enterobacter* [31%], *Enhydrobacter* [11%] and *Lactococcus* [9%]. *Lactococcus* has been identified as one of the predominant bacteria in camel milk from other countries [53] while *Enterobacter*, from the family *Enterobacteriaceae* are an important indication of contaminated meat and dairy foods [54]. The highly ethanol tolerant species *Enterobacter hormaechei* has also been detected in other forms of fermented drinks from around the world, similar to our findings [55, 56]. Other species in the Garis microbiome included *Cronobacter sakazakii,* which can be a life-threatening food-borne pathogen [57] and *Citrobacter sedlakii*.

Fungi reported in Garis include *Saccharomyces* and *Candida* [28] [45] [58] [59] [60]. In line with our findings, the fungus, *Kluyveromyces [79.65%]* was found to be abundant in Garis [40, 61] after which *Wickerhamiella* [13.09%] followed in abundance. *Kluyveromyces* is an important dairy yeast which can produce large amounts of ethanol and is heat tolerant [62] while some *Wickerhamiella* species produce sophorolipids which harbour anticancer activity [63].

The deep yoghurt, Mish is a fermentation product that can be said to contain two stages of fermentation whereby an already fermented product; 'Rob' is re- fermented with fresh milk until it resembles a soft cheese or a 'deep yoghurt' consistency. Some studies have advocated the contamination of Mish by coliform and psychrotrophic bacteria, *Staphylococcus* as well as yeasts and moulds [64]. In this study we found the Mish sample to be significantly enriched in *Bacillus thuringiensis, Lactobacillus delbrueckii* and *Acinetobacter johnsonii*. *Bacillus thuringiensis* grow well in the semi-solid state of Mish [65] and has been isolated from cheese samples from Turkey [66], fermented locust beans from Africa and soybeans from India [67]. Its significant abundance in Mish in Sudan may have potential benefits such as cytotoxic activity against cancer cells [68] while other species such as *Acinetobacter johnsonii* harbour extensive antibiotic resistance [69].

We also analysed the Gergosh dairy/legume fermented biscuit which has received limited attention in the literature. *Peptostreptococcus anaerobius, Enterococcus faecalis* and

Lactobacillus plantarum were detected during the preparation stages while *Streptococcus* were abundant in this food type.

Kawal, fermented from the Cassia obtusifolia plant had by far the most enriched microbiome profile amongst all Sudanese fermented foods. Kawal contains sulphur amino acids such as cysteine and methionine, a unique finding in plant foods [70], adding considerably to its nutritional value. Kawal has been shown to prevent malnutrition and the development of Kwashiorkor disease during times of famine in Sudan [71]. Protein loss is further prevented by the reincorporation of ammonia during fermentation. Riboflavin, iron, vitamin C and beta carotene are enriched in Kawal while calcium is particularly high [70]. Nutrients such as magnesium, manganese and zinc increase after fermentation as well as butyric and acetic acids. The fermentation process of Kawal considerably improves its palatability and shelf-life without spoilage [72].

The microbiota of Kawal has previously been reported to be composed of *Propionibacterium, Bacillus subtilis, Lactobacillus plantarum* and *Staphylococcus sciuri* while fungi detected have included *Saccharomyces, Rhizopus* and *Candida krusei* [45, 70, 73, 74]. Microorganisms in Kawal originate from the leaf, particularly as an important step in its production is that leaves remain unwashed, however utensils, air, humans, and animals, could also contribute to its microbial composition. A unique aspect of Kawal fermentation is the buffering capacity that occurs during its formation where the pH in Kawal remains between 6-7 and may even become alkaline. This is an important factor in its rich abundance of microbial content and has been associated with key species such as *Bacillus subtilis* [75]. We found an abundance of *Lactobacillus, Bifidobacterium, Pediococcus, Acinetobacter, Aeriscardovia, Citrobacter, Enterobacter, Enterococcus, Weissella, Bacillus* and *Brevibacillus* in the Kawal samples analysed in this study.

Indeed, Kawal had a high composition of *Lactobacillus* [17%] after which an abundance of *Enterobacter* [12%] and *Pediococcus* [12%] could be appreciated. Furthermore, the *Lactobacillus* species in Kawal were found to vary from the other foods included in this study such as *Lactobacillus acidipiscus* and *farciminis* but *plantarum* was also detected. *Lactobacillus acidipiscus* has been isolated from both African and Asian fermented fish [76, 77] as well as the fermented camels milk, Garis [60]. *Pediococcus acidilactici* also found in other Sudanese fermented foods and *Bacillus subtilis* were abundant in Kawal, a finding similar to other studies [78, 79]. Both *Lactobacillus farciminis* and *Pediococcus acidilactici* have

shown inhibitory activity against *Listeria monocytogenes* [80] while *Lactobacillus farciminis* has shown extensive antioxidant activity [81].

Plant-based fermented foods predominantly Kawal were unique in their high *Bifidobacterium* abundance [94.42%] compared to the other Sudanese fermented food categories but also contained the highest abundance of *Escherichia-Shigella* [46.08%]. A large diversity of fungal genera and species were further detected in Kawal and are presented in [Table 2].

Sigda, a product obtained from sesame indicum seed fermentation was found to be rich in *Pediococcus* [44%] similar to other reports [70] but unlike other studies, *Streptococcus* and *Bacillus,* the latter associated with spoilt Sigda, were not abundant [70]. While no fungal analysis was achieved in the Sigda sample in this study, *Debaryomyces, Torulopsis, Candida* and *Saccharomyces* have been detected from previous literature [70, 82].

Fish in the diet of the Sudanese is a major source of protein and energy for many of the population. Faseekh [salted moist fermented fish] is made from pebble or tiger fish from the Characidae family utilised due to their leanness and lesser fat content [83]. Fat is an undesired property to some fermented fish products due to its ability to hasten the liquefaction process and cause putrefaction. Kajeik is a sun-dried fermented produce from the fish genera, Clarias, Heterotis or Auchenoglanis [84], while Mendeshi are smaller in size and are produced from the Mormyrus fish species [45]. The fish type utilised in the production of fish fermented foods is said to influence the final microbial composition. Abundances in *Tetragenococcus [32%] and Peptoniphilus [15%]* found in the Faseekh sample. *Tetragenococcus* species are salt-tolerant and contribute to a high amino acid content, improved taste and reduced spoilage [85, 86]. Interestingly, in a case report by Hsieh et al. [2022], *Peptoniphilus* were cultured from a pharyngeal infection involving a fish bone impaction in a 52 year old male [87]. *Peptoniphilus* abundance in our Faseekh fermented fish sample may indicate that fish bones are a source for this genus. *Staphylococcus* [3%] have previously been reported in the microbiome of Faseekh samples, likely due to exterior contamination [83].

In this study Kajeik was found to be abundant in *Sporosarcina [17%], Savagea [13%], Sarcina [6%], and Proteus [6%]*. The fungal composition was abundant in *Aspergillus [35%], Rhodotorula [11%] and Trichosporon [11%]*. Previous studies have isolated *Enterobacteriaceae* and *Streptococcus* from Kajeik as well as the fungi, *Aspergillus, Alternaria* and *Penicillium* [84]. Mendeshi samples were found a differing microbiome abundance that

included *Acinetobacter* [16%], *Cetobacterium* [10%], *Psychrobacter* [8%], *Aeromonas* [7%] *and Streptococcus* [6%].

Sporosarcina and *Psychrobacter* are regarded as salt-tolerant bacteria previously isolated from Korean fermented seafood [20, 88]. *Cetobacterium*, found uniquely to Sudanese fish samples are strongly associated with the gut of freshwater fish [89, 90] and have vitamin producing properties [91]. *Lysinibacillus* [5%] detected in the Mendeshi samples have been isolated from the large intestine of freshwater fish and have shown antimicrobial potential [92]. Our results are also consistent with fish associated genera reported from dried and smoked fish in the Cameroon and that have included *Lysinibacillus*, *Macrococcus* and *Kurthia* [93].

While fish fermented foods of Sudan are low in fat, Miriss fat sheath on the other hand is a product with a high saturated fat content. During fermentation, lipolysis occurs modifying the fats firmness and coherency to a soft and mellow consistency. While Miriss fermentation occurs predominantly under anaerobic settings it is also exposed to aeration, stimulating enzymatic activity, and producing its peculiar taste. Bile acids are also known to be added by some Miriss producers to prevent the growth of anaerobic bacteria while the addition of Combu [potash] [1] achieves the bright white colour of Miriss while also offering bactericidal activity.

The Miriss is known to be a rich source of protein and fat but its microbiota has not been previously reported [1]. *Bacteroides* the most abundant genus in this food type [29%] are capable of breaking down macromolecules in the anaerobic settings found in the Miriss environment [94]. We further detected the species, *Bacteroides coprosuis, Megasphaera cerevisiae, Vagococcus humatus* and *Halomonas hamiltonii* in Miriss and found an abundance in the fungus *Dipodascus* [89.4%] in this fermented food product.

Along with Miriss, the Musran [fermented intestine] has not been previously investigated for its microbiome composition. Musran is another product of Sudanese animal fermentation that turns hard, while the fat covering the intestines undergoes darkening during drying. Musran was found to be abundant in *Escherichia_Shigella* [29%], *Paeniclostridium* [29%] and *Clostridium sensu stricto 1* [17%] as well as *Streptococcus* [9%]. In the Shermout or air-dried meat, there was an enrichment of *Acinetobacter [13%], Psychrobacter* [7%], Myroides [6%] and Aeromonas [6%]. Unlike other studies however, we could not detect *Lactobacillus* in Shermout [95]. Interestingly, the Shermout contained an [8%] abundance of Triticum aestivum [bread wheat] likely added to produce a cheaper product.

One sample of Combu or potash was analysed in this study which is known to be added to many Sudanese fermented foods. Combu was found to be enriched in *Lactobacillus* [23%], *Streptococcus* [14%], *Enterobacter* [10%] and *Citrobacter* [10%]. This addition from Combu to Sudanese fermented foods could add further benefit to the final food's composition.

Conclusion

While some communities in Sudan have come to favour a more Westernised or international diet, this study has shown that the microbiome and mycobiome composition of Sudanese fermented foods may contribute to numerous positive health benefits that should be appreciated but be better developed. Lactic acid bacteria isolated from Sudanese fermented foods have been found to harbour antimicrobial activity against *Staphylococcus aureus* and *Escherichia coli* [5], while *Lactobacillus* isolates from the fermented camel milk, Garis and air-dried meat, Shermout, have previously shown potential in the treatment of urinary tract infection [60, 95]. The Kawal fermented product has been shown to have considerable anti-fungal activity against the endemic disease of Sudan, *Madurella mycetomatis* [96] and was by far the most enriched food product in microbiome and mycobiome composition.

On the other hand, Sudanese fermented foods require better development in their preparation to prevent food contamination. Previous studies have shown pasteurisation to considerably improve the shelf-life of Garis for example but this process remains under-utilised amongst its consumers [97]. Other studies have highlighted that Sudanese fermented food products such as Mish or deep yoghurt are a hazardous source of pathogenic bacteria [98] while in this study we found several pathogens amongst these foods that included *Escherichia_Shigella*, *Wohlfahrtiimonas* and *Ignatzschineria.* Furthermore, Aflatoxin contamination in Sudanese fermented foods maybe associated with the fungal genera, *Alternaria, Aspergillus, Fusarium and Penicillium* [99].

Nevertheless, the rich abundance of lactic acid bacteria and fungi found throughout these foods have the potential to be technologically viable in many future applications that can improve both food safety and security in Sudan [100]. *Lactobacillus, Bifidobacterium, Bacillus* and *Pediococcus,* were all found to be enriched in varying abundance throughout Sudanese fermented foods and are an invaluable addition to global probiotic marketing. Fungal genera such as *Rhizopus* and *Aspergillus* found in this study are regarded as principal components in many worldwide fermentation products [101]. Future research is necessary to understand the

full probiotic potential that the microbiome and mycobiome offers in Sudanese fermented foods [2, 102-104].

3.6 References

1. Ibnouf FO. 2020. Can Indigenous Foods Play Role as 'the Food of Survival'? European Journal of Agriculture and Food Sciences 2.
2. Battcock M. 1998. Fermented fruits and vegetables: a global perspective. Food & Agriculture Org.
3. Dirar HA. 1994. Commentary: The fermented foods of the Sudan. Ecology of Food and Nutrition 32:207-218.
4. Yousif NM, Huch M, Schuster T, Cho G-S, Dirar HA, Holzapfel WH, Franz CM. 2010. Diversity of lactic acid bacteria from Hussuwa, a traditional African fermented sorghum food. Food Microbiology 27:757-768.
5. Elyas Y, Yousif N, Mohamed Ahmed I. 2015. Screening of lactic acid bacteria from Sudanese fermented foods for bacteriocin production. Journal of microbiology, biotechnology and food sciences 4:373-378.
6. Saeed R, Elyas Y, Yousif N, Eltayeb M, Mohamed Ahmed I. 2014. Incidence of Antibiotic Resistance of Lactic Acid Bacteria [LAB] Isolated from Various Sudanese Fermented Foods. J Food Nutr Disor 2014 3:1-7.
7. Eltayeb MM, Eltigani SA, Taniguchi T. 2020. Pyrosequencing scrutiny of bacterial and fungal communities in two Sudanese sorghum-based fermented foods. Annals of Microbiology 70:1-10.
8. Walsh AM, Crispie F, Kilcawley K, O'Sullivan O, O'Sullivan MG, Claesson MJ, Cotter PD. 2016. Microbial Succession and Flavor Production in the Fermented Dairy Beverage Kefir. mSystems 1:e00052-16.
9. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957-2963.
10. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-1.
11. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26:266-7.
12. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project:

improved data processing and web-based tools. Nucleic Acids Res 41:D590-6.

13. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13:581-583.
14. Martin M. 2011. CUTADAPT removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17.
15. De Roos J, De Vuyst L. 2018. Acetic acid bacteria in fermented foods and beverages. Current opinion in biotechnology 49:115-119.
16. Sievers M, Swings J. Acetobacter, p 1-7, Bergey's Manual of Systematics of Archaea and Bacteria doi:https://doi.org/10.1002/9781118960608.gbm00876.
17. Wu Y, Ma H, Zheng M, Wang K. 2015. Lactic acid production from acidogenic fermentation of fruit and vegetable wastes. Bioresource technology 191:53-58.
18. Dong Z, Li J, Wang S, Dong D, Shao T. 2022. Time of Day for Harvest Affects the Fermentation Parameters, Bacterial Community, and Metabolic Characteristics of Sorghum-Sudangrass Hybrid Silage. Msphere 7:e00168-22.
19. Lavefve L, Cureau N, Rodhouse L, Marasini D, Walker LM, Ashley D, Lee SO, Gadonna-Widehem P, Anton PM, Carbonero F. 2021. Microbiota profiles and dynamics in fermented plant-based products and preliminary assessment of their in vitro gut microbiota modulation. Food Frontiers 2:268-281.
20. Guan L, Cho KH, Lee J-H. 2011. Analysis of the cultivable bacterial community in jeotgal, a Korean salted and fermented seafood, and identification of its dominant bacteria. Food microbiology 28:101-113.
21. Abd Elmoneim OE, Schiffler B, Bernhardt R. 2005. Effect of fermentation on the functional properties of sorghum flour. Food Chemistry 92:1-5.
22. Abdalla AA, El Tinay AH, Mohamed BE, Abdalla AH. 1998. Effect of traditional processes on phytate and mineral content of pearl millet. Food Chemistry 63:79-84.
23. Rashwan AK, Yones HA, Karim N, Taha EM, Chen W. 2021. Potential processing technologies for developing sorghum-based food products: An update and comprehensive review. Trends in Food Science & Technology 110:168-182.

24. Tinay AE, Gadir AA, Hidai ME. 1979. Sorghum fermented kisra bread. I—nutritive value of kisra. Journal of the Science of Food and Agriculture 30:859-863.
25. Hamad SH, Böcker G, Vogel RF, Hammes WP. 1992. Microbiological and chemical analysis of fermented sorghum dough for Kisra production. Applied Microbiology and Biotechnology 37:728-731.
26. Sulieman AME, Ali E. 2022. Production and Quality Assessment of Kissra, a Sudanese Fermented Sorghum Product, p 393-412. *In* Elhadi Sulieman AM, Adam Mariod A [ed], African Fermented Food Products- New Trends doi:10.1007/978-3-030-82902-5_26. Springer International Publishing, Cham.
27. Ali AA, Mustafa MM. 2009. Use of starter cultures of lactic acid bacteria and yeasts in the preparation of Kisra, a Sudanese fermented food. Pakistan Journal of Nutrition 8:1349-1353.
28. Sulieman AME, Mustafa WA, Osman OA. 2022. Selected Fermented Cereal Products of Sudan, p 293-311, African Fermented Food Products-New Trends. Springer.
29. Newlove AA, Mahunu GK, Tahir HE. 2022. Fermented Millet for Porridge Production: A Model for Improved Gastrointestinal Health, p 445-458. *In* Elhadi Sulieman AM, Adam Mariod A [ed], African Fermented Food Products- New Trends doi:10.1007/978-3-030-82902-5_29. Springer International Publishing, Cham.
30. Ahmed AI, Abdalla AA, Ibrahim KA, El-Tinay A, Shambat S. 2010. Effect of traditional processing on phosphorus content and some anti nutritional factors of pearl millet [Pennisetum glaucum L.]. Research Journal of Agriculture and Biological Sciences 6:176-180.
31. Tahir HE, Xiaobo Z, Mahunu GK, Mariod AA. 2022. Bioactive Components of Fermented Food Products: Phytochemicals, Phytosterol and Vitamins, p 429-444. *In* Elhadi Sulieman AM, Adam Mariod A [ed], African Fermented Food Products- New Trends doi:10.1007/978-3-030-82902-5_28. Springer International Publishing, Cham.
32. Adebo OA. 2020. African Sorghum-Based Fermented Foods: Past, Current and Future Prospects. Nutrients 12:1111.
33. Svensson L, Sekwati-Monang B, Lutz DL, Schieber A, Ganzle MG. 2010. Phenolic acids and flavonoids in nonfermented and fermented red sorghum

[Sorghum bicolor [L.] Moench]. Journal of Agricultural and Food Chemistry 58:9214-9220.

34. KAZANAS N, Fields M. 1981. Nutritional improvement of sorghum by fermentation. Journal of Food Science 46:819-821.
35. Shambat S. 2010. Influence of traditional processing on minerals HCl–extractability of pearl millet [Pennisetum glaucum]. Research Journal of Agriculture and Biological Sciences 6:530-534.
36. Mohammed SI, Steenson LR, Kirleis AW. 1991. Isolation and characterization of microorganisms associated with the traditional sorghum fermentation for production of Sudanese kisra. Applied and environmental microbiology 57:2529-2533.
37. Rahman IA, Hamad S, Osman M, Dirar H. 2010. Characterization and distribution of microorganisms associated with kisra bread preparation from three sorghum varieties in Sudan. Current Research in Bacteriology 3:138-147.
38. Hamad S, Dieng M, Ehrmann MA, Vogel RF. 1997. Characterization of the bacterial flora of Sudanese sorghum flour and sorghum sourdough. Journal of Applied Microbiology 83:764-770.
39. Mugula JK, Narvhus JA, Sørhaug T. 2003. Use of starter cultures of lactic acid bacteria and yeasts in the preparation of togwa, a Tanzanian fermented food. International Journal of Food Microbiology 83:307-318.
40. Abdelgadir W, Nielsen DS, Hamad S, Jakobsen M. 2008. A traditional Sudanese fermented camel's milk product, Gariss, as a habitat of Streptococcus infantarius subsp. infantarius. International journal of food microbiology 127:215-219.
41. Nout MR. 2003. Traditional fermented products from Africa, Latin America and Asia. Yeasts in food: Beneficial and detrimental aspects:451-473.
42. Hanif MT, Abbas W, Nadeem SF, Ali S. FERMENTATIVE PRODUCTS OF GENUS CLOSTRIDIUM FOR VARIOUS INDUSTRIAL APPLICATIONS.
43. Kannambath R, Stanley P, Mohan P, Mandal J. 2021. Isolation of Comamonas Species from the Stool Samples of Patients with Underlying Gastrointestinal Pathology. Journal of Gastrointestinal Infections 11:47.
44. Bhatti MD, Kalia A, Sahasrabhojane P, Kim J, Greenberg DE, Shelburne SA. 2017. Identification and whole genome sequencing of the first case of

Kosakonia radicincitans causing a human bloodstream infection. Frontiers in Microbiology 8:62.

45. Dirar HA. The indigenous fermented foods of the Sudan : a study in African food and nutrition, p. *In* [ed],

46. Angelov AI, Petrova G, Angelov AD, Stefanova P, Bokossa IY, Tchekessi CK, Marco ML, Gotcheva V. 2017. Molecular identification of yeasts and lactic acid bacteria involved in the production of Beninese fermented food degue. The Open Biotechnology Journal 11.

47. Voidarou C, Antoniadou M, Rozos G, Tzora A, Skoufos I, Varzakas T, Lagiou A, Bezirtzoglou E. 2020. Fermentative foods: Microbiology, biochemistry, potential human health benefits and public health issues. Foods 10:69.

48. Agri H, Karthikeyan R, Kiranmayee B, Jayakumar V, Yadav A, Vinodh Kumar O, Sinha DK, Singh BR. 2022. Stenotrophomonas maltophilia: An Overlooked Enemy Disguised as a Friend. Acta Scientific MICROBIOLOGY [ISSN: 2581-3226] 5.

49. Jakobi M, WINKELMANN G, Kaiser D, Kempter C, JUNG G, Berg G, Bahl H. 1996. Maltophilin: A new antifiingal compound produced by Stenotrophomonas maltophilia R3089. The Journal of antibiotics 49:1101-1104.

50. Abdelgadir WS, Ahmed TK, Dirar HA. 1998. The traditional fermented milk products of the Sudan. International Journal of Food Microbiology 44:1-13.

51. Muthukumaran MS, Mudgil P, Baba WN, Ayoub MA, Maqsood S. 2022. A comprehensive review on health benefits, nutritional composition and processed products of camel milk. Food Reviews International:1-37.

52. Sulieman AME, Alayan AA. 2022. Nutritional, Antimicrobial and Bioactive Components of Gariss, a Fermented Camel Milk Product, p 175-187, African Fermented Food Products-New Trends. Springer.

53. Shori AB. 2017. Camel milk and its fermented products as a source of potential probiotic strains and novel food cultures: A mini review. PharmaNutrition 5:84-88.

54. Mladenović KG, Grujović MŽ, Kiš M, Furmeg S, Tkalec VJ, Stefanović OD, Kocić-Tanackov SD. 2021. Enterobacteriaceae in food safety with an emphasis on raw milk and meat. Applied Microbiology and Biotechnology 105:8615-8627.

55. Wei C-L, Chao S-H, Tsai W-B, Lee P-S, Tsau N-H, Chen J-S, Lai W-L, Tu JC-Y, Tsai Y-C. 2013. Analysis of bacterial diversity during the fermentation of inyu, a high-temperature fermented soy sauce, using nested PCR-denaturing gradient gel electrophoresis and the plate count method. Food microbiology 33:252-261.
56. Pati S, Mohanty MK, Mohapatra S, Samantaray D. 2022. Bioethanol production by Enterobacter hormaechei through carbon monoxide-rich syngas fermentation. International Journal of Energy Research 46:20096-20106.
57. Kim D-H, Jeong D, Song K-Y, Kang I-B, Kim H, Seo K-H. 2018. Culture supernatant produced by Lactobacillus kefiri from kefir inhibits the growth of Cronobacter sakazakii. Journal of dairy research 85:98-103.
58. Ahmed O, El Owni O. 2015. Isolation and Identification of Lactic Acid Bacteria in Sudanese Fermented Camel Milk [Gariss].
59. Abdelgadir W, Hamad S, Moller P, Jakobsen M. 2001. Characterisation of the dominant microbiota of Sudanese fermented milk Rob. International Dairy Journal 11:63-70.
60. Ibrahim A, Altayb H, Elgadi ZA. 2017. Antimicrobial Activity of Lactic Acid Bacteria Isolated from Camel Milk against Pathogenic Bacteria Isolated from Urinary Tract Infected Patients, Sudan.
61. Abdelgadir WS, Ahmed TK, Dirar HA. 1998. The traditional fermented milk products of the Sudan. Int J Food Microbiol 44:1-13.
62. Beniwal A, Saini P, Kokkiligadda A, Vij S. 2017. Physiological growth and galactose utilization by dairy yeast Kluyveromyces marxianus in mixed sugars and whey during fermentation. 3 Biotech 7:349.
63. Chen J, Song X, Zhang H, Qu Y. 2006. Production, structure elucidation and anticancer properties of sophorolipid from Wickerhamiella domercqiae. Enzyme and microbial technology 39:501-506.
64. Abdalla M, Ahmed A. 2010. Evaluation of microbiological quality of Sudanese fermented dairy product ‘mish’during storage. Adv J Food Sci Technol 2:155-158.
65. Capalbo DM. 1995. Bacillus thuringiensis: fermentation process and risk assessment: a short review. Memórias do Instituto Oswaldo Cruz 90:135-138.

66. Molva C, Sudagidan M, Okuklu B. 2009. Extracellular enzyme production and enterotoxigenic gene profiles of Bacillus cereus and Bacillus thuringiensis strains isolated from cheese in Turkey. Food Control 20:829-834.
67. Sarkar PK, Hasenack B, Nout MJR. 2002. Diversity and functionality of Bacillus and related genera isolated from spontaneously fermented soybeans [Indian Kinema] and locust beans [African Soumbala]. International Journal of Food Microbiology 77:175-186.
68. Melo ALdA, Soccol VT, Soccol CR. 2016. Bacillus thuringiensis: mechanism of action, resistance, and new applications: a review. Critical reviews in biotechnology 36:317-326.
69. Carvalheira A, Silva J, Teixeira P. 2021. Acinetobacter spp. in food and drinking water–a review. Food Microbiology 95:103675.
70. Harper DB, Collins M. 1992. Leaf and seed fermentations of Western Sudan. Applications of Biotechnology in Traditional Fermented Foods:105-113.
71. Mbaiguinam M, Mahmout Y, Tarkodjiel M, Delobel B, Bessiere J-M. 2005. Constituents of Kawal, fermented Cassia obtusifolia leaves, a traditional food from Chad. African Journal of Biotechnology 4.
72. Osman N, Mohamed Ahmed I, Babiker E. 2010. Fermentation and cooking of sicklepod [Cassia obtusifolia] leaves: Changes in chemical and amino acid composition, antinutrients and protein fractions and digestibility. International Journal of Food Science & Technology 45:124-132.
73. Dirar HA, Harper DB, Collins MA. 1985. Biochemical and microbiological studies on kawal, a meat substitute derived by fermentation of Cassia obtusifolia leaves. Journal of the Science of Food and Agriculture 36:881-892.
74. Dirar HA. 1984. Kawal, meat substitute from fermentedCassia obtusifolia leaves. Economic Botany 38:342-349.
75. Wang J, Fung DY. 1996. Alkaline-fermented foods: a review with emphasis on pidan fermentation. Critical Reviews in Microbiology 22:101-138.
76. Anihouvi V, Kindossi J, Hounhouigan J. 2012. Processing and quality characteristics of some major fermented fish products from Africa: a critical review. International Research Journal of Biological Sciences 1:72-84.
77. Tanasupawat S, Shida O, Okada S, Komagata K. 2000. Lactobacillus acidipiscis sp. nov. and Weissella thailandensis sp. nov., isolated from

fermented fish in Thailand. International Journal of Systematic and Evolutionary Microbiology 50:1479-1485.

78. Hamadnalla HM. 2020. Isolation and Identification of Bacteria From Dried Fermented Leaves of Cassia obtusifolia. L [Kawal]. Clinical Medicine and Medical Research 1:16-20.

79. Lawane AI, Aly S, François T, Flibert G, Yves T. 2016. Identification and Biochemical Analysis of Microorganisms involved in the Fermentation of Kawal-A Traditional Fermented Sicklepod Leaves [Senna obtusifolia]. Int J Curr Microbiol App Sci 5:261-270.

80. Jawan R, Kasimin M, Jalal S, Faik AM, Abbasiliasi S, Ariff A. Isolation, characterisation and in vitro evaluation of bacteriocins-producing lactic acid bacteria from fermented products of Northern Borneo for their beneficial roles in food industry, p 012020. *In* [ed], IOP Publishing,

81. Verardo V, Gómez-Caravaca AM, Tabanelli G. 2020. Bioactive components in fermented foods and food by-products, vol 9, p 153. MDPI.

82. Elfaki AE, Dirar HA, Collins MA, Harper DB. 1991. Biochemical and microbiological investigations of sigda—A Sudanese fermented food derived from sesame oilseed cake. Journal of the Science of Food and Agriculture 57:351-365.

83. Osman O, Mustafa W, Sulieman AM. 2022. Selected Fermented Fish Products of the Sudan, p 203-217 doi:10.1007/978-3-030-82902-5_14.

84. KushKush ZMH, Sulieman AME. 2022. Nutrient Composition and Bioactive Components of Kejeik, p 365-378. *In* Elhadi Sulieman AM, Adam Mariod A [ed], African Fermented Food Products- New Trends doi:10.1007/978-3-030-82902-5_24. Springer International Publishing, Cham.

85. Udomsil N, Rodtong S, Choi YJ, Hua Y, Yongsawatdigul J. 2011. Use of Tetragenococcus halophilus as a starter culture for flavor improvement in fish sauce fermentation. Journal of Agricultural and Food Chemistry 59:8401-8408.

86. Li C, Li W, Li L, Chen S, Wu Y, Qi B. 2022. Microbial community changes induced by a newly isolated salt-tolerant Tetragenococcus muriaticus improve the volatile flavor formation in low-salt fish sauce. Food Research International 156:111153.

87. Hsieh JW, Dulguerov N, Mermod M. 2022. Mediastinal abscess revealed by computed tomography after pharyngeal fish-bone impaction. Radiology Case Reports 17:4478-4480.
88. Song E-J, Lee E-S, Park S-L, Choi H-J, Roh SW, Nam Y-D. 2018. Bacterial community analysis in three types of the fermented seafood, jeotgal, produced in South Korea. Bioscience, Biotechnology, and Biochemistry 82:1444-1454.
89. Tsuchiya C, Sakata T, Sugita H. 2008. Novel ecological niche of Cetobacterium somerae, an anaerobic bacterium in the intestinal tracts of freshwater fish. Letters in applied microbiology 46:43-48.
90. Zhang Z, Fan Z, Yi M, Liu Z, Ke X, Gao F, Cao J, Wang M, Chen G, Lu M. 2022. Characterization of the core gut microbiota of Nile tilapia [Oreochromis niloticus]: indication of a putative novel Cetobacterium species and analysis of its potential function on nutrition. Archives of Microbiology 204:1-12.
91. Ramírez C, Coronado J, Silva A, Romero J. 2018. Cetobacterium is a major component of the microbiome of giant Amazonian fish [Arapaima gigas] in Ecuador. Animals 8:189.
92. Seelam NS, Akula H, Katike U, Obulam VSR. Isolation, identification and characterization of Lysinibacillus sphaericus from fresh water fish intestinal tract.
93. Maïworé J, Ngoune LT, Koumba MK, Metayer I, Montet D, Durand N. 2021. Determination of bacterial population and the presence of pesticide residues from some Cameroonian smoked and dried fish. Scientific African 13:e00886.
94. Nakasaki K, Idemoto Y, Abe M, Rollon AP. 2009. Comparison of organic matter degradation and microbial community during thermophilic composting of two different types of anaerobic sludge. Bioresource technology 100:676-682.
95. Elyass ME, Shigidi M, Attitalla IH, Mahdi AA. 2017. Characterization and Optimization of Bacteriocin from Lactobacillus plantarum Isolated from Fermented Beef [Shermout]. Open Journal of Applied Sciences 7:83.
96. Ahmed S, Ahmed KE, Ali H, Omer EA, Khalid SA. 2015. Polyphenols profiles of fermented and non fermented Cassia obtusifolia leaves and their antifungal activity against Madurella mycetomatis. Planta Medica 81:PW_173.

97. Ibtisam E, Marowa II. 2009. Effect of pasteurization of milk on the keeping quality of fermented camel milk [Gariss] in Sudan. Livestock Research for Rural Development 21.
98. Abdalla W, El-Zubeir I. 2010. Microbial hazards associated with fermented milk [roub and mish] processing in Sudan. International Journal of Dairy Science 5:162-167.
99. Kumar A, Pathak H, Bhadauria S, Sudan J. 2021. Aflatoxin contamination in food crops: causes, detection, and management: a review. Food Production, Processing and Nutrition 3:1-9.
100. Mensah P. 1997. Fermentation—the key to food safety assurance in Africa? Food Control 8:271-278.
101. Banik A, Halder SK, Ghosh C, Mondal KC. 2019. Fungal Probiotics: Opportunity, Challenge, and Prospects, p 101-117. *In* Yadav AN, Singh S, Mishra S, Gupta A [ed], Recent Advancement in White Biotechnology Through Fungi: Volume 2: Perspective for Value-Added Products and Environments doi:10.1007/978-3-030-14846-1_3. Springer International Publishing, Cham.
102. Jans C, Meile L, Kaindi DWM, Kogi-Makau W, Lamuka P, Renault P, Kreikemeyer B, Lacroix C, Hattendorf J, Zinsstag J. 2017. African fermented dairy products–overview of predominant technologically important microorganisms focusing on African Streptococcus infantarius variants and potential future applications for enhanced food safety and security. International Journal of Food Microbiology 250:27-36.
103. Soemarie YB, Milanda T, Barliana MI. 2021. Fermented foods as probiotics: a review. Journal of Advanced Pharmaceutical Technology & Research 12:335.
104. Salih AMM, Hassan ZR, El Sanousi SM. 2011. Antimicrobial activity of selected Probiotic bacteria isolated from Sudanese Traditional dairy Product ‘Rob’. Pakistan Journal of Nutrition 10:637-639.

4 Altered oral microbiome in Sudanese smokeless tobacco users carries a newly emerging risk of squamous cell carcinoma development and progression

Amel Sami [1,3], Imad Elimairi [3], C. Anthony Ryan [4], Catherine Stanton [1,2] Dhrati Patangia [1] and R. Paul Ross [1]

Affiliations:
1-APC Microbiome Ireland, School of Microbiology, University College Cork, Cork T12 YN60
2 - Teagasc Food Research Centre, Moorepark, Fermoy, Cork, P61 C996, Ireland
3 - Department of Oral and Maxillofacial Surgery and Oral Medicine, Faculty of Dentistry, National Ribat University, Nile street, Khartoum, 1111, Sudan
4 - Department of Paediatrics and Child Health, University College Cork, Cork T12 DFK4

This chapter has been submitted for Cancer microbiome guest edition 2022, *Scientific Reports*

4.1 Abstract:

There are an estimated 6-10 million smokeless tobacco [Toombak] users in Sudan, the majority being males. Toombak is known to be a carcinogenic product that is likely to modify the oral microbiome spatiality into a high-risk potential for the development and progression of oral cancer, but previous studies are lacking in this field.

Here, we endeavour for the first time the exploration of the oral microbiome in key mucosal areas of the oral cavity and assess the microbiome variations in premalignant and oral squamous cell carcinoma [OSCC] samples from both users and non-users of Toombak. 16S rRNA sequencing was performed on DNA obtained from pooled saliva, oral mucosa and supragingival plaque from 78 Sudanese users and non-users of Toombak, aged between 20-70 years. In 32 of the pooled saliva samples, the mycobiome [fungal] environment was analysed through ITS sequencing. Then, 46 formalin-fixed paraffin-embedded samples of premalignant and OSCC samples were collected, and their associated microbiomes sequenced.

The oral Sudanese microbiome was found to be enriched in *Streptococcaceae,* but *Staphylococcaceae* were significantly more abundant amongst Toombak users. Genera enriched in the oral cavity of Toombak users were *Corynebacterium_1* and *Cardiobacterium* while in non-users, *Prevotella, Lactobacillus* and *Bifidobacterium* were prominent. *Aspergillus* was the most abundant fungus in the mouths of Toombak users with a marked loss of *Candida*. The genus *Corynebacterium_1* was abundant in the buccal, floor of the mouth and saliva microbiomes as well as oral cancer samples from Toombak users indicating a possible role for this genus through the early stages of cancer development. An oral cancer microbiome favouring poor survival and metastasis in those who use Toombak emerged that includes the genera *Stenotrophomonas* and *Schlegelella.*

Those utilising Toombak carry an altered oral microbiome that may be an additional risk factor for its carcinogenicity to the oral structures with significant microbiome modulations a newly emerging key driving factor of oral cancer development and progression in Toombak users. Toombak users also carry an oral cancer microbiome that may increase the potential for a poorer prognosis.

4.2 Introduction

Smokeless tobaccos harbour bacterial and fungal entities that can allow for microorganisms to reside in the oral cavity and their mucosal surfaces which are not normal of the typical oral microflora. These oral mucosal surfaces are further impacted by salivary pH, flow of saliva and oral hygiene and along with smokeless tobacco use can have a varying impact on the localised keratinised and non-keratinised tissue.

An example of smokeless tobacco is 'Toombak', used predominantly by Sudanese males that is produced from the plant, the *Nicotiana Rustica*. Leaves are harvested, subjected to fermentation and sodium bicarbonate is added to improve the taste and bioavailability of nicotine in the final product. Through these processes, Toombak has been found to carry potently elevated quantities of carcinogenic compounds that include but are not limited to tobacco-specific nitrosamines, formaldehyde, and acetaldehyde [1, 2]. Iron is also found to be abundant in Toombak and may be involved in tumour development including oral squamous cell carcinoma [OSCC] [3, 4].

Indeed, OSCC is a disease that in Sudan has been primarily attributed to Toombak use [5]. The oral microbiome of smokeless tobacco users has been shown to be modelled into a dysbiotic state as early as four weeks from tobacco use initiation [6, 7]. Smokeless tobacco use can lead to histological changes in the oral epithelium within just two days of its use, causing inflammation and ulceration [8, 9]. The roughened texture of smokeless tobacco products, including Toombak is also a factor in supporting the survival of many of the microorganisms in these products.

Furthermore, smokeless tobacco use has been shown to modify the immune response through the elevation of Interleukins 1 and 2 and the reduction of macrophages, interferon γ, and interleukin 10 in local smokeless tobacco placement sites [8]. Smokeless tobacco use is further said to activate the oncogenic RAS gene and help in the upregulation of the oxidative stress pathways; ASK1, JNK 1 and 2 and p38 [9]. Toombak users who develop OSCC were found to have increased p53 mutations, novel mutations in exons 5,6 and 7 [10] and an increased expression in keratins 13,14 and 19, that relate to abnormal proliferation and maturation of keratinocytes [11].

The relationship between the oral microbiome and cancer development has in the last decade been continuously interrogated, albeit with a narrower or absent insight from developing countries [12]. Culture studies from Sudan have highlighted enriched

Bacillus growth from the buccal mucosa of Toombak users [6]. On the other hand, Mohamed et al. [2021] identified *Malassezia* as a favourable predictor of survival amongst Sudanese OSCC patients [13].

This study serves to analyse the oral cavity microbial spatiality of the Sudanese, including smokeless tobacco users [a high risk cohort for oral cancer development] through metagenomics sequencing approaches while also helping to understand the implications that Toombak use brings post-OSCC development.

4.3 Materials and Methods

Sample collection

Ethical approval was first sought and obtained from the Sudanese ethics committee at National Ribat University Sudan and the Cork Ethics Committee, Cork, Ireland. All experiments were performed in accordance with relevant names guidelines and regulations. Informed consent was obtained from all participants.

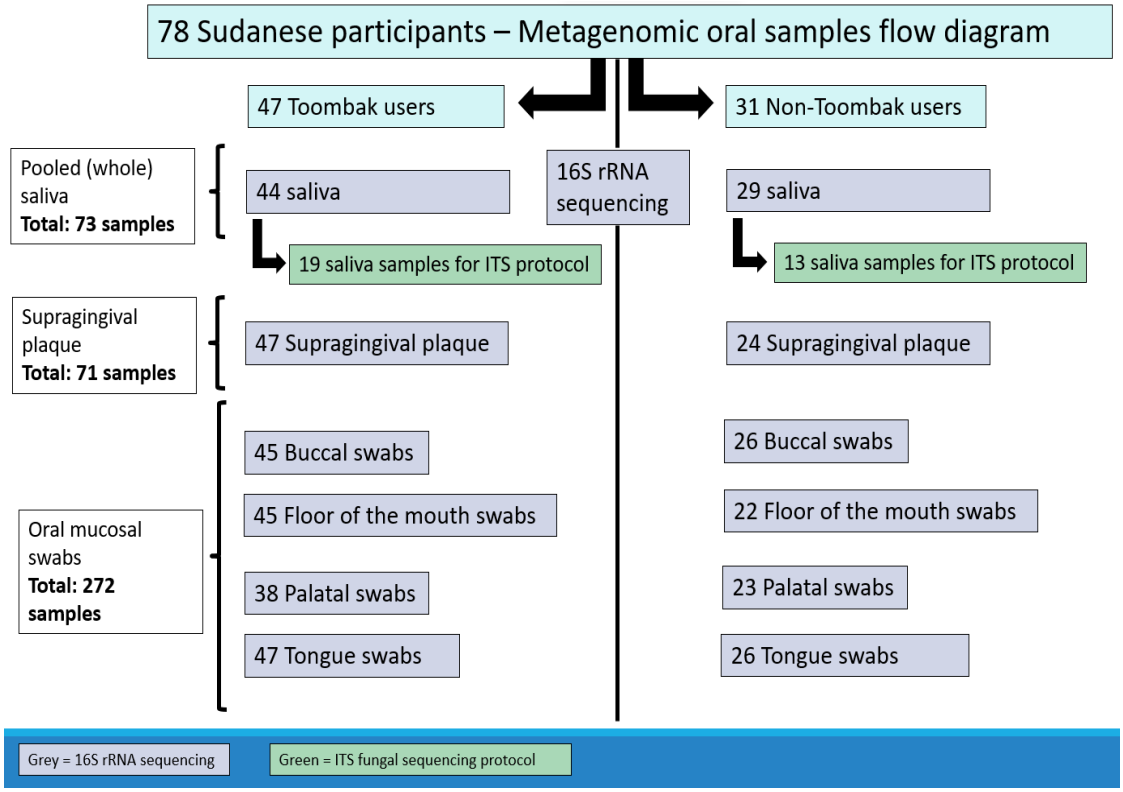
Inclusion criteria included those participants with good oral care, inactive periodontal disease, a controlled caries mouth and absence of any other dental infection. Following consent, participants did not eat or drink for at least 1 hour before sample collection. In the smokeless tobacco group, oral swab samples were obtained at least 1 hour after disregarding the most recent tobacco dip. Exclusion criteria included local conditions such as those with active periodontal disease, caries and periapical infection and systemic factors such as recent antibiotic use [< three months] and unstabilised conditions.

We achieved the collection of 72 pooled saliva, 71 supragingival plaque and 272 oral mucosal swab samples from 78 Sudanese participants. These were further categorised into 47 Toombak smokeless tobacco users and 31 non-users aged 20-70 years of age. While all Toombak users were male, in the non-user group, 18 participants were female. Pooled saliva was collected by asking the patient to remain seated, holding a sterile disposable polypropylene Falcon® [110ml] container and up to 5ml of saliva was pooled. Oral mucosal swab samples were collected by applying Puritan® Diagnostic Swabs to the required mucosal location with gentle rubbing for 10 seconds.

Supragingival plaque samples were obtained using a sterile probe which was stored in Puritan® Diagnostic Swabs for upstream analysis. The mucosal swabs were collected

from two keratinised [dorsum tongue and hard palate] and two non-keratinised locations [floor of the mouth and buccal mucosa]. Seventy-one buccal swabs [from 26 non-users and 45 Toombak users], 67 floors of the mouth swabs [from 22 non-users and 45 Toombak users], 61 hard palate swabs [from 23 non–users and 38 Toombak users], and 73 dorsum tongue swabs [from 26 non–users and 47 Toombak users] were obtained. Figure 1A summarises the metagenomic oral sample collection in this study.

Figure 1A: 78 Sudanese participants – Metagenomic oral samples flow diagram



1A: Participant samples were categorised into saliva, supragingival plaque and oral mucosa swabs

Storage and transfer

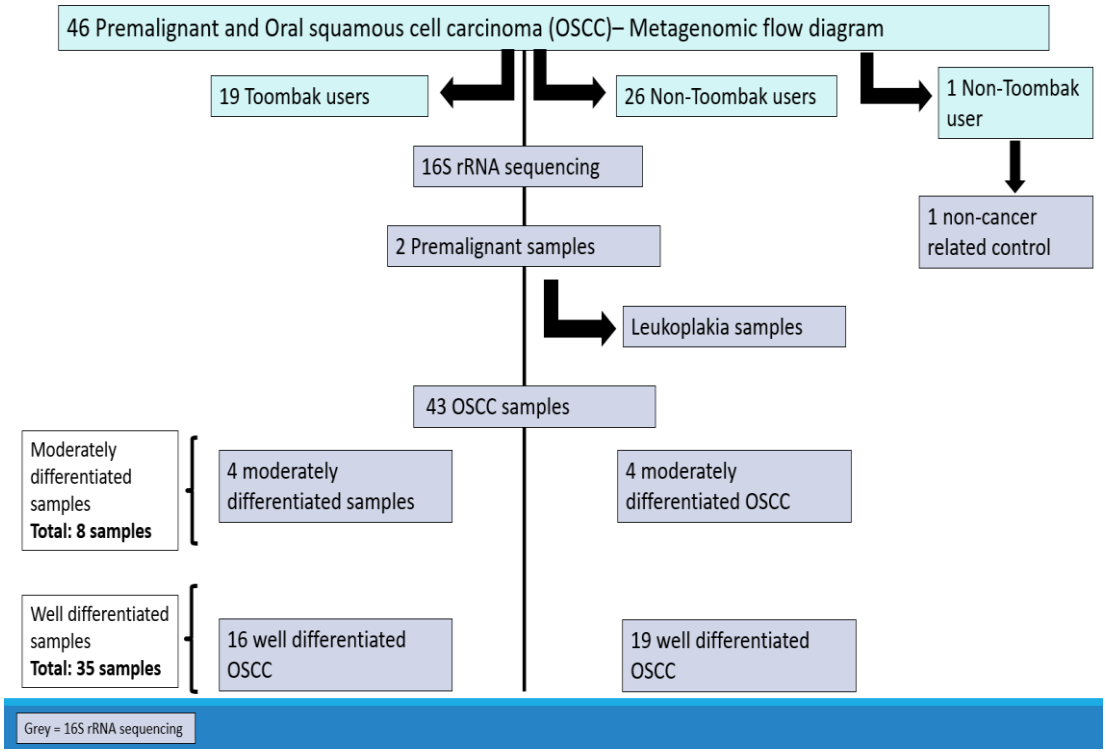
Samples were stored in an iced cooler transport box and transferred to a - 80 °C freezer in the laboratory at National Ribat University, Sudan. Finally, samples were shipped to Cork, Ireland on dry ice.

Collection of formalin-fixed paraffin-embedded oral cancer and premalignant tissue samples

Forty-six formalin-fixed paraffin-embedded samples were collected [from 43 OSCC, two premalignant lesions [Leukoplakia] and one non- premalignant condition]. The OSCC samples were further delineated as eight moderately differentiated and 35 well

differentiated OSCC samples. Samples were sourced from 29 males and 17 females, aged between 30-70 years of age. Twenty-six samples were from non-Toombak users and 19 were from Toombak users. Figure 1B summarises the metagenomic premalignant and oral squamous cell carcinoma samples collection flow obtained in this study. To yield optimum DNA from paraffin-embedded samples, we used 10µm thickness to slice sections with the first 2-3 sections being discarded. Eight sections were used for each sample. All formalin fixed paraffin embedded tissue samples were no longer than three months old from fresh sample preparation and samples were collected pre-cancer treatment. This time length was kept consistent to exclude any bias that could be introduced by time variations. Sections were immediately placed in sterile 2ml Eppendorf tubes.

Figure 1B: Premalignant and oral squamous cell carinoma Metagenomic flow diagram



1B: Samples were achieved from formalin fixed paraffin embedded tissues of premalignant, oral cancer and non – cancerous [control] sample

Metagenomic analysis

DNA extraction of pooled saliva, mucosal swab, and plaque samples

DNA extraction of samples was achieved by the DNeasy Powersoil Pro Kit [QIAGEN®, Hilden, Germany]. Eluted DNA was transferred to sterile Eppendorf

tubes and frozen at - 80^{o}C for further analysis. Swabs were first cut using a sterile scissor which allowed placement or dissolvement of samples in DNA extraction tubes. Swabs were removed after vortexing during the first stage of DNA extraction.

DNA extraction of formalin-fixed paraffin-embedded oral cancer and premalignant tissue samples

Here, QIAamp® DNA formalin-fixed paraffin embedded [FFPE] tissue kit instructions were followed. Briefly, in a microcentrifuge tube, xylene was placed with each sample to remove paraffin and centrifuged at full speed for two min, after which 1 ml of ethanol was added to remove the xylene. Samples were vortexed and centrifuged to remove residual ethanol, maintaining the pellet. Tubes were incubated at 37 ^{o}C, resuspended in 180ul tissue lysis buffer and 20µl proteinase K was added to overcome the inhibitory effects caused by formalin crosslinking of nucleic acids. All buffers were prepared according to the manufacturer's instructions and equilibrated to room temperature before protocol initiation. Samples were incubated at 56 ^{o}C for one hour and 90 ^{o}C for another hour. The incubation in the latter temperature reverses the modification of nucleic acids by formaldehyde. DNA extraction continued by adding 200µl lysis buffer and vortexing until reaching a homogenous lysate. QIAamp® MinElute columns were used to allow for the purification of high-quality DNA from the lysate and centrifuged at 6000 X g [8000 rpm] for one minute. This step was repeated until all lysates had passed through the columns, and the QIAamp® MinElute columns were empty. 500 µl of wash buffer was added and samples centrifuged for one minute at 6000 X g [8000 rpm]. Columns were placed in a clean 2ml collection tube, flow-through discarded and centrifuged at high speed; 20,000 X g [14,000 rpm] for 3 min to ensure all ethanol was removed. Finally, columns were placed in clean 1.5ml microcentrifuge tubes, incubated at room temperature for 5 minutes in 50µl buffer and then centrifuged for 1 minute to yield the final product at 20,000 X g [14,000 rpm]. Samples were stored at -80 ^{o}C for upstream analysis.

Amplicon sequencing

The Illumina Metagenomic Sequencing Library Preparation protocol for 16S rRNA gene library preparation was followed. Briefly, PCR was performed on DNA extracted samples using AMPure XP beads [Beckman Coulter], universal forward primers and

reverse primers with overhang adapters, following standard IUPAC nucleotide nomenclature, to target the 16S V3 – V4 hypervariable regions. PCR was then performed in a 50 µL reaction mixture containing 10 ng of template DNA and 2x KAPA HiFi HotStart Ready Mix. The following thermal cycling conditions were used: 3 min of initial denaturation at 95°C; 30 cycles each of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 30 s and the last step at 72°C for 5 min.

PCR clean-up was achieved with AMPure XP beads [Beckman Coulter] to purify the 16S V3 and V4 amplicon from free primers and primer dimer species. Electrophoresis was performed on the PCR-amplified products using 1.5% agarose gel with Invitrogen DNA loading dye [ThermoFisher Scientific] to verify the expected size at 550 base pairs. The gel was then visualised under 300 nm ultraviolet light. Index PCR was then continued by attaching dual indices, and Illumina sequencing adapters using Nextera XT index kits, and the following PCR thermal cycling conditions were used: 3 min of initial denaturation at 95°C; 8 cycles each of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 30 s; the last step at 72°C for 5 min. AMPure XP beads [Beckman Coulter] were utilised for final clean-up, and the final library was quantified using the Qubit® 2.0 Fluorometer and the Qubit dsDNA HS Assay kit [ThermoFisher Scientific]. Samples were then normalised, and the final library run on an Agilent high sensitivity chip and quantified by qPCR using the KAPA Illumina Library quantification kit.

In addition, the saliva samples of 13 non-users and 19 Toombak users were processed for ITS gene sequencing. This was prepared with a similar method to the 16S rRNA gene protocol and followed the methods described by Walsh et al [14]. The primers were specific to the ITS1-ITS2 regions of the ITS gene and included the Illuminas overhang adaptors [ITSF1 primer 5′ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTTGGTCATTTAGAGG AAGTAA 3′ and ITS2 primer 5′ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCTGCGTTCTTCATCG ATGC 3′].

Bacterial bioinformatics analysis

For bacterial analysis, PCR products were sequenced using the V3-V4 regions of the 16S rRNA gene on an Illumina MiSeq device two × 300 platforms [Illumina, Inc. San Diego] according to the manufacturer's instructions. The 300 base pair paired-ends FastQ product generated from 16S sequencing was merged using FLASH [fast length adjustment of short reads] using default parameters [15]. QIIME's split_libraries_fastq.py script was used for demultiplexing and filtering the fastq sequence data. Further quality filtering was performed using the USEARCH analysis tool. Single unique reads were removed; following denoising, chimera removal and grouping into operational taxonomic units [OTUs] at 97% similarity was performed using USEARCH v7 [64 bit] [16]. OTUs were aligned using Pynast [PyNAST: python nearest alignment space termination; a flexible tool for aligning sequences to a template alignment [17]. Further taxonomic ranks were set using the Basic Local Alignment Search Tool [BLAST] against the SILVA SSURef database release 132 [18]. Data were processed using Microbiome Analyst CA and Calypso version 8.84 with data filtering [removal of low abundance features less than 10%]. Data were normalised utilising total sum scaling [TSS]. The U.S National Library of Medicine, national centre for biotechnology information, was then accessed to assign species identity to those FASTA sequences with > 98% percentage homogeneity by using the BLAST® tool for bacterial 16S rRNA sequencing database [19].

Fungal bioinformatics analysis

For fungal analysis, the resulting sequencing reads were variable in length and were processed using DADA2 [v1.18.0] in R, following DADA2 pipeline workflow for ITS sequences [20, 21]. The reads were filtered to remove sequences with ambiguous 'N' bases. This was followed by primer removal from the reads using Cutadapt [v3.5] to remove primers as described in the DADA2 ITS pipeline workflow [22]. The quality of the reads was inspected, and the reads were further filtered and trimmed to retain reads that were a minimum of 50 base pair long with default parameters. Dereplicating and merging of paired-end reads was done following the workflow, following which chimeras were removed and taxonomy was assigned to the resulting ASVs [Amplicon Sequence Variants] using the UNITE ITS database. The resulting ASV table was used for further analysis with Microbiome Analyst.

Data was not rarefied but normalised utilising total sum scaling [TSS]. Calypso version 8.84 and Microbiome analyst CA [23] were also used for the statistical and interpretational metagenomic data.

4.4 Results and Discussion

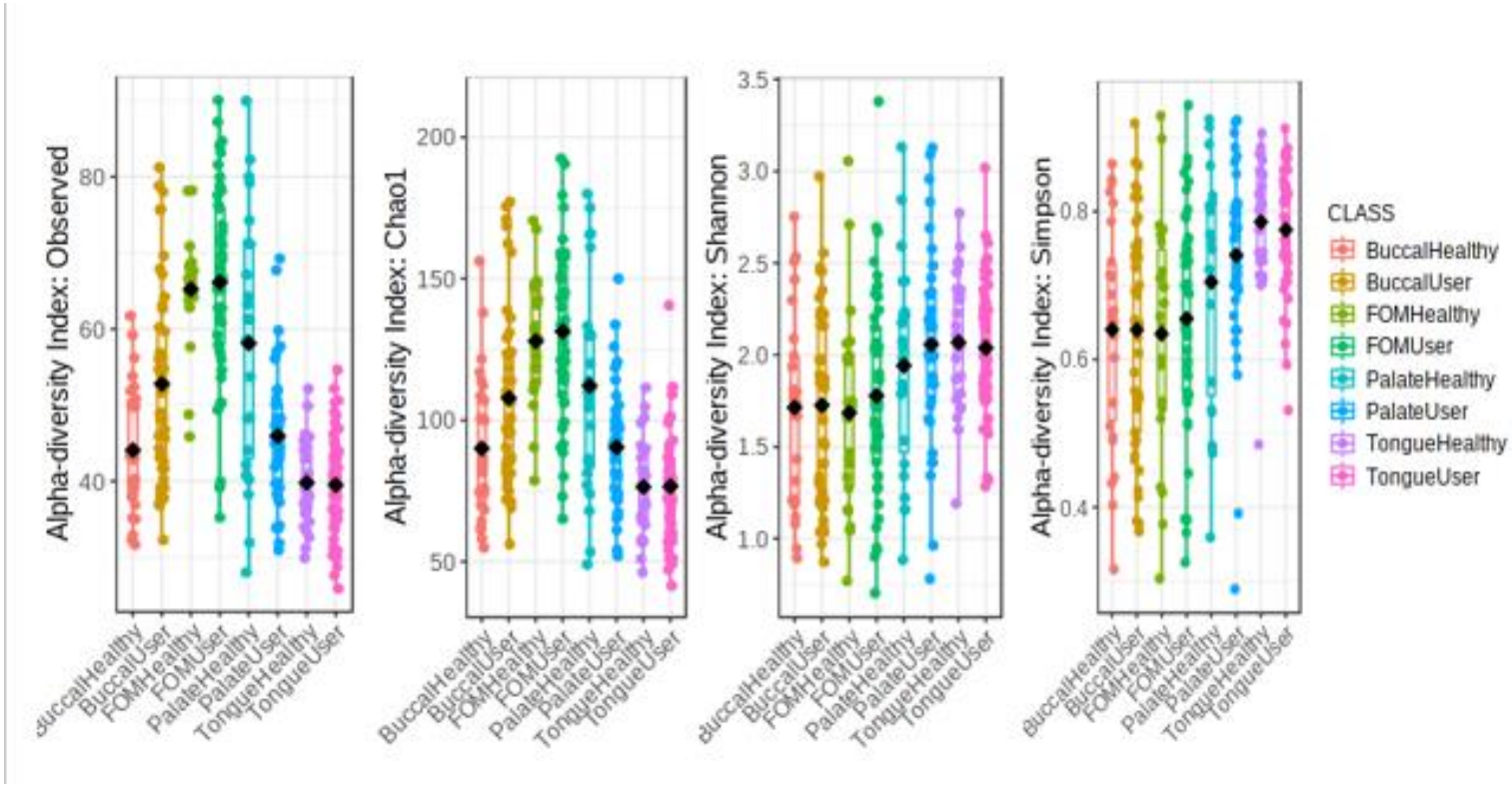
For the plaque and oral mucosal samples, a total of 1344 observations were encountered with a minimum read of 8.0 [controls] and a maximum read of 2008537.0.

Alpha and Beta diversity

Samples were grouped by mucosal location [Buccal, Floor of the mouth, Palate and Tongue] and status of Toombak use [Healthy=non-user, User= Toombak user]. We found that mucosal location in the oral cavity and keratinisation status is a stronger predictor of microbiome variation compared to Toombak use. Non-keratinised mucosal locations [buccal and floor of the mouth] had similar patterns in microbiome diversity while keratinised locations also harboured close resemblances [palate and tongue]. However, utilising Toombak alters the oral microbiome regardless of optimum oral health.

Four alpha diversity measures, Chao1, Shannon and Simpson index, and richness were assessed [Figure 2]. Alpha diversity was found to be significantly varied between groups. Chao 1, a measure of variation between the rare or unobserved species, showed that the tongue mucosa had the lowest Chao 1 diversity in both users and non-users of Toombak. In the palatal mucosa, those who used Toombak, however, had lower Chao 1 indices compared to non-users. Shannon indices were more distinct by location rather than Toombak use [p=0.00076]. Simpson index showed a higher but narrower tongue and palatal microbiome compared to the remaining groups [Simpson's index community diversity closest to 1]. Richness [p=3.6849e-34] was highest on the floor of the mouth mucosa and lowest on the tongue mucosa. The median richness of the buccal mucosa microbiome was markedly lower for non-users compared to users of smokeless tobacco.

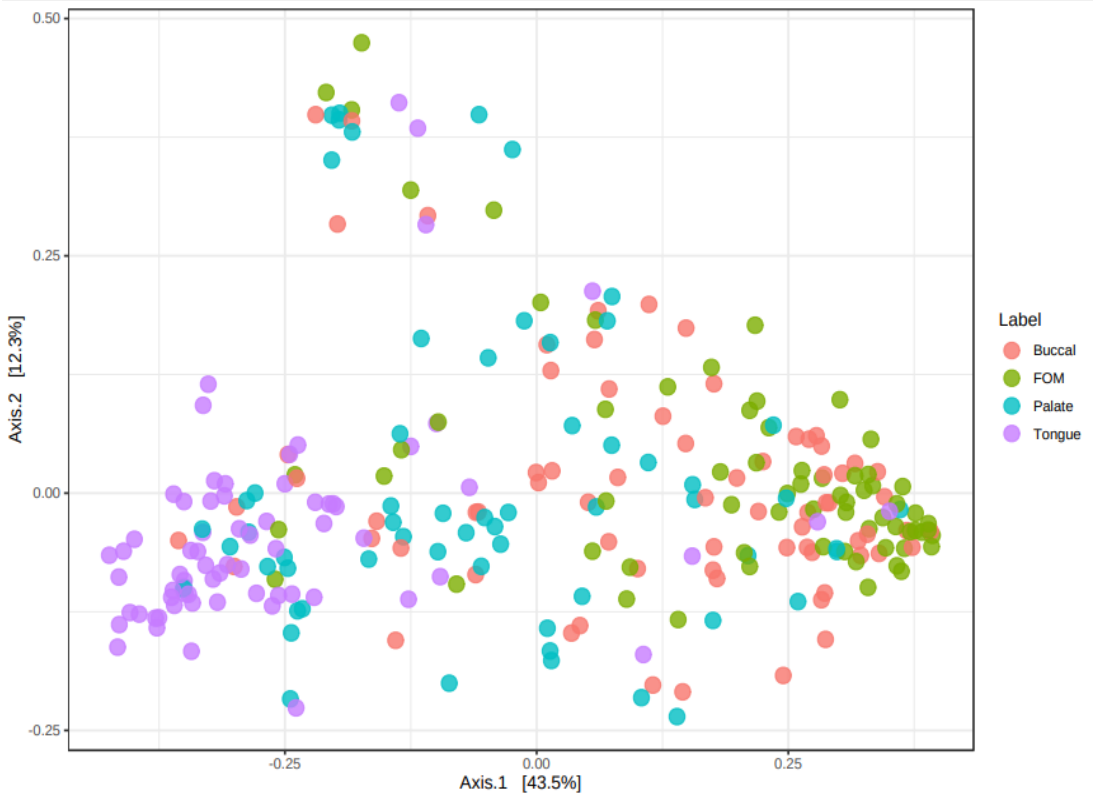
Figure 2:



2: Shannon index [p=0.00076], Richness or Observed [p=3.6849e-34] and Simpson's index [p=3.3603e-08] were all significant between groups. Richness was lowest for tongue microbiome and similar in the remaining locations. The tongue and palatal mucosa [Simpson's index community diversity closest to 1] show the closest clustering.

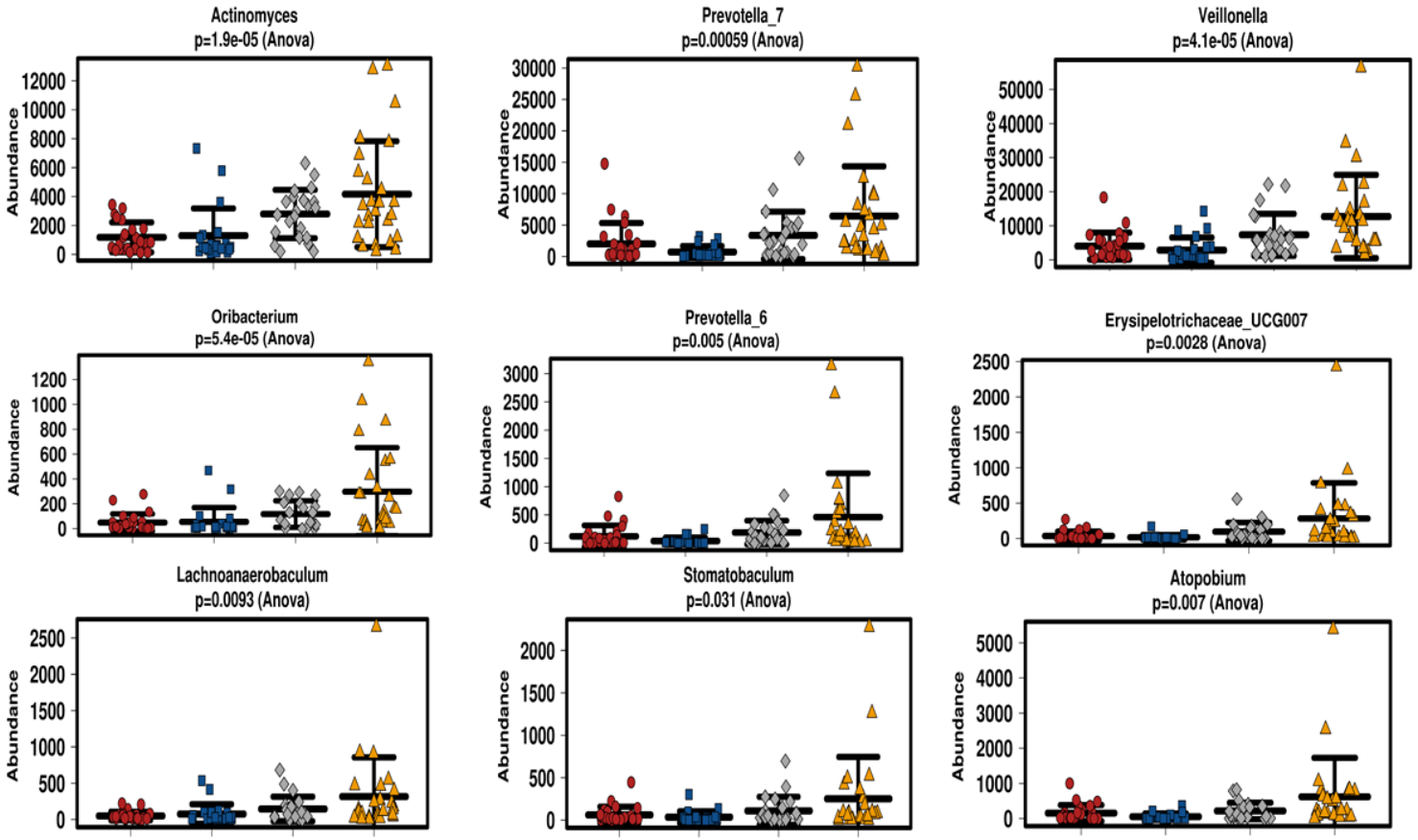
Beta diversity [Figure 3a] between the oral mucosal regions was significantly distinct [$p < 0.001$ $R^2 = 0.22$], particularly between the dorsum tongue and the buccal/floor of the mouth, which were found to cluster together. The palatal microbiome was 'sandwiched' between the microbiome of the other mucosal environments. Such diversity could be due to the genera and species composition found in the various mucosal habitats. The tongue microbiome harboured an abundance of *Actinomyces, Prevotella_6, Prevotella_7, Veillonella, Oribacterium, Erysipelotrichaceae_UCG007, Lachnoanaerobaculum, Stomatobaculum* and *Atopobium* [Figure 3b]. The floor of the mouth mucosa harboured increases in the genera *Actinobacillus* and *Bergeyella*, while the palatal mucosa was increased in *Kingella,* after which the buccal cheek mucosa was found to be significantly increased in the genus *Shigella* [Figure 3C].

Figure 3a:



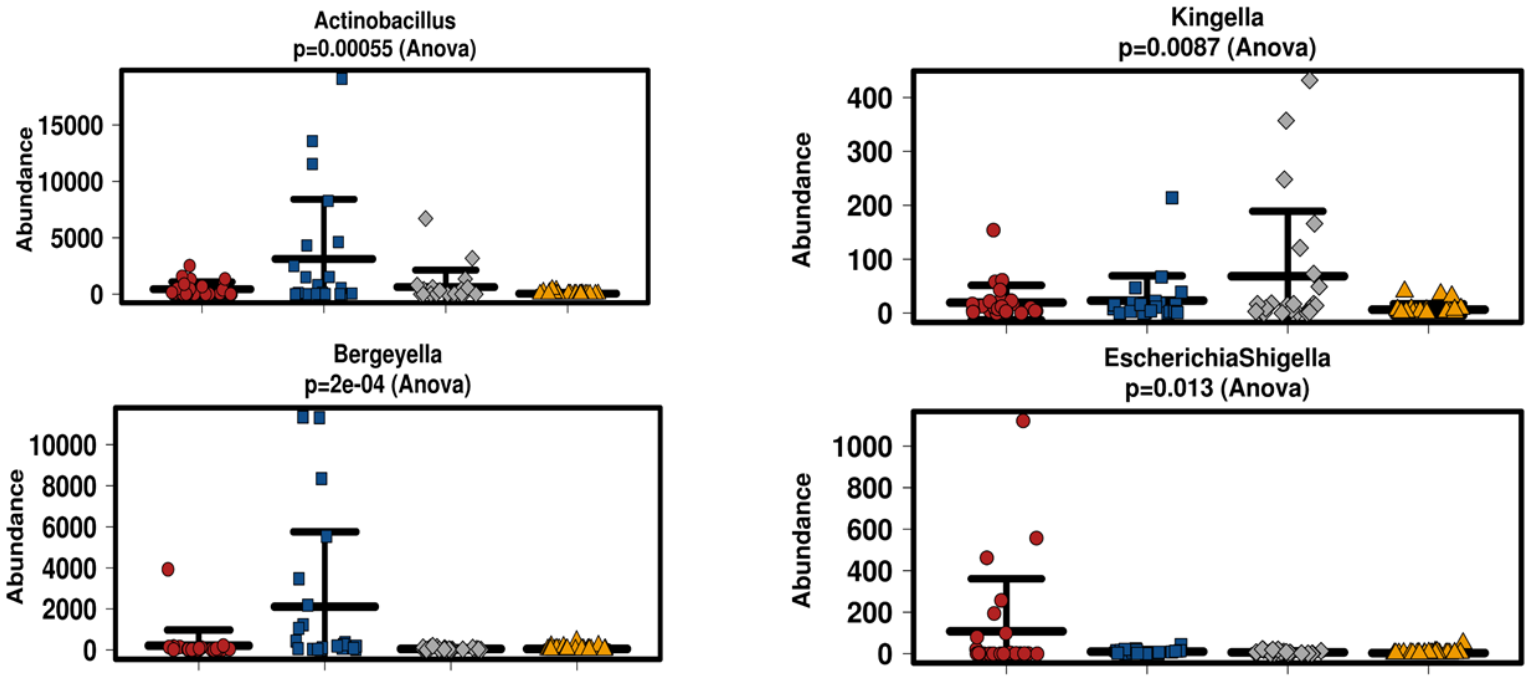
3a: β diversity between the oral mucosal regions. Significant β diversity between the oral mucosal regions [p value < 0.001]. R^2 = 0.22. The tongue microbiome [purple] is most distinct. Non-keratinised buccal and floor of the mouth mucosa are the most clustered together while the palatal microbiome is centrally bridged.

Figure 3b:



3b: Tongue microbiome abundances amongst the oral Sudanese microbiomes. Red = buccal, blue = floor of the mouth, grey = palate and gold = tongue. Actinomyces, Prevotella, Veillonella are amongst the most significantly abundant genera of tongue microbiome.

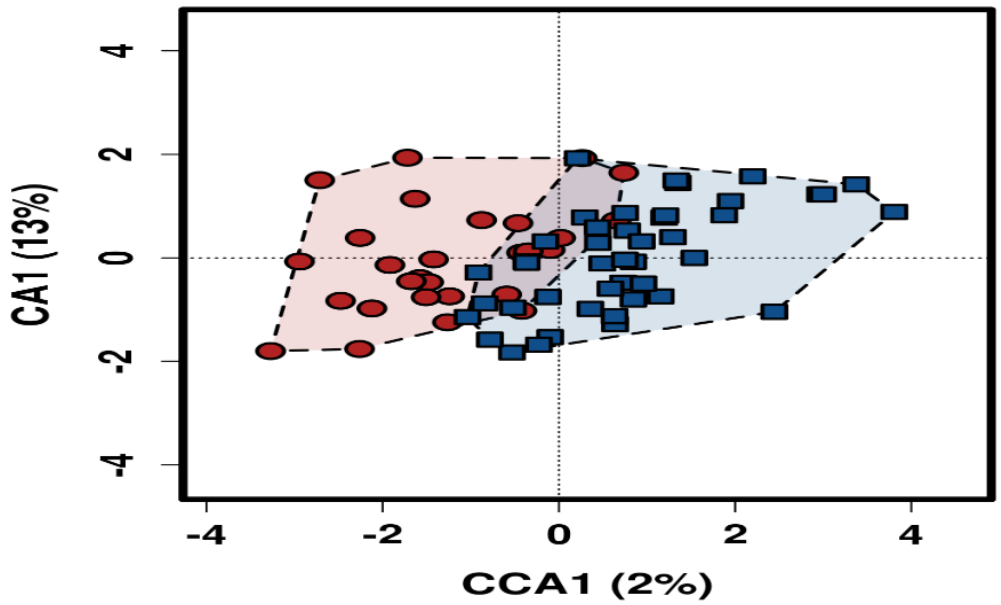
Figure 3c:



3c: Abundant genera in the other mucosal locations. Red = buccal, blue = floor of the mouth, grey = palate, and gold = tongue. Actinobacillus and Bergeyella are most abundant in the floor of the mouth, Kingella in the palate and Shigella in the buccal mucosa.

We further used canonical correspondence analysis [CCA] as a multivariate error-reducing technique analysing the abundance of the oral microbiome within users and non-users of Toombak on the known gradient oral health minimising matrix errors in OTU- oral health environmental relationships [24]. CCA plotting [Figure 3d] showed that after accounting for oral health, Toombak and non-users of Toombak still harboured distinct oral microbiome variations.

Figure 3d:

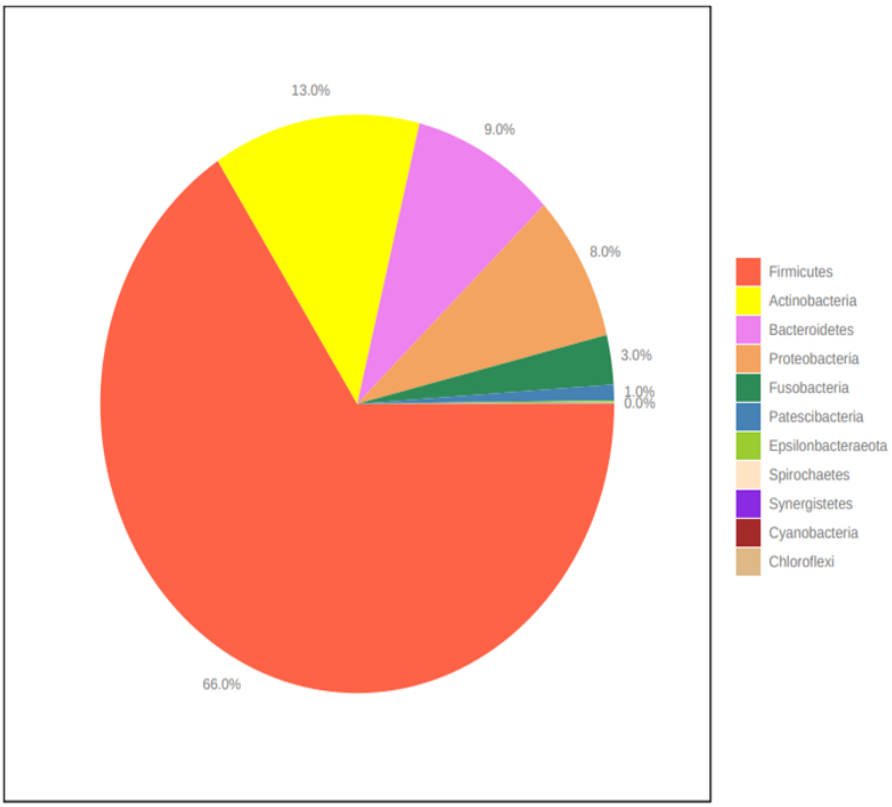


3d: Canonical correspondence analysis [CCA] plot of users and non-users of Toombak. Red= non-user, blue = Toombak user. Oral health was utilised as the known gradient in minimising matrix errors in OTU – oral health relationships. Accounting for oral health, Toombak harboured oral microbiome differences that could be appreciated compared to non-users.

Pro-carcinogenic phyla and premalignant associated classes of bacteria are abundant in the saliva of Toombak users

The most abundant phyla in pooled saliva were found to be *Firmicutes* [66%], *Actinobacteria* [13%], *Bacteroidetes* [9%], *Proteobacteria* [8%] and *Fusobacteria* [3%] [Figure 4a].

Figure 4a:

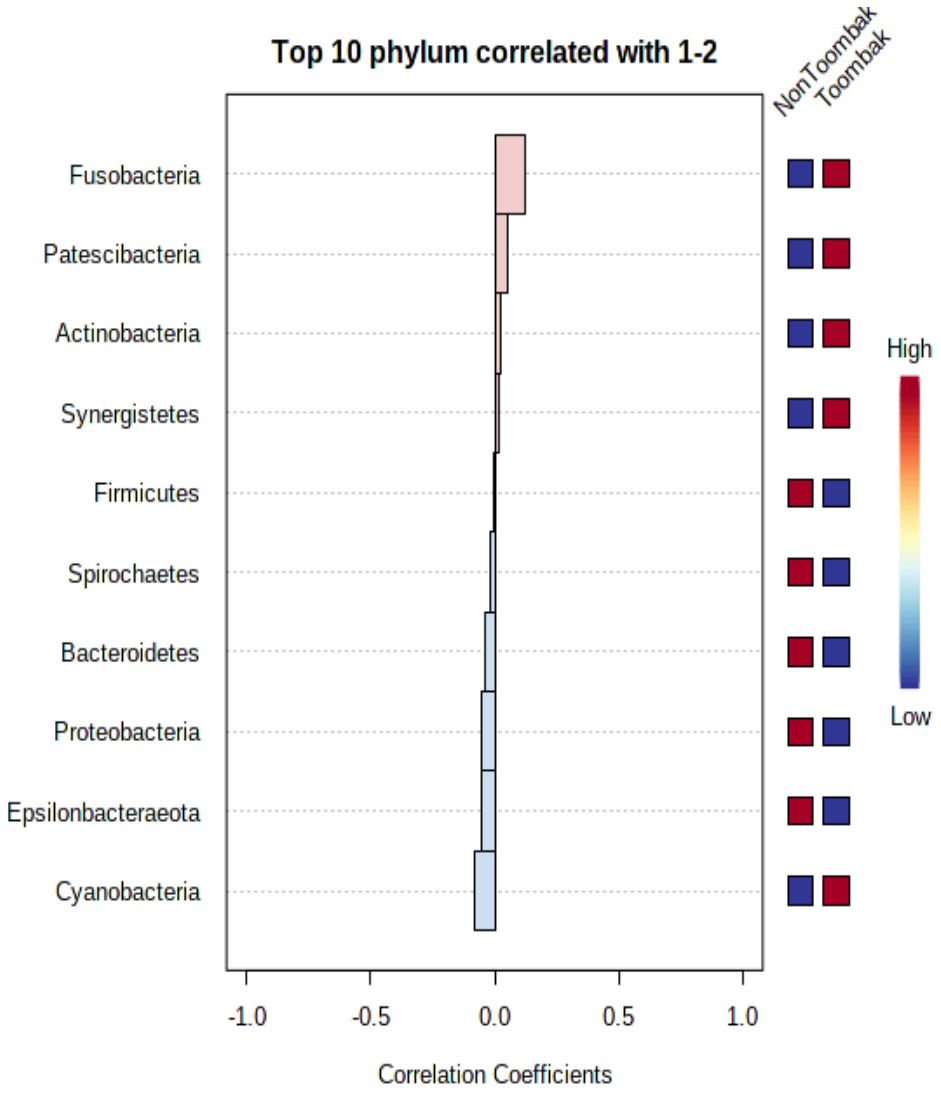


4a: Pie chart percentage compostion of phyla from pooled saliva of both users and non-users of Toombak. Firmicutes [66%], Actinobacteria [13%], Bacteroidetes [9%], Proteobacteria [8%] and Fusobacteria [3%] are the most abundant phyla.

We further utilised correlation plotting to visualise the top ten phyla varied in abundance between the saliva of users and non-users of Toombak. Here, the phyla *Fusobacteria* and *Patescibacteria* [also known as *Candidate Phyla Radiation*] were the most correlated for Toombak use while *Cyanobacteria* were associated with non-users of Toombak [Figure 4b]. Both *Fusobacteria* and *Patescibacteria* have been shown to be significantly enriched in various forms of gastric cancer while *Fusobacteria* in particular have been shown to promote cancer progression and the invasion of cancer into surrounding tissue [25]. They are said to be late colonisers in healthy individuals and their abundance amongst Toombak users could have a role to play in the early developmental changes of OSCC [26]. This is likely due to

Fusobacteria possessing FadA proteins that are associated with the attachment, invasion, and adherence of cancer cells [27].

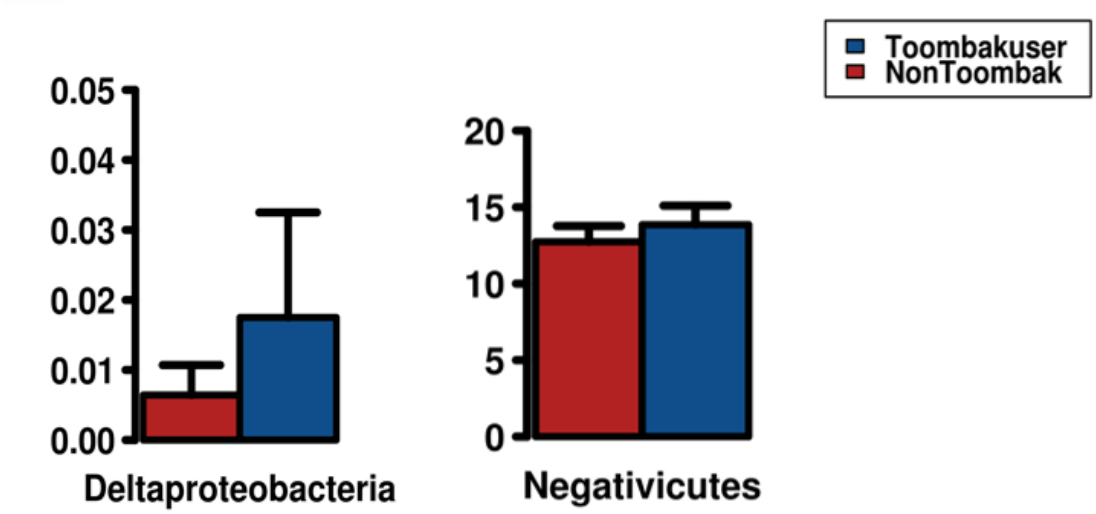
Figure 4b:



4b: Correlation plot of ten most abundant phyla in the Sudanese oral cavity of users and non-users of Toombak. Fusobacteria are the most correlated with Toombak use compared to Cyanobacteria in non-users.

Gracilibacteria were also found to be abundant in Toombak users. These are part of the *Patescibacteria* group that contain unique antimicrobial peptides and heavy metal [nickel] resistance pathways which may make them more tolerant to the heavy metal content of Toombak [28, 29]. Classes of bacteria associated with premalignancy were found to be increased in Toombak users. *Negativicutes* amongst Toombak users may predispose them to the development of oral leukoplakia, a premalignant lesion [30, 31]. *Deltaproteobacteria*, a class of diverse sulphate-reducing bacteria can be pathogenic and cancer-promoting due to their ability to produce hydrogen sulphide [32] [33] [Figure 4c] and were found to be high in Toombak users.

Figure 4c:



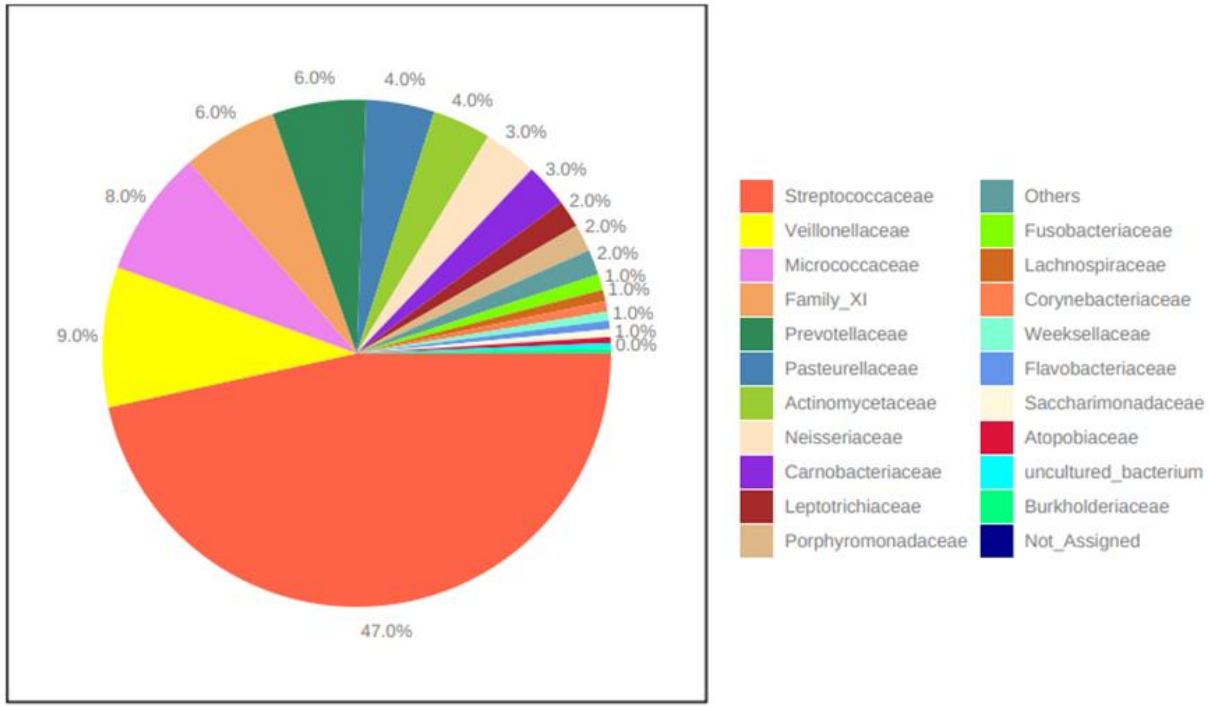
4c: Log-bar values of class composition of pooled saliva of both users and non-users of Toombak. Increased abundances in the classes Deltaproteobacteria and a slight increase in Negativicutes were found among Toombak users

Microbial family and genera discrepancies between saliva and plaque of users and non-users of Toombak

Staphylococcaceae were the most significantly differentiated family between users and non-users of Toombak [q=0.037]. At family level, the saliva of users and non-users of Toombak indicated similar profiles of *Streptococcaceae* [47%], *Veillonellaceae* [9%] and *Micrococcaceae* [8%] [Figure 4d]. Sixty-five core microbiome genera were identified in the saliva of users and non-users of Toombak.

Three genera, however, *Streptobacillus*, *Shuttleworthia* and *Eubacterium yuri* group, were unique to non-users of smokeless tobacco. In comparison, four genera, *Fretibacterium, Filifactor, Corynebacterium_1* and *Alysiella*, were unique to Toombak users.

Figure 4d:


6.0%
4.0%
4.0%
3.0%
3.0%
2.0%
2.0%
2.0%
1.0%
1.0%
1.0%
1.0%
0.0%
6.0%
8.0%
9.0%
47.0%
Streptococcaceae
Veillonellaceae
Micrococcaceae
Family_XI
Prevotellaceae
Pasteurellaceae
Actinomycetaceae
Neisseriaceae
Carnobacteriaceae
Leptotrichiaceae
Porphyromonadaceae
Others
Fusobacteriaceae
Lachnospiraceae
Corynebacteriaceae
Weeksellaceae
Flavobacteriaceae
Saccharimonadaceae
Atopobiaceae
uncultured_bacterium
Burkholderiaceae
Not_Assigned

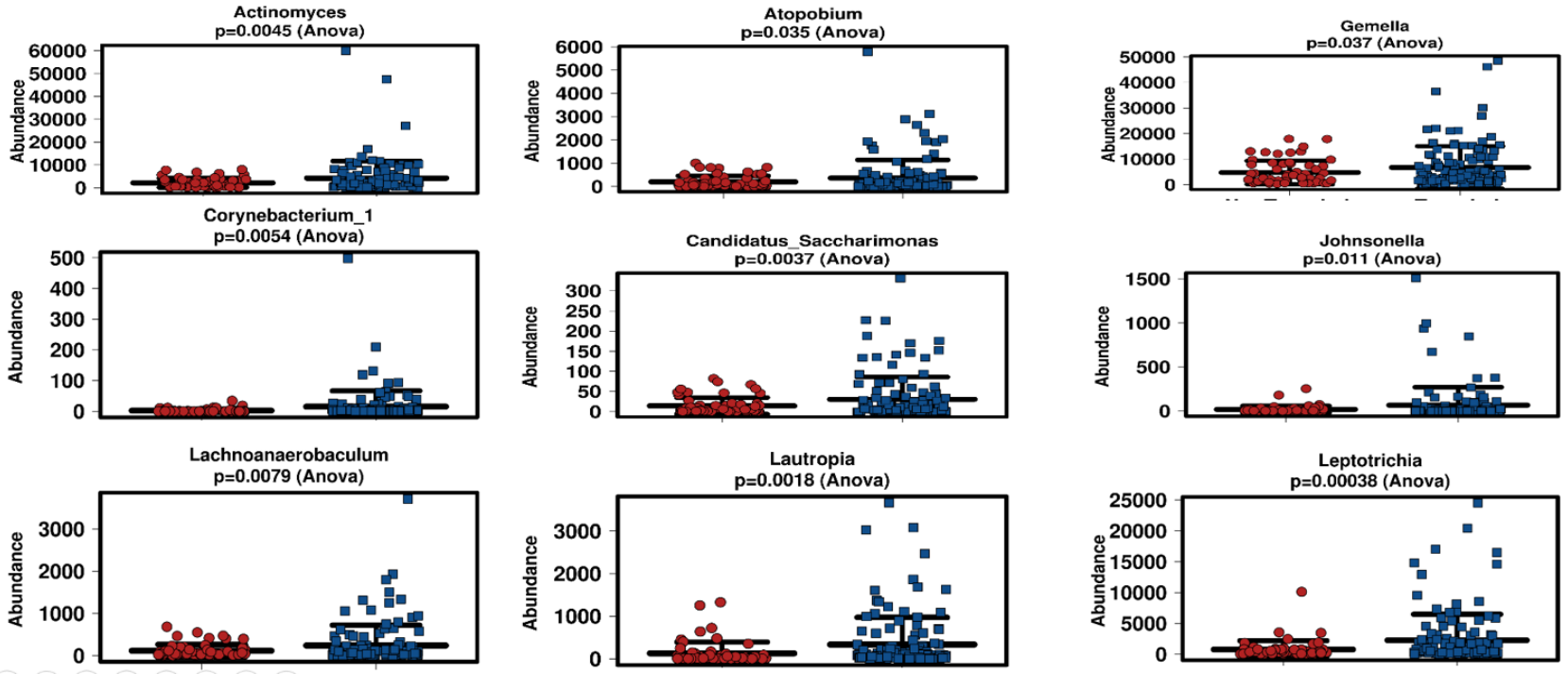
4d: Pie chart percentage composition of families from pooled saliva of both users and non-users of Toombak. Streptococcaceae [47%], Veillonellaceae [9%] and Micrococcaceae [8%] were the most abundant while Staphylococcaceae were the most significant differentiating family between users and non-users of Toombak with a p-adjusted value of 0.037.

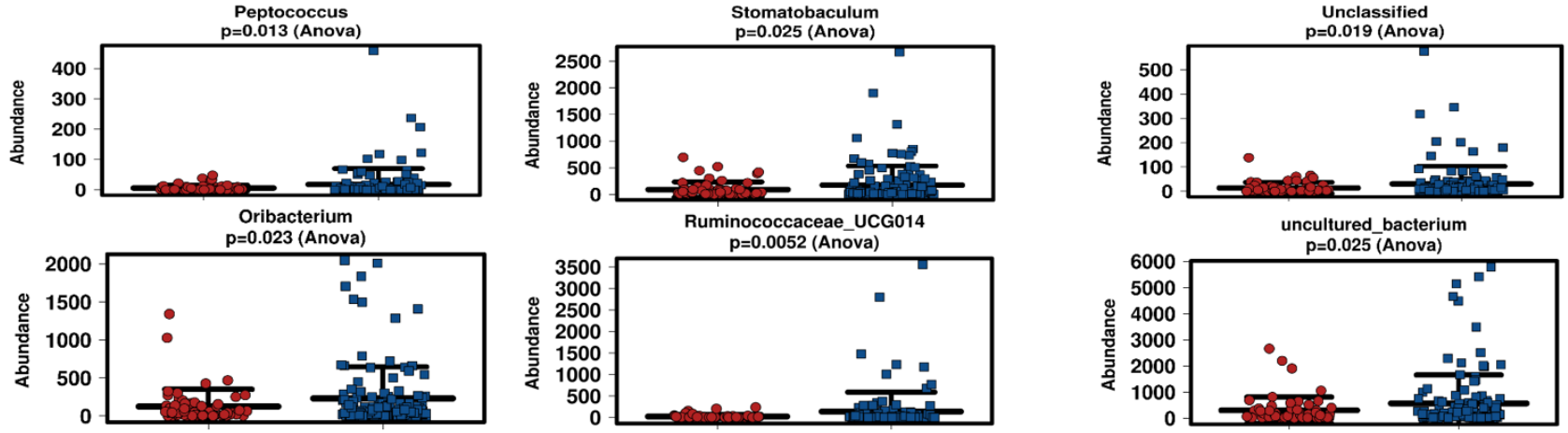
The most significantly abundant microbial genera amongst Toombak users are highlighted in Figure 5a which include *Actinomyces* [p= 0.0045], *Corynebacterium_1* [p=0.0054], *Lachnoanerobaculum* [p=0.0079], *Lautropia* [p=0.0018], *Atopobium* [p=0.035], *Gemella* [p=0.037], *Johnsonella* [p=0.011], *Leptotrichia* [p=0.00038], *Candidatus_Saccharimonas* [p=0.0037], *Peptococcus* [p=0.013], *Ruminococcaceae_UCG014* [p= 0.0052], *Oribacterium* [p=0.023] and unclassified and uncultured bacterium.

In a study assessing the tongue microbiome of smokeless tobacco users in Saudi Arabia, species of *Actinomyces* and *Oribacterium* were also found to be significantly abundant in smokeless tobacco users [34] and in smokeless tobacco users from Guam [35]. Elevated levels of *Atopobium* has been found in OSCC biopsies, and *Leptotrichia*, *Gemella* and *Oribacterium* in the saliva of OSCC patients [36]. *Jonhsonella*, found to be abundant in Toombak users, has been highly associated with oral tumour sites indicating specific microbiome dispositional trends in those that utilise Toombak towards oral cancer development [37].

In a culture-based study from Sudan by Ali et.al. [2014], oral swab samples from long-term smokers showed abundances in *Peptococcus*, a genus also found to be significantly increased in those who develop oral cancer [38]. *Lachnoanaerobaculum* was also abundant in the saliva of Toombak users, similar to the findings from other studies [39]. *Lachnoanaerobaculum* was abundant in smokeless tobacco users as well as oral cancer patients in studies from India [40, 41]. This genus has also been found to be significantly increased in a group of pipe or ‘medwakh’ smokers from the United Arab Emirates [42]. Another elevated genus found in Toombak users, *Leptotrichia* has been associated with more severe levels of oral epithelial dysplasia [43] including in those who developed pancreatic cancer [44].

Figure 5a:


Actinomyces
p=0.0045 (Anova)
Abundance
60000
50000
40000
30000
20000
10000
0
Atopobium
p=0.035 (Anova)
Abundance
6000
5000
4000
3000
2000
1000
0
Gemella
p=0.037 (Anova)
Abundance
50000
40000
30000
20000
10000
0
Corynebacterium_1
p=0.0054 (Anova)
Abundance
500
400
300
200
100
0
Candidatus_Saccharimonas
p=0.0037 (Anova)
Abundance
300
250
200
150
100
50
0
Johnsonella
p=0.011 (Anova)
Abundance
1500
1000
500
0
Lachnoanaerobaculum
p=0.0079 (Anova)
Abundance
3000
2000
1000
0
Lautropia
p=0.0018 (Anova)
Abundance
3000
2000
1000
0
Leptotrichia
p=0.00038 (Anova)
Abundance
25000
20000
15000
10000
5000
0



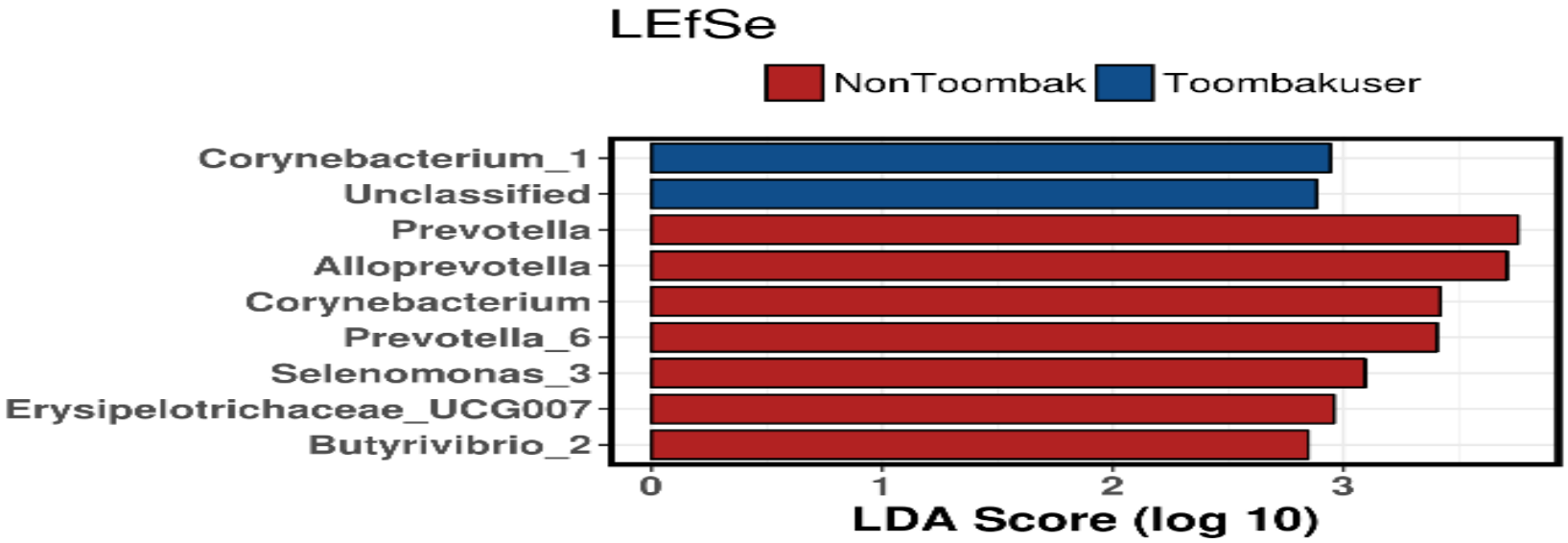
5a: The most abundant genera with p value significance amongst Toombak users and non-users. Actinomyces [p= 0.0045], Corynebacterium_1 [p=0.0054], Lachnoanerobaculum [p=0.0079], Lautropia [p=0.0018], Atopobium [p=0.035], Gemella [p=0.037], Johnsonella [p=0.011], Leptotrichia [p=0.00038], Candidatus_Saccharimonas [p=0.0037], Peptococcus [p=0.013], Ruminococcaceae_UCG014 [p= 0.0052], Oribacterium [p=0.023] and unclassified and uncultured bacterium.

Leptotrichia abundance in the oral cavity has been associated with head and neck cancer [45] and they were significantly increased in a cohort of heavy smokers in one study from the United Arab Emirates [46].

Utilising Mann Whitney Kruskal Wallis with p adjusted values [q value]; two genera were significantly increased in Toombak users; *Corynebacterium_1* [q= 0.0286] and *Staphylococcus* [q= 0.0286] while *Scardovia* was significantly abundant in non-users of Toombak [q=1.9435e-4]. Both *Coryenbacterium_1* and *Staphylococcus*, found to be abundant in the saliva of Toombak users have also been found to be abundant in the Toombak microbiome composition [1]. *Corynebacterium* has also been found to be significantly increased in the saliva of smokeless tobacco users from other studies [47, 48]. LEfSe plotting on the pooled saliva dataset further indicated *Corynebacterium_1* and unclassified bacterium to be discriminant for Toombak use *while Prevotella, Selenomonas, Erysipelotrichaceae* and *Butyrivibrio _2* are discriminant for the saliva of non-users of Toombak [Figure 5b].

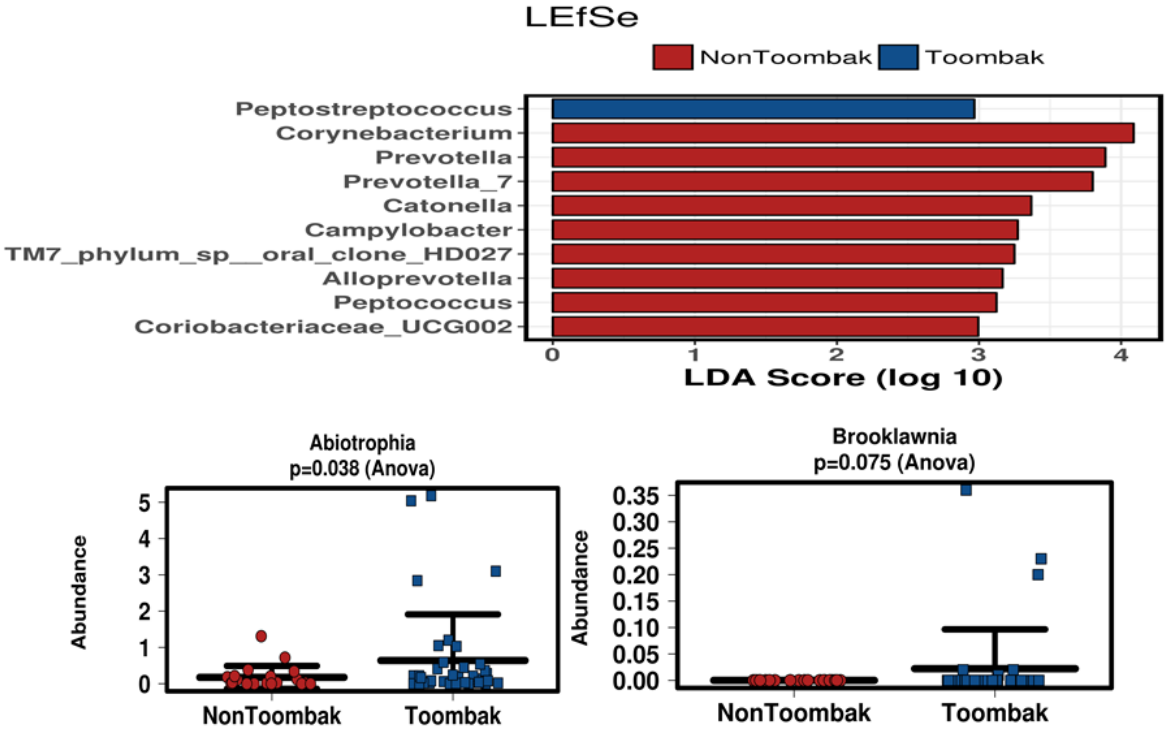
In supragingival plaque, LEfSe plotting showed that *Peptostreptococcus* was the distinct genus amongst Toombak users while the genera *Corynebacterium, Prevotella* and *Catonella* were the most distinguishing plaque genera of non-users [Figure 5c]. Fifty eight genera were shared between users and non-users of Toombak. *Comamonas* was only found in the plaque of Toombak users while *Abiotrophia* [p=0.038] and *Brooklawnia* [p=0.075] were significantly abundant in Toombak users [Figure 5d]. Non-users of Toombak harboured *Scardovia, Peptococcus, Lactobacillus, Fretibacterium, Catonella* and *Bifidobacterium* in the plaque microbiome composition.

*Figure 5b**:***



5b*: LEfSe plotting on pooled saliva dataset indicated that Corynebacterium_1 and unclassified bacterium were discriminant for Toombak use while Prevotella, Selenomonas, Erysipelotrichaceae and Butyrivibrio_2 were discriminant for non-users.*

*Figures 5c***:**


LEfSe
NonToombak
Toombak
Peptostreptococcus
Corynebacterium
Prevotella
Prevotella_7
Catonella
Campylobacter
TM7_phylum_sp__oral_clone_HD027
Alloprevotella
Peptococcus
Coriobacteriaceae_UCG002
0
1
2
3
4
LDA Score (log 10)
Abiotrophia
p=0.038 (Anova)
Abundance
5
4
3
2
1
0
NonToombak
Toombak
Brooklawnia
p=0.075 (Anova)
Abundance
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
NonToombak
Toombak

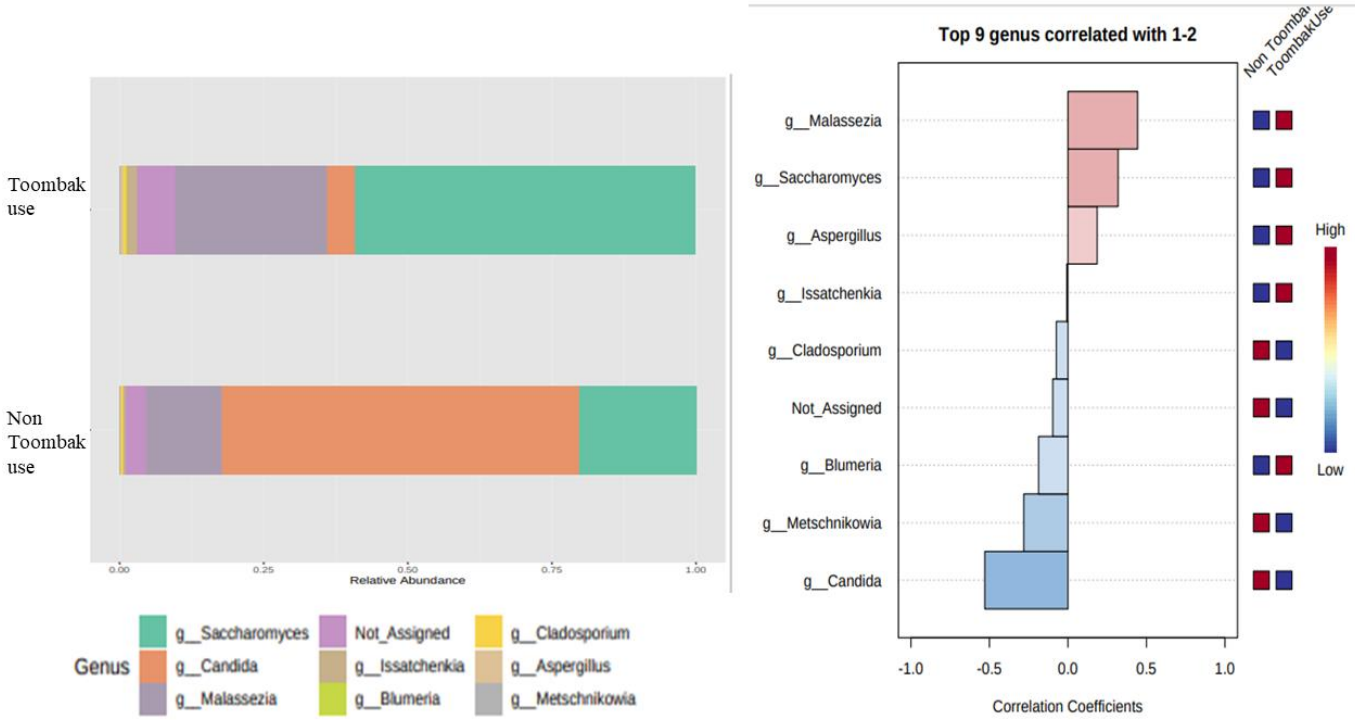
5c: Supragingival microbiome in the Sudanese population. LEfSe plotting showed that Peptostreptococcus were the most distinct genus amongst Toombak users while the genera Corynebacterium, Prevotella and Catonella were the most distinguishing in non-users in plaque samples. 4d Abiotrophia [p=0.038] and Brooklawnia [p=0.075] were significantly abundant in the plaque of Toombak users.

Aspergillus abundance amongst Toombak users likely exposes them to carcinogenic aflatoxin compounds

In this study, we found more than a three-fold enrichment of *Aspergillus* in the oral cavity of Toombak users [78.93%] compared to non-users [21.07%]. Aflatoxins produced by *Aspergillus* species can have long-term adverse effects in the causation of cancer [49]. The abundance of *Aspergillus* in the oral cavity is now strongly associated with premalignant conditions such as Lichen planus and thus could predispose Toombak users to developing conditions that favour OSCC development [50]. *Aspergillus* abundance has also been associated with lymph node involvement in OSCC patients from Sudan [13].

In Toombak users the mycobiome was found to be abundant in *Blumeria* [61.7%], *Issatchenkia* [61.52%] and *Saccharomyces* [62.11%] compared to non-users [Figure 6a]. *Metschnikowia* [36.22%] and *Cladosporium* [65.07%] were lower in abundance amongst Toombak users compared to non-users. *Malassezia* was found to be positively correlated with Toombak use while *Candida* was associated with non-users [Figure 6b]. Indeed, there was a marked loss of *Candida* in the oral cavity in those utilising smokeless tobacco [4.33%] compared to non-users [95.67%] [q=4.758E-4]. Interestingly, in a recent study of salivary mycobiome of Sudanese OSCC patients, *Candida* was also reduced amongst Toombak users [13]. One reason for this could be the inhibitory effects that high nicotine levels [found in Toombak] have been found to have on the growth of *Candida* [51].

Figure 6a-6b:



6a: Fungal entities in pooled saliva amongst Sudanese users and non-users of Toombak. Marked loss of Candida presence in the oral cavity of Toombak users and increased Saccharomyces and Aspergillus. Unassigned species also abundant amongst Toombak users. 6b: Correlation differentiators of fungal inhabitance in the oral cavity between Toombak and non-users. Abundances in Malassezia, Saccharomyces and Aspergillus were correlated positively with Toombak use while Candida were the most correlated with non-users of Toombak.

Fungal species are found to be more virulent with smokeless tobacco use.

Candida albicans was found to be the most dominant species amongst non-users, while *Candida tropicalis* was the most abundant species amongst Toombak users. *Candida tropicalis* is one of the most virulent of the *Candida* species and are resistant to many antifungal medication [52, 53]. This species has also been previously isolated from OSCC samples [54]. *Malassezia restricta* was found to be more abundant in Toombak users [63.02%] compared to non-users [36.98%] where q=0.046. Abundance of *Malassezia* may be linked to better survival rates amongst OSCC patients [13].

Toombak use allows for significant microbiome variations throughout the four oral mucosal locations

In four mucosal locations, low abundance features were removed based on 10% prevalence and low variance features were further removed based on 5% standard deviation. For data OTU analysis in both users and non-users of Toombak, 274 OTUs were assessed from the buccal mucosa, 208 OTUs from the dorsum tongue mucosa, 254 OTUs from the floor of the mouth and 271 OTUs were assessed from the palatal mucosa. We further evaluated the core microbiome of each mucosal location between users and non-users of Toombak. The buccal mucosa had the highest similarity of core microbiome between users and non-users of Toombak; 126 core genera while the dorsum tongue had the least similarity in core microbiome between users and non-users of Toombak; 42 core genera. 53 core microbiome genera were present on the floor of the mouth and 55 core microbial genera on the palate between users and non-users of Toombak.

The buccal mucosa [inner cheek lining]

The relative abundance of *Actinobacteria* and *Fusobacteria* were increased in the buccal mucosa of Toombak users compared to non-users of Toombak. *Moraxellaceae* [q=0.037], *Leptotrichiaceae* [q= 0.019], and *Staphylococcaceae* [q=0.037] were the most significantly abundant families amongst the buccal microbiome of Toombak users. *Staphylococcaceae* exhibited the highest fold change increase [4.259], while *Enterobacteriaceae* [0.037] using Deseq2 coverage were increased amongst non-users of Toombak in this mucosal location with a fold change increase of 7.691. *Leptotrichia* [q=0.029], *Staphylococcus* [q=0.04] and *Cutibacterium* [q=0.04] were the most

significantly abundant genera amongst the buccal mucosa of Toombak users while LEfSe highlighted *Staphylococcus, Cutibacterium and Corynebacterium_1 [LDA > 2]* as the most discriminant genera of the buccal microbiome amongst this group. In the non-users of Toombak, *Shigella* [q= 0.03] and *Prevotella_6* [q= 0.04] were significantly abundant while *Scardovia* and *Prevotella_6* could distinguish the buccal microbiome of non-users [Figure 7a].

Genera found in abundance in the buccal microbiome of Toombak users may be related to the increased risk of oral tumour development [47]. *Fusobacteria* and the genera *Staphylococcus* and *Corynebacterium_1 a*re able to reduce nitrate and nitrite to produce carcinogenic tobacco-specific nitrosamine compounds [29]. Studies from Indian smokeless tobacco users have also highlighted an abundance of *Staphylococcus* amongst OSCC smokeless tobacco users [55]. *Cutibacterium* are likely introduced by users fingers into the oral cavity during Toombak placement [1]. Biofilm formation is enhanced with smokeless tobacco use which can improve epithelial adherence of *Staphylococcus* [46]. In addition, oral *Staphylococcus* abundance amongst Toombak users may be linked to an increased fungal load in the oral cavity [56] likely through chemical interplay involving metabolites, quorum sensing effectors and stable biofilm development [57]. Oral *Staphylococcus* abundance found in Toombak users may also be a reservoir for the development of systemic infection, is a source of methicillin-resistant strains and could pose an increased susceptibility to blood-related infections [58, 59]. *Staphylococcus* may further play a sinister role in the acidic hypoxic environments of oral cancer [47] and have been isolated from OSCC sites [60].

Species in abundance in the buccal microbiome of Toombak users included *Gemella morbillorum* [98.28% per ident], *Prevotella nigrescens* [99.35% per ident], *Staphylococcus caprae* [100% per ident] and *Lachnoanaerobaculum gingivalis* [99.32% per ident]. *Gemella morbillorum* has been significantly associated with inflammatory gingival responses and have been detected with high abundance on the surfaces of some oral cancer tissue [60] [61]. They are also known to contribute to the acidic and hypoxic environments of oral cancer, promoting their growth [62, 63]. *Gemella morbillorum* has also been the cause of infective endocarditis amongst a smokeless tobacco user from Somalia [64]. The ability of *Staphylococcus caprae* to produce B haemolysins and enterotoxins can lead to significant cell damage amongst oral cells [65]. On the other hand, species distinguishing the buccal mucosa of non-

users of Toombak included *Rothia mucilaginosa* [98.88% per ident], *Veillonella atypica* [98.28% per ident], *Streptococcus sobrinus* [99.35% per ident.], *Prevotella salivae* [99.35% per ident], *Prevotella pallens* [99.35% per ident] and *Bifidobacterium dentium* [99.33% per ident].

The dorsum tongue mucosa

Actinomyces [p= 0.013] and uncultured bacterium [p=0.031] were enriched in the tongue of Toombak users. *Tannerella* [q=0.0085] and *Cardiobacterium* [q= 0.0097] in Toombak users, were more abundant in this mucosal location with a mean change of 5 and 13 respectively compared to non-users. A *Fusobacteria* unidentified species *[OTU_1157]* was also associated with Toombak use [LDA=2.69]. In non-users, *Bifidobacterium* [q=0.0049] was significantly increased by a mean change of 1.15 while five unique genera in non-users of smokeless tobacco were identified; *Olsenella* [p=0.04], *Lactobacillus, F0058, Aggregatibacter* [p=0.02], and *Actinobacillus* [p=0.01]. Pattern search of the top 25 genera revealed that *Bergeyella* [p=0.02], *uncultured bacteria* and *Leptotrichia [p=0.007]* were the most positively correlated genera associated with Toombak use, while *Scardovia [p=0.002]* and *Lactobacillus* were positively correlated with non-users. LEfSe plotting further highlighted that *Haemophilus* [p=0.004, LDA=3.7], *Actinobacillus* [0.003, LDA = 3.8], and *Scardovia* were discriminant genera for the dorsum tongue of non-users [Figure 7b] while *Actinomyces* distinguished the tongue of Toombak users [LDA = 3.9].

The species *Leptotrichia wadei* [q= 0.023, 99.55% per ident], *Schaalia odontolytica* [p= 0.0028] and *Streptococcus equinus* [p= 0.043, 99.78% per ident] were abundant in the microbiome of the dorsum tongue of Toombak users while *Haemophilus haemolyticus* [q=0.001, 99.78% per ident] and *Streptococcus sobrinus* [q= 0.023, 99.35% per ident] were enriched amongst non-users of Toombak. *Corynebacterium argentoratense* [LDA=3.7] and *Actinobacillus pleuropneumoniae* [LDA=2.64] were also associated with the tongue of non-users.

The floor of the mouth mucosa

Significant enrichment of the genera *Corynebacterium_1* [p=0.0028], *Staphylococcus* [q=0.01] and *Candidatus_Saccharimonas* [p=0.026] was found in the floor of the mouth in those utilising Toombak while *Scardovia* [q=0.04] was high in non-users. LEfSe plotting distinguished uncultured bacterium [LDA =3.1] *and Porphyromonas*

gingivalis [LDA =3.09] as the most distinctive microbiome components in the floor of the mouth of Toombak users [Figure 7c]. Five unique genera were present in Toombak users: *Selenomonas_4, Filifactor, Eubacterium_nodatum group, Candidatus_Saccharimonas*, *Anaeroglobus* while one genus was unique to non-users*; Catonella*. Pattern search further revealed that *Corynebacterium_1, Candidatus Saccharimonas, Stomatobaculum, Parvimonas*, and *Oribacterium* were positively correlated with Toombak use while *Prevotella_2, Lactobacillus, Scardovia, Actinobacillus, Comamonas, and Aggregatibacter* were the positively correlated genera with non-users of Toombak.

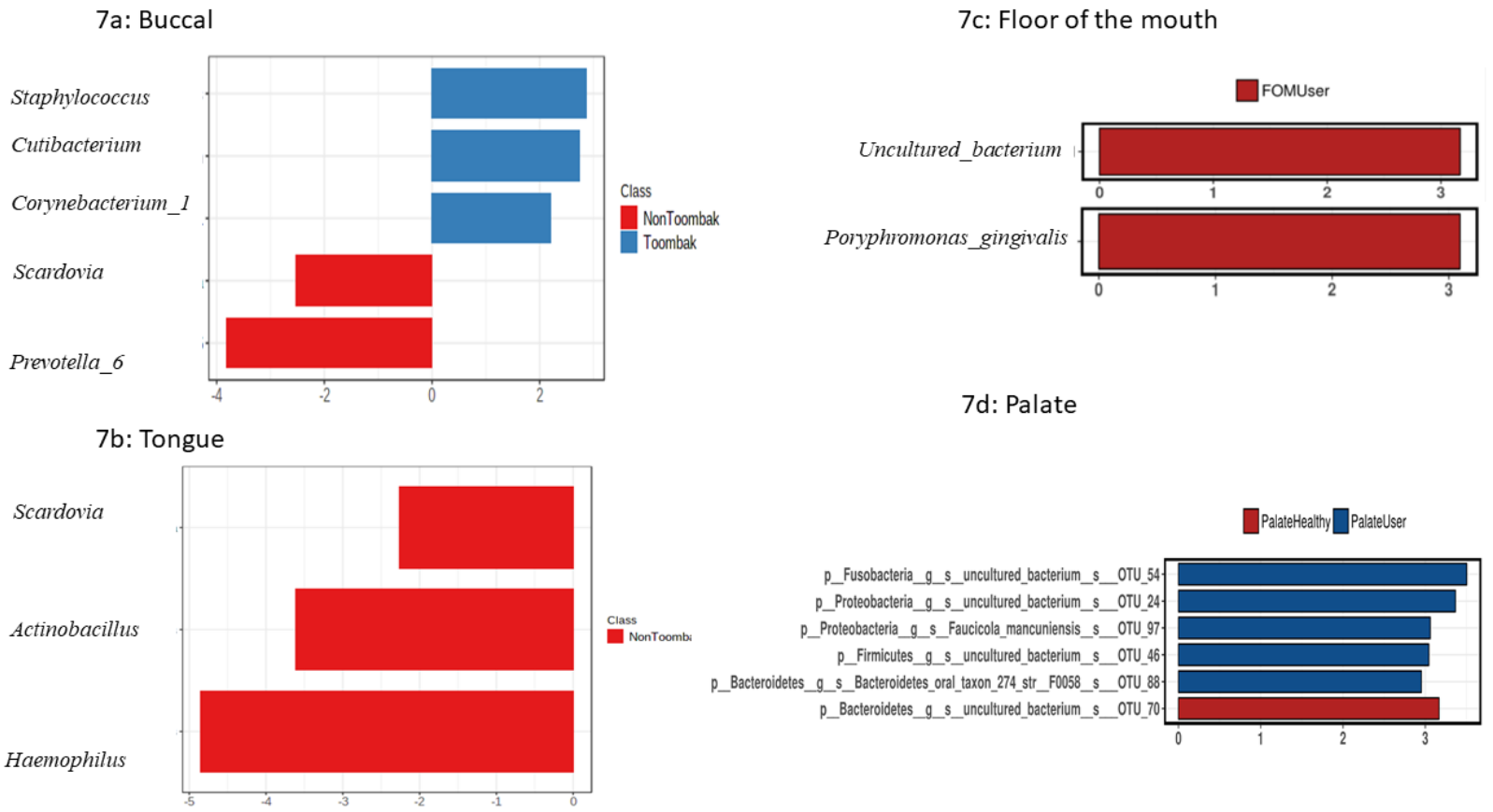
Species inhabiting the floor of the mouth of Toombak users included *Stomatobaculum longum, Eubacterium infirmum, Prevotella denticola,* and *Bifidobacterium dentium,* while *Corynebacterium argentoratense* was evident in the floor of the mouth of non-users.

The palatal mucosa

In the palatal mucosa, positive microbiome correlations with Toombak use included *Candidatus Saccharimonus* [p=0.086], *Lautropia* [q=0.002, LDA=3.3], *Johnsenella* [q=0.001], *Actinomyces*, *Capnocytophaga*, *Leptotrichia*, *Faucicola* [q=0.001], and *Peptococcus* while non-users of Toombak had positive correlations with *Prevotella* and *Prevotella_6, Selenomonas, Anaeroglobus, Lactobacillus, Moraxella, Shigella, Scardovia* [q=0.012] and *Simonsiella* [q=0.04] which were all evident in non-users. An unclassified *Gracilibacteria* oral taxon 873 [p=0.031] was found to be unique to the palatal microbiome of Toombak users. Several OTUs from the *Patescibacteria* phylum were significantly abundant amongst Toombak users and included OTU_112 [p=0.032], OTU_144 [p=0.023], OTU_24 [p=0.039], OTU_113 [p=0.031] and OTU_114 [p=0.043].

The species, *Leptotrichia hofstadii* was highly correlated with the palatal mucosa of Toombak users [99.32% ident] and *Streptococcus salivarius* was abundant. *Actinomyces israelii, Actinomyces cardiffensis and Faucicola mancuniensis,* were detected only in the palate of Toombak users. LEfSe plotting of species discriminant of Toombak users highlighted *Faucicola mancuniensis,* a *Bacteroidetes oral taxon 274 strain F008* and uncultured bacterium from the phyla *Fusobacteria* and *Proteobacteria* [Figure 7d].

Figure 7a – 7d:


7a: Buccal
Staphylococcus
Cutibacterium
Corynebacterium_1
Scardovia
Prevotella_6
-4
-2
0
2
Class
NonToombak
Toombak
7b: Tongue
Scardovia
Actinobacillus
Haemophilus
-5
-4
-3
-2
-1
0
Class
NonToombi
7c: Floor of the mouth
FOMUser
Uncultured_bacterium
Poryphromonas_gingivalis
0
1
2
3
7d: Palate
PalateHealthy
PalateUser
p__Fusobacteria__g__s__uncultured_bacterium__s__OTU_54
p__Proteobacteria__g__s__uncultured_bacterium__s__OTU_24
p__Proteobacteria__g__s__Faucicola_mancuniensis__s__OTU_97
p__Firmicutes__g__s__uncultured_bacterium__s__OTU_46
p__Bacteroidetes__g__s__Bacteroidetes_oral_taxon_274_str__F0058__s__OTU_88
p__Bacteroidetes__g__s__uncultured_bacterium__s__OTU_70
0
1
2
3

7: LEfSe plotting of the four oral mucosal locations.7a: Buccal microbiome. Staphylococcus, Corynebacterium_1 and Cutibacterium were the most discriminant genera in Toombak users while Prevotella and Scardovia were found in non-users. 7b: Tongue microbiome. Haemophilus, Actinobacillus, and Scardovia were discriminant in the dorsum tongue of non-users of Toombak. 7c: The floor of the mouth. uncultured bacterium [LDA =3.1] and Porphyromonas gingivalis [LDA =3.09] are the most distinct microbiome components in the floor of the mouth of Toombak users. 7d: Palate. Uncultured bacterium from the phyla Fusobacteria and Proteobacteria, and the species, Faucicola mancuniensis are seen in Toombak users while uncultured bacterium from the phyla Bacteroidetes are discriminant of non-users.

In comparison, *Bifidobacterium longum* [99.33% ident], *Veillonella atypica* [98.28% ident] and *Alloscardovia omnicolens* [99.33% ident] were positively correlated with the palatal mucosal microbiome of non-users while *Haemophilus paraurethrae, Haemophilus quentini,* and *Prevotella denticola* were found in non-users and *Actinomyces graevenitzii* in both groups. *Actinomyces cardiffensis* and *Actinomyces israelii* were only found on the palatal microbiome of Toombak users. Deseq 2 highlighted four OTUs abundant amongst Toombak users with q values < 0.05 that included OTU 113 [weak similarity with *Terrisporobacter*], OTU 427 [high similarity with *Neisseria* species], *Lautropia mirabilis* [BLASTn 99.35% ident] and *Leptotrichia*. *Poryphromonas cataniae* [98% ident] were significantly abundant in Toombak users.

Actinomyces abundance in the saliva, tongue, buccal cheek, and palate of Toombak users

We found an abundance of *Actinomyces* in the saliva [q=0.0045], tongue [p=0.013], and palate [p=0.001] of Toombak users. Reports on the effects of smokeless tobacco on *Actinomyces* growth has been contradictory. Smokeless tobacco use has been found to both reduce [66, 67] and enrich *Actinomyces* growth [68]. *Actinomyces* has also been found in abundance in moist Indian smokeless tobacco products [69]. *Actinomyces massiliensis* [p=0.016] was abundant in the buccal microbiome of Toombak users while *Actinomyces graevenitzii* was found to be enriched in the tongue in both users and non-users of Toombak. *Actinomyces* is associated with the discolouration of teeth [70] while *Actinomyces graevenitzii* has been associated with halitosis [71]. *Actinomyces graevenitzii* and *Staphylococcus* in co-culture were also found to significantly reduce neutrophil recruitment a factor that could participate in immune dysregulation amongst Toombak users [72]. Furthermore, many *Actinomyces* species are potent nitrate-reducing bacteria and thus may contribute to increasing tobacco-specific nitrosamine production in Toombak [73]. In other studies, *Actinomyces meyeri* was found to be abundant amongst smokeless tobacco users from Saudi Arabia [34]. *Actinomyces israelii* and *Actinomyces cardiffensis* were found to be unique to the palatal microbiome of Toombak users.

Lautropia abundance in the saliva, buccal cheek and palate microbiomes of Toombak users could relate to compromised local immunity

In this study, *Lautropia* was abundant in the saliva, buccal cheek, and palate microbiomes of Toombak users with *Lautropia mirabilis* the species identified. *Lautropia* has been associated with benign lesions in the oral cavity such as fibroepithelial polyp [74] as well as malignant diseases including non-small cell lung cancer, oral and oesophageal cancers [75, 76]. In a study assessing Srilankan betel quid users, *Lautropia* was not abundant amongst smokeless tobacco users but rather in healthy controls [77]. *Lautropia* has been detected as part of the core oral microbiome amongst healthy Nigerians [78], however, *Lautropia mirabilis* abundance may also be associated with compromised local immunity that could be caused by Toombak use. The *Lautropia* genus was also found to be abundant in the oral microbiome of a series of HIV-infected children [79].

Prevotella abundance amongst non-users of Toombak

Although *Prevotella* was present in both users and non-users, they were significantly more abundant in Toombak non-users [$p < 0.05$] and were found in the saliva, tongue, buccal cheek and palatal microbiomes. *Alloprevotella* amongst non-users was also higher compared to Toombak users. *Prevotella* is often a commensal oral microorganism and in the oral cavity has been found to be more abundant amongst those with African and Indian heritage [80-82]. *Prevotella* abundance in the oral cavity has been associated with the development of other diseases that includes rheumatoid arthritis, metabolic syndrome, inflammatory bowel and cardiovascular disease but in other studies its abundance is associated with the reduction of hypersensitivity reactions [83, 84].

Species modifications may harbour distinct susceptibilities in the oral cavity of Toombak users

The most common bacteria in the oral cavity were found to vary in species due to smokeless tobacco use. This may allow for a complex integration of bacteria with a more sinister ability to transfer and integrate features such as antibiotic-resistance genes eliciting distinct susceptibilities in the oral microbiome of Toombak users. Common to both users and non-users of Toombak were the species *Prevotella salivae* and *Prevotella pallens* which were found to be abundant in the tongue and palate of

both groups and the buccal microbiome of non-users. They are common to the oral microbiome of adults [85]. *Prevotella scopos* and *Prevotella veroralis were* also evident in the tongue and palate microbiomes in both groups. In one study *Prevotella salivae* was increased in the buccal mucosa of smokers [86]. Although *Prevotella pallens* is commonly associated with oral health [87], in some studies, it has been associated with halitosis [oral malodour] as well as OSCC [88, 89].

Prevotella nigrescens however was abundant only in Toombak users. Intra-orally, this species has been associated with inflammatory changes to the mucosa [90] and alveolar bone loss [83], while extra-orally, it has been detected in samples of occluded arteries [vascular disease] [91]. *Prevotella nigrescens* abundance did not differ in a study of betel nut users and non-users in Thailand [92]. *Prevotella nigrescens* and *Prevotella pallens* have also been associated with penicillin resistance [93].

In the tongue, *Streptococcus sobrinus* was abundant in non-users, while *Streptococcus equinus* was abundant in Toombak users. *Streptococcus salivarius* was abundant in the palate of Toombak users [90]. Species from the *Streptococcus bovis/Streptococcus equinus* complex are some of the most antibiotic-resistant *Streptococcus* species found in the oral cavity [94]. It is interesting that on the dorsum tongue in Toombak users, *Streptococcus sobrinus* [a caries-associated bacterium but with no known antibiotic resistance] is replaced with *Streptococcus equinus*. *Streptococcus salivarius* found in abundance in the palate of Toombak users has been shown to elicit the release of the pro-inflammatory interleukin 8 [95] and is known to be potent acetaldehyde producers [34]. Non-sucrose-containing smokeless tobacco forms, such as Toombak have also been shown to be a possible source of growth substrate for various *Streptococcus* species including *Streptococcus salivarius* [96]. Therefore, it is probable that the use of smokeless tobacco Toombak contributes to a modified *Streptococcus* inhabitance in the oral cavity that leads to specific alterations in the various oral mucosal locations. *Streptococcus* species associated with antibiotic resistance increase in those who use Toombak compared to non-users which could be explained by the increased microbiome entry from the Toombak itself into the oral cavity as well as alterations from mechanical, chemical, and pH changes that arise from Toombak use. *Streptococcus mitis* was the most abundant species in all sites of the oral cavity, while *Streptococcus porcorum* was abundant in the palate and tongue in both users and non-users. *Streptococcus mitis* species are regarded as commensals in the oral cavity but

some strains harbour unusually high levels of resistance to β lactam antibiotics [97]. In one study, *Streptococcus mitis* was found to be abundant in those with OSCC [98] while in another study, *Streptococcus mitis* was shown to have therapeutic potential in the treatment of oral cancer [99].

The family, *Veillonellaceae* were found to be in similar abundance in both groups. The genus *Veillonella* is generally associated with good oral health [100] and were abundant in the tongue of both users and non-users. *Veillonella rogosae* was found on all mucosal surfaces. *Veillonella atypica* was abundant in the palate and tongue of Toombak users and non-users and the buccal cheek of non-users. In Toombak users, *Veillonella atypica* may have a role to play in nitrate/nitrite reduction from smokeless tobacco into the carcinogenic tobacco-specific nitrosamines [101].

Non-users of smokeless tobacco harbour genera with pro-mucosal integrity potential that include *Scardovia, Bifidobacterium* and *Lactobacillus*

Associated with healthy mucosal integrity are the genera *Scardovia,* [102] and *Lactobacillus* [56] while *Bifidobacterium* help in the maintenance of epithelial integrity and promotion of healthy gingival barriers [103]. Interestingly, *Scardovia* is found to be depleted in oral cancer [104] and in this study was enriched in the supragingival plaque of non-users and could distinguish the buccal mucosal microbiome of the same group. Therefore, the absence of this genus amongst Toombak users may highlight a lack of mucosal protection against oral cancer. *Lactobacillus* was enriched in the plaque of non-users of Toombak and was part of the core genera of the tongue microbiome in non-users. *Lactobacillus* was also abundant in the paraffin-embedded tissue OSCC samples taken from non-users of Toombak. *Bifidobacterium* was enriched in the plaque of non-users, *Bifidobacterium dentium* in the tongue and buccal microbiome and *Bifidobacterium longum* on the palate of non-users.

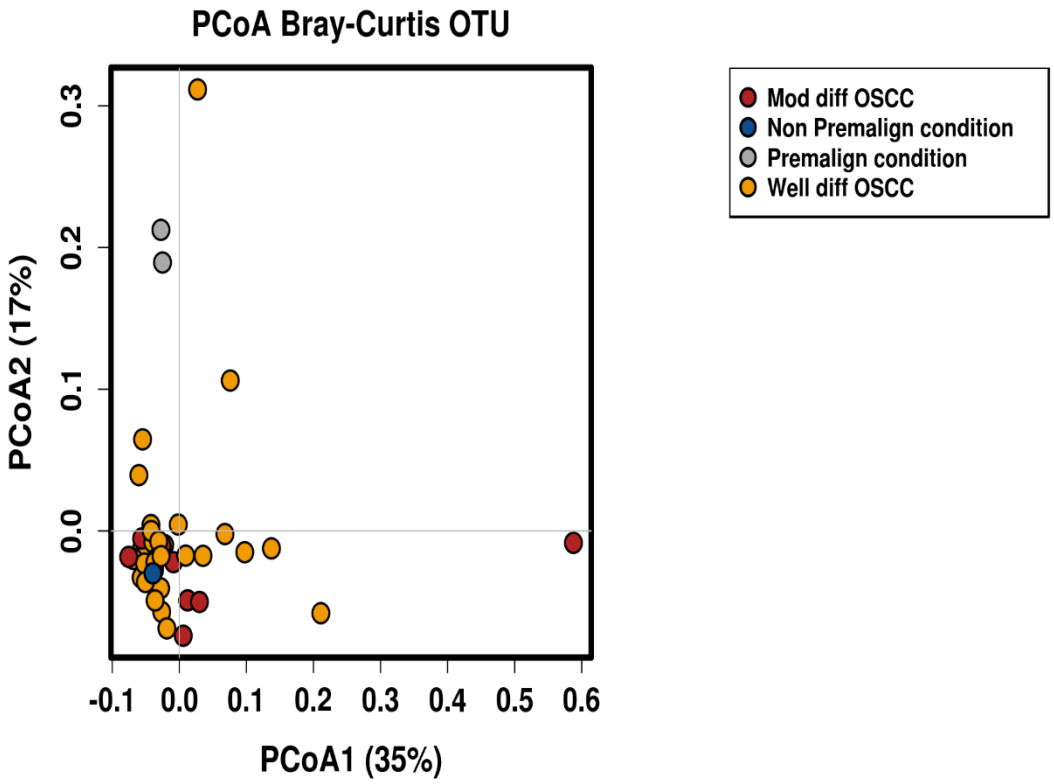
Lactobacillus and *Bifidobacterium* could have many beneficial activities in the oral cavity that may be lost with Toombak use. They can produce antimicrobial compounds and have a key role in supporting the oral immune defence mechanisms. IgA levels in saliva have been shown to increase in those with *Lactobacillus* and *Bifidobacterium* abundances [105]. *Bifidobacterium dentium* is ‘mucin’ friendly preventing the disruption of the mucous barrier [106]. *Bifidobacterium* is reported to reduce DNA

damage and may delay or even prevent the onset of cancers [107]. Both *Bifidobacterium* and *Lactobacillus* can help prevent the colonisation of pathogenic bacteria in the oral cavity [108]. We found *Prevotella nigrescens* for example to be abundant in the Toombak group but not in non-users. *Lactobacillus* has been shown to have eliminating action against *Prevotella nigrescens* and thus its reduced presence amongst Toombak users could explain why this species was more abundant in Sudanese smokeless tobacco users [100].

Toombak-associated OSCC significantly carries a more aggressive microbiome

Clustering of the premalignant samples together could be appreciated [Figure 8a] but Beta diversity was non-significant between groups; moderate [n=8], well-differentiated [n=35], premalignant samples [n=2] and the control [n=1].

Figure 8a:

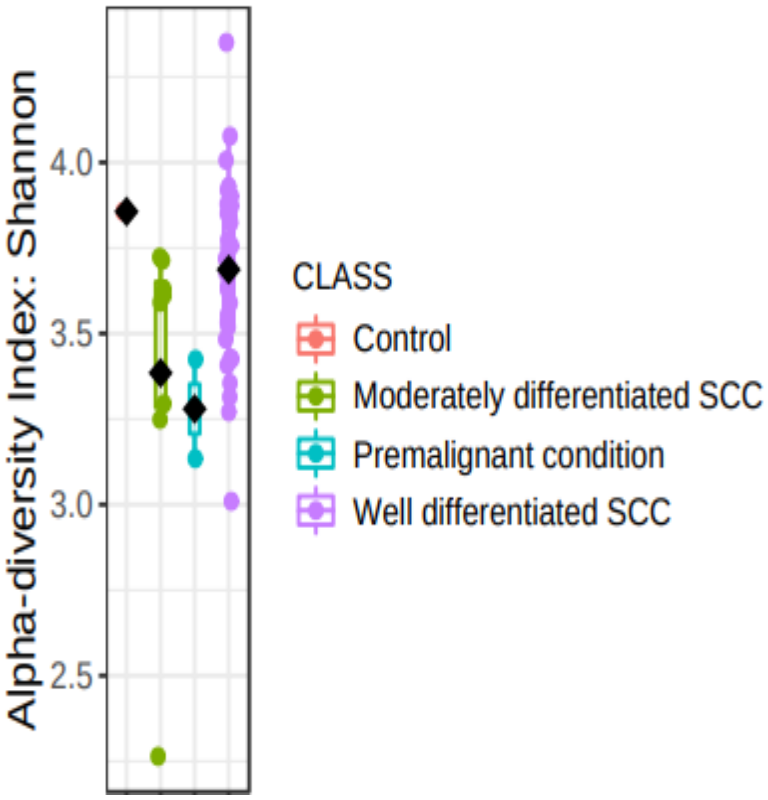


8a: Premalignant and oral squamous cell carcinoma metagenomic profiles. B diversity of oral squamous cell carcinoma cancer samples from Sudanese participants utilising paraffin embedded blocks. Differentiation between premalignant and malignant samples may be

observed. Malignant samples further classified into moderate [n=8] and well-differentiated OSCC [n=35].

It is likely, that the microbiome of moderate and well-differentiated OSCC samples is similar in microbial habitat. Shannon index between groups was statistically significant [p=0.03], where the control or non-cancerous sample had the highest median alpha diversity compared to the remaining groups of oral cancer and premalignant conditions [Figure 8b].

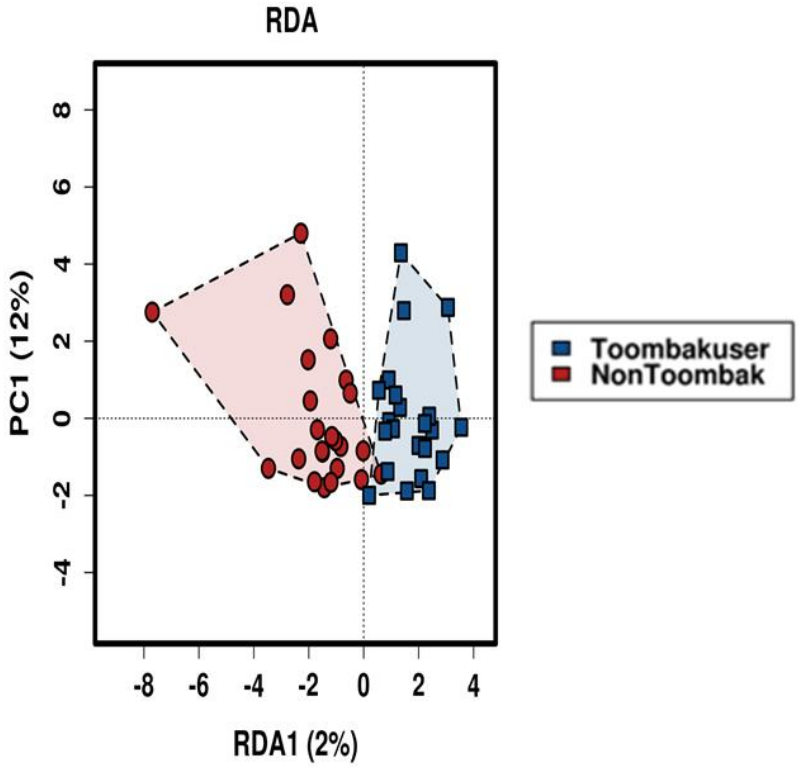
Figure 8b:



8b: Premalignant and oral squamous cell carcinoma metagenomic profiles. Shannon diversity [p=0.03]. Non premalignant samples had the highest Shannon diversity, and the premalignant samples had the lowest alpha diversity.

Redundancy analysis indicated that the oral cancer microbiome composition was influenced by Toombak use [Figure 8c].

Figure 8c:



8c: Premalignant and oral squamous cell carcinoma metagenomic profiles. redundancy analysis of oral cancer samples. Toombak use highlights microbiome variation between cancer samples compared to non-users. Oral cancer samples from non-Toombak users had increased abundances of Lactobacillus [p=0.024], Collinsella [p=0.014], and Flavobacterium [p=0.0087] while oral cancers from Toombak users had abundances in Corynebacterium_1 [p=0.044], Mogibacterium [p=0.0013], Enterococcus [p=0.012], Empedobacter [p=0.03], Stenotrophomonas [p=0.043] and Schlegelella [p=0.048].

We further applied DeSeq2 to compare mean and subsequent fold changes of the most important genera amongst Toombak and non-users of Toombak OSCC samples. OSCC from non-Toombak users had increased abundances of *Lactobacillus* [p=0.024], *Collinsella* [p=0.014], and *Flavobacterium* [p=0.0087] while OSCC samples from Toombak users had abundances in *Corynebacterium_1* [p=0.044], *Mogibacterium* [p=0.0013], *Enterococcus* [p=0.012], *Empedobacter* [p=0.03], *Stenotrophomonas* [p=0.043] and *Schlegelella* [p=0.048]. *Corynebacterium_1* was

significantly increased amongst those cancer samples from Toombak users [p=0.044] with a reduced fold change of -1.446 amongst non–users' cancer samples. Interestingly, *Corynebacterium_1* has been found to be the most abundant genus in Toombak [1] and in this study, was found to be significantly increased in the saliva, buccal and floor of the mouth microbiomes of Toombak users. *Coprococcus_2* [p=0.05], *Rothia* [p= 1.5e-10], *Eubacterium_nodatum* [p=0.0015], *Peptostreptococcus* [p=6.5e-07], *Ruminiclostridium* [p=0.0042] and *Stomatobaculum* [p=0.013] were the most abundant genera amongst premalignant samples compared to oral cancer samples.

Conclusion

In this study we showed that the microbiome and mycobiome of the oral cavity is significantly altered with Toombak smokeless tobacco use. In Toombak users, non-normally residing genera become abundant throughout the soft tissue mucosal locations users that include *Staphylococcus* while in non-users of Toombak, *Prevotella, Lactobacillus* and *Bifidobacterium* are found to be more prevalent. Toombak use was also found to significantly alter species determinants in genera such as *Prevotella, Streptococcus*, *Actinomyces* and *Veillonella.*

The mycobiome is further seen to lose counterbalance with Toombak use, with a significant enhancement of *Aspergillus* and loss of *Candida*. Such findings can pave the way for the development of 'oral fungal and bacterial response panels' to track OSCC progression throughout Toombak use. Further whole genomic sequencing approaches can answer how such modified communities challenge the specific functionality of the microorganism in the response to Toombak use.

While no significant local behaviours were known to otherwise alter the oral microbiome of the participants included in this study, factors such as dietary intake could play an indirect role in the host-oral microbiome connections reflected in this study requiring further research to outline other impingements on the oral microbiome of the Sudanese.

The results of this study further contribute to a new understanding in OSCC development and its progression amongst smokeless tobacco users worldwide. We have further shown that the microbiome of Toombak related OSCC may be more aggressive resulting in an increased risk of recurrence and metastasis.

4.5 References

1. Sami A, Elimairi I, Patangia D, Watkins C, Ryan CA, Ross RP, Stanton C. 2021. The ultra-structural, metabolomic and metagenomic characterisation of the sudanese smokeless tobacco 'Toombak'. Toxicology Reports 8:1498-1512.
2. Idris AM, Nair J, Ohshima H, Friesen M, Brouet I, Faustman EM, Bartsch H. 1991. Unusually high levels of carcinogenic tobacco-specific nitrosamines in Sudan snuff [toombak]. Carcinogenesis 12:1115-8.
3. Torti SV, Manz DH, Paul BT, Blanchette-Farra N, Torti FM. 2018. Iron and Cancer. Annu Rev Nutr 38:97-125.
4. Arthur RA, Dos Santos Bezerra R, Ximenez JPB, Merlin BL, de Andrade Morraye R, Neto JV, Fava NMN, Figueiredo DLA, de Biagi CAO, Jr., Montibeller MJ, Guimarães JB, Alves EG, Schreiner M, da Costa TS, da Silva CFL, Malheiros JM, da Silva LHB, Ribas GT, Achallma DO, Braga CM, Andrade KFA, do Carmo Alves Martins V, Dos Santos GVN, Granatto CF, Terin UC, Sanches IH, Ramos DE, Garay-Malpartida HM, de Souza GMP, Slavov SN, Silva WA, Jr. 2021. Microbiome and oral squamous cell carcinoma: a possible interplay on iron metabolism and its impact on tumor microenvironment. Braz J Microbiol 52:1287-1302.
5. Ahmed HG. 2013. Aetiology of oral cancer in the Sudan. J Oral Maxillofac Res 4:e3.
6. Sajid M, Srivastava S, Joshi L, Bharadwaj M. 2021. Impact of smokeless tobacco-associated bacteriome in oral carcinogenesis. Anaerobe 70:102400.
7. Buduneli N. 2021. Environmental factors and periodontal microbiome. Periodontology 2000 85:112-125.
8. Willis D, Popovech M, Gany F, Zelikoff J. 2012. Toxicology of Smokeless Tobacco: Implications for Immune, Reproductive, and Cardiovascular Systems. Journal of Toxicology and Environmental Health, Part B 15:317-331.
9. Alsanosy RM. 2014. Smokeless tobacco [shammah] in Saudi Arabia: a review of its pattern of use, prevalence, and potential role in oral cancer. Asian Pacific Journal of Cancer Prevention 15:6477-6483.

10. Ibrahim SO, Vasstrand EN, Johannessen AC, Idris AM, Magnusson B, Nilsen R, Lillehaug JR. 1999. Mutations of the p53 gene in oral squamous-cell carcinomas from sudanese dippers of nitrosamine-rich toombak and non-snuff-dippers from the Sudan and Scandinavia. International Journal of Cancer 81:527-534.
11. IBRAHIM SO, BERTELSEN B, KALVENES MB, IDRIS AM, VASSTRAND EN, NILSEN R, JOHANNESSEN AC. 1998. Expression of keratin 13, 14 and 19 in oral squamous cell carcinomas from Sudanese snuff dippers: Lack of association with human papillomavirus infection. APMIS 106:959-969.
12. Saxena R, Prasoodanan P K V, Gupta SV, Gupta S, Waiker P, Samaiya A, Sharma AK, Sharma VK. 2022. Assessing the Effect of Smokeless Tobacco Consumption on Oral Microbiome in Healthy and Oral Cancer Patients. Frontiers in Cellular and Infection Microbiology 12.
13. Mohamed N, Litlekalsøy J, Ahmed IA, Martinsen EMH, Furriol J, Javier-Lopez R, Elsheikh M, Gaafar NM, Morgado L, Mundra S, Johannessen AC, Osman TA, Nginamau ES, Suleiman A, Costea DE. 2021. Analysis of Salivary Mycobiome in a Cohort of Oral Squamous Cell Carcinoma Patients From Sudan Identifies Higher Salivary Carriage of Malassezia as an Independent and Favorable Predictor of Overall Survival. Front Cell Infect Microbiol 11:673465.
14. Walsh AM, Crispie F, Kilcawley K, O'Sullivan O, O'Sullivan MG, Claesson MJ, Cotter PD. 2016. Microbial Succession and Flavor Production in the Fermented Dairy Beverage Kefir. mSystems 1:e00052-16.
15. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957-2963.
16. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-1.
17. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26:266-7.
18. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project:

improved data processing and web-based tools. Nucleic Acids Res 41:D590-6.

19. McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20-5.
20. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13:581-583.
21. Team RC. 2014. R: A language and environment for statistical computing. MSOR connections 1.
22. Martin M. 2011. CUTADAPT removes adapter sequences from high-throughput sequencing reads. EMBnetjournal 17.
23. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. 2017. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Research 45:W180 - W188.
24. Kovács L, Kézér FL, Bakony M, Hufnágel L, Tőzsér J, Jurkovich V. 2015. Associations between Heart Rate Variability Parameters and Housing- and Individual-Related Variables in Dairy Cows Using Canonical Correspondence Analysis. PloS one 10:e0145313-e0145313.
25. Harrandah AM, Chukkapalli SS, Bhattacharyya I, Progulske-Fox A, Chan EKL. 2020. Fusobacteria modulate oral carcinogenesis and promote cancer progression. Journal of oral microbiology 13:1849493-1849493.
26. Verma D, Garg PK, Dubey AK. 2018. Insights into the human oral microbiome. Archives of Microbiology 200:525-540.
27. Fujiwara N, Kitamura N, Yoshida K, Yamamoto T, Ozaki K, Kudo Y. 2020. Involvement of Fusobacterium Species in Oral Cancer Progression: A Literature Review Including Other Types of Cancer. International journal of molecular sciences 21:6207.
28. Espinoza JL, Harkins DM, Torralba M, Gomez A, Highlander SK, Jones MB, Leong P, Saffery R, Bockmann M, Kuelbs C, Inman JM, Hughes T, Craig JM, Nelson KE, Dupont CL, Relman DA. 2018. Supragingival Plaque Microbiome Ecology and Functional Potential in the Context of Health and Disease. mBio 9:e01631-18.

29. Sami A, Elimairi I, Patangia D, Watkins C, Ryan CA, Ross RP, Stanton C. 2021. The ultra-structural, metabolomic and metagenomic characterisation of the sudanese smokeless tobacco 'Toombak'. Toxicol Rep 8:1498-1512.
30. Rands CM, Starikova EV, Brüssow H, Kriventseva EV, Govorun VM, Zdobnov EM. 2018. ACI-1 beta-lactamase is widespread across human gut microbiomes in Negativicutes due to transposons harboured by tailed prophages. Environmental Microbiology 20:2288-2300.
31. Gopinath D, Kunnath Menon R, Chun Wie C, Banerjee M, Panda S, Mandal D, Behera PK, Roychoudhury S, Kheur S, George Botelho M, Johnson NW. 2021. Salivary bacterial shifts in oral leukoplakia resemble the dysbiotic oral cancer bacteriome. Journal of Oral Microbiology 13:1857998.
32. Waite DW, Chuvochina M, Pelikan C, Parks DH, Yilmaz P, Wagner M, Loy A, Naganuma T, Nakai R, Whitman WB, Hahn MW, Kuever J, Hugenholtz P. 2020. Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. Int J Syst Evol Microbiol 70:5972-6016.
33. Campbell A, Campbell J, Schwientek P, Woyke T, Sczyrba A, Allman S, Beall C, Griffen A, Leys E, Podar M. 2013. Multiple Single-Cell Genomes Provide Insight into Functions of Uncultured Deltaproteobacteria in the Human Oral Cavity. PloS one 8:e59361.
34. Halboub E, Al-Ak'hali MS, Alamir AH, Homeida HE, Baraniya D, Chen T, Al-Hebshi NN. 2020. Tongue microbiome of smokeless tobacco users. BMC Microbiology 20:201.
35. Hernandez BY, Zhu X, Goodman MT, Gatewood R, Mendiola P, Quinata K, Paulino YC. 2017. Betel nut chewing, oral premalignant lesions, and the oral microbiome. PLOS ONE 12:e0172196.
36. Chattopadhyay I, Verma M, Panda M. 2019. Role of Oral Microbiome Signatures in Diagnosis and Prognosis of Oral Cancer. Technology in cancer research & treatment 18:1533033819867354-1533033819867354.
37. Pushalkar S, Ji X, Li Y, Estilo C, Yegnanarayana R, Singh B, Li X, Saxena D. 2012. Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. BMC Microbiology 12:144.

38. Ali MABM. 2014. Isolation and Identification of some Oral Microorganisms from healthy Sudanese Smokers and Oral Cancer PatientsUniversity of Gezira.
39. Okada N, Murakami A, Sato M, Nakamura S, Fujii S, Sogabe K, Takahashi M, Okada A, Abe A, Fujii H, Abe M, Azuma M, Ishizawa K. 2022. First reported case of Lachnoanaerobaculum gingivalis bacteremia in an acute myeloid leukemia patient with oral mucositis during high dose chemotherapy. Anaerobe 76:102610.
40. Sawant S, Dugad J, Parikh D, Srinivasan S, Singh H. 2021. Identification & correlation of bacterial diversity in oral cancer and long-term tobacco chewers- A case-control pilot study. Journal of Medical Microbiology 70.
41. Sarkar A, Stoneking M, Nandineni MR. 2017. Unraveling the human salivary microbiome diversity in Indian populations. PLOS ONE 12:e0184515.
42. Al-Marzooq F, Al Kawas S, Rahman B, Shearston JA, Saad H, Benzina D, Weitzman M. 2022. Supragingival microbiome alternations as a consequence of smoking different tobacco types and its relation to dental caries. Scientific reports 12:1-14.
43. Amer A, Galvin S, Healy CM, Moran GP. 2017. The Microbiome of Potentially Malignant Oral Leukoplakia Exhibits Enrichment for Fusobacterium, Leptotrichia, Campylobacter, and Rothia Species. Frontiers in Microbiology 8.
44. Torres PJ, Fletcher EM, Gibbons SM, Bouvet M, Doran KS, Kelley ST. 2015. Characterization of the salivary microbiome in patients with pancreatic cancer. PeerJ 3:e1373.
45. Mougeot JC, Beckman MF, Langdon HC, Lalla RV, Brennan MT, Bahrani Mougeot FK. 2021. Haemophilus pittmaniae and Leptotrichia spp. Constitute a Multi-Marker Signature in a Cohort of Human Papillomavirus-Positive Head and Neck Cancer Patients. Front Microbiol 12:794546.
46. Al Bataineh MT, Dash NR, Elkhazendar M, Alnusairat DaMH, Darwish IMI, Al-Hajjaj MS, Hamid Q. 2020. Revealing oral microbiota composition and functionality associated with heavy cigarette smoking. Journal of Translational Medicine 18:421.
47. Srivastava A, Mishra S, Garg PK, Dubey AK, Deo SVS, Verma D. 2022. Comparative and analytical characterization of the oral bacteriome of

smokeless tobacco users with oral squamous cell carcinoma. Applied Microbiology and Biotechnology 106:4115-4128.

48. Proctor DM, Fukuyama JA, Loomer PM, Armitage GC, Lee SA, Davis NM, Ryder MI, Holmes SP, Relman DA. 2018. A spatial gradient of bacterial diversity in the human oral cavity shaped by salivary flow. Nature Communications 9:681.
49. Magnussen A, Parsi MA. 2013. Aflatoxins, hepatocellular carcinoma and public health. World J Gastroenterol 19:1508-12.
50. Li Y, Wang K, Zhang B, Tu Q, Yao Y, Cui B, Ren B, He J, Shen X, Van Nostrand JD, Zhou J, Shi W, Xiao L, Lu C, Zhou X. 2019. Salivary mycobiome dysbiosis and its potential impact on bacteriome shifts and host immunity in oral lichen planus. Int J Oral Sci 11:13.
51. Pavia CS, Plummer MM. 2020. Clinical implications of nicotine as an antimicrobial agent and immune modulator. Biomedicine & Pharmacotherapy 129:110404.
52. da Costa KR, Ferreira JC, Komesu MC, Candido RC. 2009. Candida albicans and Candida tropicalis in oral candidosis: quantitative analysis, exoenzyme activity, and antifungal drug sensitivity. Mycopathologia 167:73-9.
53. Deepa A, Nair BJ, Sivakumar T, Joseph AP. 2014. Uncommon opportunistic fungal infections of oral cavity: A review. J Oral Maxillofac Pathol 18:235-43.
54. Sankari SL, Mahalakshmi K, Kumar VN. 2020. A comparative study of Candida species diversity among patients with oral squamous cell carcinoma and oral potentially malignant disorders. BMC Res Notes 13:488.
55. Srivastava A, Mishra S, Garg PK, Dubey AK, Deo S, Verma D. 2022. Comparative and analytical characterization of the oral bacteriome of smokeless tobacco users with oral squamous cell carcinoma. Applied Microbiology and Biotechnology:1-14.
56. Laheij AMGA, Raber-Durlacher JE, Koppelmans RGA, Huysmans M-CDNJM, Potting C, van Leeuwen SJM, Hazenberg MD, Brennan MT, von Bültzingslöwen I, Johansson J-E, de Soet JJ, Haverman TM, Buijs MJ, Brandt BW, Rozema FR, Blijlevens NMA, Zaura E. 2019. Microbial changes in relation to oral mucositis in autologous hematopoietic stem cell transplantation recipients. Scientific Reports 9:16929.

57. Carolus H, Van Dyck K, Van Dijck P. 2019. Candida albicans and Staphylococcus Species: A Threatening Twosome. Frontiers in Microbiology 10.
58. McCormack MG, Smith AJ, Akram AN, Jackson M, Robertson D, Edwards G. 2015. Staphylococcus aureus and the oral cavity: an overlooked source of carriage and infection? Am J Infect Control 43:35-7.
59. Pasman R, Krom BP, Zaat SAJ, Brul S. 2022. The Role of the Oral Immune System in Oropharyngeal Candidiasis-Facilitated Invasion and Dissemination of Staphylococcus aureus. Front Oral Health 3:851786.
60. Minarovits J. 2021. Anaerobic bacterial communities associated with oral carcinoma: intratumoral, surface-biofilm and salivary microbiota. Anaerobe 68:102300.
61. Rai AK, Panda M, Das AK, Rahman T, Das R, Das K, Sarma A, Kataki AC, Chattopadhyay I. 2021. Dysbiosis of salivary microbiome and cytokines influence oral squamous cell carcinoma through inflammation. Archives of Microbiology 203:137-152.
62. Hettmann A, Demcsák A, Decsi G, Bach Á, Pálinkó D, Rovó L, Nagy K, Takács M, Minarovits J. 2015. Infectious agents associated with head and neck carcinomas. Advances in Microbiology, Infectious Diseases and Public Health:63-80.
63. Lee A, Ghaname CB, Braun TM, Sugai JV, Teles RP, Loesche WJ, Kornman KS, Giannobile WV, Kinney JS. 2012. Bacterial and salivary biomarkers predict the gingival inflammatory profile. Journal of periodontology 83:79-89.
64. Massoure PL, Lions C, Caumes JL, Spadoni S, Gaillard PE, Bougere J. 2010. [Lethal aortic endocarditis due to Gemella morbillorum in a Djiboutian khat user]. Rev Med Interne 31:e7-9.
65. Fowoyo PT, Ogunbanwo ST. 2016. Virulence and toxigenicity of coagulase-negative staphylococci in Nigerian traditional fermented foods. Can J Microbiol 62:572-8.
66. Bhatt S, Rajesh G, Pai MB, Rao A, Shenoy R. 2014. Effects of refined extracts of smokeless tobacco on Cariogenic microorganisms: An in-vitro study. International Journal of Advanced Research 2:847-852.

67. Liu M, Jin J, Pan H, Feng J, Cerniglia CE, Yang M, Chen H. 2016. Effect of smokeless tobacco products on human oral bacteria growth and viability. Anaerobe 42:152-161.
68. Vishwakarma A, Verma D. 2021. Microorganisms: crucial players of smokeless tobacco for several health attributes. Applied Microbiology and Biotechnology 105:6123-6132.
69. Sajid M, Srivastava S, Kumar A, Kumar A, Singh H, Bharadwaj M. 2021. Bacteriome of Moist Smokeless Tobacco Products Consumed in India With Emphasis on the Predictive Functional Potential. Front Microbiol 12:784841.
70. Mortazavi H, Baharvand M, Khodadoustan A. 2014. COLORS IN TOOTH DISCOLORATION: A NEW CLASSIFICATION AND LITERATURE REVIEW. International Journal of Clinical Dentistry 7.
71. Bernardi S, Karygianni L, Filippi A, Anderson AC, Zürcher A, Hellwig E, Vach K, Macchiarelli G, Al-Ahmad A. 2020. Combining culture and culture-independent methods reveals new microbial composition of halitosis patients' tongue biofilm. MicrobiologyOpen 9:e958.
72. Jalali F, Ellett F, Balani P, Duncan MJ, Dewhirst FE, Borisy GG, Irimia D. 2021. No man's land: Species-specific formation of exclusion zones bordering Actinomyces graevenitzii microcolonies in nanoliter cultures. MicrobiologyOpen 10:e1137.
73. Gordon JH, LaMonte MJ, Genco RJ, Zhao J, Li L, Hovey KM, Tsompana M, Buck MJ, Andrews CA, Mcskimming DI. 2019. Is the oral microbiome associated with blood pressure in older women? High Blood Pressure & Cardiovascular Prevention 26:217-225.
74. Perera M, Al-Hebshi N, Perera I, Ipe D, Ulett G, Speicher D, Chen T, Johnson N. 2018. Inflammatory bacteriome and oral squamous cell carcinoma. Journal of dental research 97:725-732.
75. Zhang W, Luo J, Dong X, Zhao S, Hao Y, Peng C, Shi H, Zhou Y, Shan L, Sun Q. 2019. Salivary microbial dysbiosis is associated with systemic inflammatory markers and predicted oral metabolites in non-small cell lung cancer patients. Journal of Cancer 10:1651.
76. Schmidt BL, Kuczynski J, Bhattacharya A, Huey B, Corby PM, Queiroz EL, Nightingale K, Kerr AR, DeLacure MD, Veeramachaneni R, Olshen AB,

Albertson DG. 2014. Changes in abundance of oral microbiota associated with oral cancer. PLoS One 9:e98741.

77. Uehara O, Hiraki D, Kuramitsu Y, Matsuoka H, Takai R, Fujita M, Harada F, Paudel D, Takahashi S, Yoshida K, Muthumala M, Nagayasu H, Chiba I, Abiko Y. 2021. Alteration of oral flora in betel quid chewers in Sri Lanka. Journal of Microbiology, Immunology and Infection 54:1159-1166.
78. Anukam KC, Onwuzor I, Olise N, Duru M, Agbakoba N. 2018. Oral bacteriome compositions identified by 16S rRNA metagenomics in a randomly selected “healthy” nigerian male and female subjects. International Journal of Research and Reports in Dentistry 1:1-11.
79. Rossmann SN, Wilson PH, Hicks J, Carter B, Cron SG, Simon C, Flaitz CM, Demmler GJ, Shearer WT, Kline MW. 1998. Isolation of Lautropia mirabilis from Oral Cavities of Human Immunodeficiency Virus-Infected Children. Journal of Clinical Microbiology 36:1756-1760.
80. Yang Y, Zheng W, Cai Q, Shrubsole MJ, Pei Z, Brucker R, Steinwandel M, Bordenstein SR, Li Z, Blot WJ, Shu X-O, Long J, Xu J. 2019. Racial Differences in the Oral Microbiome: Data from Low-Income Populations of African Ancestry and European Ancestry. mSystems 4:e00639-19.
81. Downes J, Hooper SJ, Wilson MJ, Wade WG. 2008. Prevotella histicola sp. nov., isolated from the human oral cavity. International Journal of Systematic and Evolutionary Microbiology 58:1788-1791.
82. Acharya A, Chan Y, Kheur S, Kheur M, Gopalakrishnan D, Watt R, Mattheos N. 2017. Salivary microbiome of an urban Indian cohort and patterns linked to subclinical inflammation. Oral Diseases 23:926-940.
83. Larsen JM. 2017. The immune response to Prevotella bacteria in chronic inflammatory disease. Immunology 151:363-374.
84. Murugesan S, Elanbari M, Bangarusamy DK, Terranegra A, Al Khodor S. 2021. Can the Salivary Microbiome Predict Cardiovascular Diseases? Lessons Learned From the Qatari Population. Front Microbiol 12:772736.
85. Könönen E, Gursoy UK. 2021. Oral Prevotella Species and Their Connection to Events of Clinical Relevance in Gastrointestinal and Respiratory Tracts. Front Microbiol 12:798763.

86. Karabudak S, Ari O, Durmaz B, Dal T, Basyigit T, Kalcioglu MT, Durmaz R. 2019. Analysis of the effect of smoking on the buccal microbiome using next-generation sequencing technology. J Med Microbiol 68:1148-1158.
87. Fteita D, Könönen E, Gürsoy M, Ma X, Sintim HO, Gürsoy UK. 2018. Quorum sensing molecules regulate epithelial cytokine response and biofilm-related virulence of three Prevotella species. Anaerobe 54:128-135.
88. Jia YJ, Liao Y, He YQ, Zheng MQ, Tong XT, Xue WQ, Zhang JB, Yuan LL, Zhang WL, Jia WH. 2021. Association Between Oral Microbiota and Cigarette Smoking in the Chinese Population. Front Cell Infect Microbiol 11:658203.
89. Riggio MP, Lennon A, Rolph HJ, Hodge PJ, Donaldson A, Maxwell AJ, Bagg J. 2008. Molecular identification of bacteria on the tongue dorsum of subjects with and without halitosis. Oral Dis 14:251-8.
90. Saxena R, Prasoodanan PKV, Gupta SV, Gupta S, Waiker P, Samaiya A, Sharma AK, Sharma VK. 2022. Assessing the Effect of Smokeless Tobacco Consumption on Oral Microbiome in Healthy and Oral Cancer Patients. Front Cell Infect Microbiol 12:841465.
91. Iwai T, Inoue Y, Umeda M, Huang Y, Kurihara N, Koike M, Ishikawa I. 2005. Oral bacteria in the occluded arteries of patients with Buerger disease. Journal of Vascular Surgery 42:107-115.
92. Dahlén G, Nauclér C, Nordwall S, Suksu-art N. 2010. Oral microflora in betel-chewing adults of the Karen tribe in Thailand. Anaerobe 16:331-336.
93. Mättö J, Asikainen S, Väisänen ML, Von Troil-Lindén B, Könönen E, Saarela M, Salminen K, Finegold SM, Jousimies-Somer H. 1999. Beta-lactamase production in Prevotella intermedia, Prevotella nigrescens, and Prevotella pallens genotypes and in vitro susceptibilities to selected antimicrobial agents. Antimicrob Agents Chemother 43:2383-8.
94. Pompilio A, Di Bonaventura G, Gherardi G. 2019. An Overview on Streptococcus bovis/Streptococcus equinus Complex Isolates: Identification to the Species/Subspecies Level and Antibiotic Resistance. Int J Mol Sci 20.
95. Mostefaoui Y, Bart C, Frenette M, Rouabhia M. 2004. Candida albicans and Streptococcus salivarius modulate IL-6, IL-8, and TNF-α expression and secretion by engineered human oral mucosa cells. Cellular Microbiology 6:1085-1096.

96. Vellappally S, Fiala Z, Šmejkalová J, Jacob V, Shriharsha P, Fiala Z, Smejkalová J. 2007. Influence of tobacco use in dental caries development. Central European journal of public health 15.
97. Anonymous. 1998. Streptococcus mitis with Unusually High Level Resistance to β-Lactam Antibiotics. Microbial Drug Resistance 4:45-49.
98. Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, Goodson JM. 2005. The salivary microbiota as a diagnostic indicator of oral cancer: A descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. Journal of Translational Medicine 3:27.
99. Baraniya D, Jain V, Lucarelli R, Tam V, Vanderveer L, Puri S, Yang M, Al-Hebshi NN. 2020. Screening of health-associated oral bacteria for anticancer properties in vitro. Frontiers in cellular and infection microbiology 10:575656.
100. Hojo K, Mizoguchi C, Taketomo N, Ohshima T, Gomi K, Arai T, Maeda N. 2007. Distribution of salivary Lactobacillus and Bifidobacterium species in periodontal health and disease. Bioscience, biotechnology, and biochemistry 71:152-157.
101. Doel JJ, Benjamin N, Hector MP, Rogers M, Allaker RP. 2005. Evaluation of bacterial nitrate reduction in the human oral cavity. European journal of oral sciences 113:14-19.
102. McDaniel S, McDaniel J, Howard KM, Kingsley K. 2021. Molecular Screening and Analysis Reveal Novel Oral Site-Specific Locations for the Cariogenic Pathogen Scardovia wiggsiae. Dentistry Journal 9:73.
103. Takahashi N, Sulijaya B, Yamada-Hara M, Tsuzuno T, Tabeta K, Yamazaki K. 2019. Gingival epithelial barrier: regulation by beneficial and harmful microbes. Tissue Barriers 7:e1651158.
104. Börnigen D, Ren B, Pickard R, Li J, Ozer E, Hartmann EM, Xiao W, Tickle T, Rider J, Gevers D, Franzosa EA, Davey ME, Gillison ML, Huttenhower C. 2017. Alterations in oral bacterial communities are associated with risk factors for oral and oropharyngeal cancer. Scientific reports 7:17686-17686.
105. Lin CW, Chen YT, Ho HH, Hsieh PS, Kuo YW, Lin JH, Liu CR, Huang YF, Chen CW, Hsu CH, Lin WY, Yang SF. 2021. Lozenges with probiotic strains enhance oral immune response and health. Oral Dis doi:10.1111/odi.13854.
106. Engevik MA, Luk B, Chang-Graham AL, Hall A, Herrmann B, Ruan W, Endres BT, Shi Z, Garey KW, Hyser JM, Versalovic J. 2019. Bifidobacterium

dentium Fortifies the Intestinal Mucus Layer via Autophagy and Calcium Signaling Pathways. mBio 10.

107. Grover HS, Luthra S. 2011. Probiotics–the nano soldiers of oral health. J Indian Acad Clin Med 13:s48-s54.

108. Meurman JH. 2005. Probiotics: do they have a role in oral medicine and dentistry? European journal of oral sciences 113:188-196.

5 Traditional memory and reshaping of Sudanese gut metagenomics in relation to diet and migratory movements to westernised countries

Amel Sami [1,3], Imad Elimairi [3], Kiera Healy [1], C. Anthony Ryan [4], R. Paul Ross [1] and Catherine Stanton [1,2]

Affiliations:

1- APC Microbiome Institute, School of Microbiology, University College Cork, Cork T12 YN60

2 - Teagasc Food Research Centre, Moorepark, Fermoy, Cork, P61 C996, Ireland

3 - Department of Oral and Maxillofacial Surgery and Oral Medicine, Faculty of Dentistry, National Ribat University, Nile street, Khartoum, 1111, Sudan

4 - Department of Paediatrics and Child Health, University College Cork, Cork T12 DFK4

5.1 Abstract

Migratory movement of populations from Africa to westernised countries like Ireland can impact the traditional gut microbiome and promote microbial variations that may be atypical to a specific population. This study highlights new insights into the African gut microbiome with the relevance of the westernisation effects on diet, metabolomics and microbiome shaping that may influence future health directions.

Twenty-five Irish/ Caucasian participants living in Ireland [Irel], 49 Sudanese participants who migrated to Ireland [SudIrel] and 67 participants living in Sudan [Sud], aged between 6-70 years of age were included in this multi-directional study that consisted of three components. The first component involved the modification and analysis of the European prospective investigation into cancer and nutrition Norfolk food frequency questionnaire [EPIC - Norfolk FFQ] for the Sudanese population, employed for nutrient and food group comparisons across the three cohorts. In addition, a further sublayer validation was achieved utilising the online 24-hour recall automated self-administered 24 questionnaires [ASA-24]. The second component consisted of DNA extraction and 16SrRNA sequencing of the V3-V4 hypervariable regions in 141 faecal samples, including PICRUSt analysis, while the third component was achieved in 40 faecal water samples through semi-polar metabolite and short-chain fatty acid analysis to determine metabolomic variations.

For nutritional analysis, eta^2 was utilised to determine the strength of associations between FFQ and ASA24 formats, while correlation plotting was utilised to compare nutritional and microbial correlations. Microbiome statistical comparisons included Metagenomeseq, alpha and beta diversity and one-way Anova, while LEfSe plotting highlighted significant discriminants in PICRUSt and metabolomic findings.

Six food types, including meat, fruit and vegetables, were consumed at a reduced intake in the Sud cohort with a high strength of association and large effect size [eta^2 =>0.14]. Three food types, egg and egg products [eta^2 =0.04], nuts and seeds [eta^2=0.05] and non-alcoholic beverages [eta^2=0.108], were consumed highly in Sud, albeit with low and medium effect sizes. The average energy intake [measured in

kcal/day] between groups was statistically different [p <0.001] and was found to be lowest in Sud [1077 kcal/day]. The diet of the Sud cohort was found to be low in several nutrients, including carbohydrates, fibre, protein, fat, and vitamin D [p<0.001] while at the same time, it was high in vitamin B5 [pantothenic acid] and intermediate in Vitamin K levels. The genus *Megamonas* was the most positively associated genus in Sud to several nutrients that included calcium, protein, and fat intake. At the same time, reduced *Lachnospiraceae* in Sud may be a novel factor in the endemically associated iodine deficiency in Sudan.

The gut microbiome of Sud and Irel differed significantly in β diversity [p < 0.001], while SudIrel predominantly exhibited a 'sandwiched' status between Sud and Irel with either a loss or gain from the traditional microbiome composition. Genera enriched in Sud and depleted in Irel included *Prevotella* [q=3.169e-06] and *Lactobacillus* [q=2.918e-09], while in Irel, *Lachnospira* [q=1.623e-09] and *Bacteroides*: [q=3.056e-05] were abundant compared to the Sudanese groups.

PCA of metabolomic compounds highlighted 27 metabolites that varied significantly between Irel and Sud, 21 metabolites between Sud and SudIrel and none between Irel and SudIrel. The short-chain fatty acids, isobutyric [p=0.0], and isovaleric acids [0.002], were significantly lower in Irel. LEfSe plotting of PICRUSt data indicated multi-drug resistance activity pathways in the Irel cohort [K09686], mineral and heavy metal transport activity in Sud [K06199] and SudIrel, multiple prominent carbohydrate metabolism pathways.

Distinctions and preservations in the gut microbiome are evident in African and European populations, while those who migrate show a 'sandwiched' microbiome composition integrated by memory and reshaping. Genera findings may be related to specific nutrient alterations amongst the cohorts, while differing genera may also be responsible for the same nutrient profiles, high or low depending on migration and geographical status.

5.2 Introduction

The gut microbiome and metabolome of the Sudanese population [> 40 million] have never been previously explored. The Sudanese diet primarily consists of fermented foods, legumes, dairy, millet, and sorghum-based products. The Sudanese gut microbiome has likely sculpted its way through generations, determined by tradition but also in adaption to the changing environmental impacts at national and international levels.

Current African gut microbiome research is currently limited in exploration of around 168 human microbiome studies [using next generation sequencing] involving African participants from 33 out of 54 countries. South Africa, Kenya, and Uganda [1] represent the African countries with the most microbiome data, primarily in malnutrition and infection [2]. Of note is that conclusions from westernised countries and Asia alone cannot fully comprehend the gut microbiome landscape of Africans, particularly the Sudanese.

Problematically, Sudan now sees high incidences of non-communicable diseases such as cardiovascular disease and type two diabetes. Cardiovascular disease is thought to account for 73% of acute admissions compared to negligent in the last century, while the prevalence of type two diabetes in rural Sudan is only around 2.5% compared to 19% in urban communities [3] [4]. In addition, those Sudanese who migrate to westernised countries may harbour a distinct susceptibility profile to diseases and westernised related diseases such as IBD [5].

This study aimed to understand the current gut microbiome and metabolome landscape of the Sudanese living both in their home country and in a westernised environment. In this analysis, we question whether microbiome memory could be retained while also analysing what components of the microbiome are reshaped after moving to Ireland. We compare those living in Ireland and are Caucasian in background [Irel], those Sudanese living in Ireland [SudIrel] and those Sudanese living in Sudan [Sud]. We explore the metabolomic findings and, furthermore, analyse Sudanese nutritional

habits through the first-time utilisation of a modified food frequency questionnaire and the sublayer validation of its use on the Sudanese.

5.3 Materials and methods

Participant recruitment

Data were collected from the two countries, Sudan and Ireland. Consent was achieved from the Sudanese ethical committee at National Ribat University and Cork University College Cork ethics board. We recruited 141 participants, 6-70 years of age. These were further categorised as 25 Irish/ Caucasian participants living in Ireland [Irel], 49 Sudanese participants who migrated to Ireland [SudIrel] and 67 participants living in Sudan, Africa [Sud]. Participants were graded according to the American Society of Anaesthesiologists [ASA] physical status classification system, grades 1 and 2 [6] [Table 1]. For the SudIrel group, participants with a migration history to Ireland for at least one year were included.

Dietary questionnaires

The European prospective investigation into cancer and nutrition Norfolk food frequency questionnaire [EPIC - Norfolk FFQ] is a semi-quantitative questionnaire aimed at recording average nutrient intakes over a retrospective period of at least 12 months. It was modified and employed for macro and micronutrient and food group analysis covering over 290 foods across the three cohorts, Sud, SudIrel and Irel [7].

The University of Cambridge, UK, FETA software, version 2-53, was downloaded from ‘www.epic-norfolk.org.uk’ to record nutrient values, and output one was utilised [average daily nutrient and food group intakes] in wide format. Twenty-four foods were added to the original FETA database for those living in Sudan, while one food was added to the questionnaire for those living in Ireland. This was achieved by modifying the FETA input/look-up files. Data were entered manually on digital spreadsheets [Microsoft Excel] by two independent researchers, in Sudan and Ireland to account for errors. Each recorded food was framed within a nine-point scale where ‘1’ is for never or < once a month and nine is for >6 times a day. A code of ‘-9’ indicated a data cell not recorded by the participant.

Table 1: Participants included in this study were graded according to the American Society of Anaesthesiologists [ASA] physical status classification system, grades 1 and 2. DM=diabetes mellitus, HTN= hypertension. BMI = body mass index. OSA = obstructive sleep apnoea.

ASA Classification	**Definition**	**Adult examples including but not limited to:**	**Paediatric examples including but not limited to:**
ASA 1	A normal healthy patient	Healthy, non-smoking, no or minimal alcohol use	Healthy [no acute or chronic disease], normal BMI percentile for age
ASA 2	Patients with mild systemic disease	Mild diseases only without substantive functional limitations. Current smoker, social alcohol drinker, pregnancy, obesity [30<BMI<40], well-controlled DM/HTN, mild lung disease	Asymptomatic congenitcal cardiac disease, well controlled dysrytmias, asthma without exacerbation, well controlled epilepsy, non-insulin dependant diabetes mellitus, abnormal BMI percentile for age mild/moderate OSA, oncologic state in remission autism with mild limitations

Nutrient data retrieved were then imported in CSV format, further processed, and compared through SPSS version 28. Nutrient intakes were analysed for mean, minimum and maximum levels of nutrient intake and compared through parametric tests. Nutrient intakes were recorded in mcg, mg, and g. The eta coefficient [eta^2] was used to measure the correlation strength between the categorical and scale nutrient variables. For example, an eta^2 value of under 0.01-0.05 signified a small effect size, 0.06-0.013, a medium effect size and 0.14 or higher was a large effect size.

Sublayer validation of Sudanese EPIC- Norfolk FFQ through the automated self-administered 24-hour recall system [ASA-24]

The EPIC Norfolk FFQ is a validated food frequency questionnaire but predominantly utilised on the European population. We carried out the sublayer validation of this questionnaire for its use on the Sudanese population using an additional online recall system [ASA-24] on 32 Sudanese subjects. These participants were asked to fill in the EPIC- Norfolk FFQ and the ASA-24 on two different days. Results were analysed using non–parametric tests to account for any non-normal distributions. In addition, similar mean nutrient levels between the EPIC-Norfolk FFQ and the ASA-24 were evaluated using independent samples Mann – Whitney U test. Non-significant values were taken as a positive outcome on validation.

Faecal sample collection

Faecal samples were collected and placed in a sterile falcon 50ml tube, labelled and frozen immediately at minus 80°C. Those samples from Sudan were prepared for shipment under dry ice to Cork, Ireland.

Faecal DNA Extraction

Faecal samples, positive and negative controls underwent DNA extraction by DNeasy Powersoil Kit [Qiagen, Hilden, Germany] utilising manufacturer's instructions. In summary, 0.25g of each faecal sample was placed in a 2ml screw capsule containing 0.25g of sterile zirconium beads. 1ml of lysis buffer was added to samples, and they were homogenised for three minutes using a bead beater. Samples were incubated at 70^{o}C for 15 minutes alongside gentle shaking and centrifugation [16,000 X g] at 4^{o}C for five minutes. The supernatant was transferred to a new 2ml Eppendorf tube, and 300µl of lysis buffer was added to

each sample with the repeated above steps. Then, 10M ammonium acetate [260µl] was added to each tube with further centrifugation [16,000 X g] at 4^{o}C for 10 minutes. The supernatant was then transferred into two 1.5ml Eppendorf tubes, and 650µl isopropanol was added with mixtures left overnight at minus 20^{o}C.

The following day, samples were centrifuged at 4^{o}C for 15 minutes at 16,000 X g. The supernatant was removed, and the nucleic acid pellets were washed with 70% ethanol. After all, ethanol was evaporated, 100µl TRIS EDTA buffer was added, and the two aliquots were re-combined. 2µl of DNase, RNase-free endonuclease was added, and samples were incubated at 37^{o}C for 15 minutes. Next, 15 µl proteinase K and 200µl buffer were mixed and samples were incubated at 70^{o}C for a further 10 minutes after which 200µl of ethanol were added.

Samples were then transferred to QIamp columns and centrifuged at 16,000 X g for one minute. The flow-through was discarded, and 500µl buffer was added with further centrifugation. The dry column was also centrifuged for one minute, and 200µl buffer was added, incubating at room temperature for two minutes. Finally, DNA was eluted, transferred to Eppendorf tubes and frozen at minus 20^{o}C. Samples were quantified using the Qubit dsDNA High Sensitivity Kit [Invitrogen, Life Technologies, Grand Island, NY].

Library preparation [Amplification and Sequencing]

The 16S V3 – V4 hypervariable regions were targeted using AMPure XP beads [Beckman Coulter, Brea, CA, USA] and universal forward and reverse primers with overhang adapters that followed standard IUPAC nucleotide nomenclature. Amplification was achieved with the MiSeq protocol for the preparation of metagenomics libraries [Illumina, Inc., San Diego, CA, USA]. PCR was performed in a 50 µL reaction mixture containing 10 ng template DNA and 2x KAPA HiFi HotStart Ready Mix. The following thermal cycling conditions were used: 3 min of initial denaturation at 95°C; 30 cycles each of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 30 s; and the last step at 72°C for 5 min. To purify the 16S V3-V4 amplicon from free primers and primer dimer species, PCR clean-up was carried out with AMPure XP beads [Beckman Coulter, Brea, CA, USA]. Electrophoresis verified the size of the product expected at 550 base pairs, and this was achieved by testing PCR amplified products on 1.5% agarose gel with Invitrogen

DNA loading dye [ThermoFisher Scientific]. The gel was visualised with 300 nm ultraviolet light. Samples were positive on the agarose gel while the control samples were negative. Index PCR was continued by attaching dual indices and Illumina sequencing adapters using Nextera XT index kits. The following PCR thermal cycling conditions were used: 3 min of initial denaturation at 95°C; 8 cycles each of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and elongation at 72°C for 30 s; and the last step at 72°C for 5 min. AMPure XP beads [Beckman Coulter, Brea, CA, USA] were utilised for final clean-up and final library quantified with the Qubit® 2.0 Fluorometer and the Qubit dsDNA HS Assay kit [ThermoFisher Scientific].

Taxonomy and bioinformatics

Samples were normalised, and the final library was run on an Agilent high-sensitivity chip and quantified by qPCR using the KAPA Illumina Library quantification kit. The 300 bp paired-end FASTQ product generated from 16S sequencing was merged using Fast Length Adjustment of Short Reads [FLASH] using default parameters [8]. Quantitative Insights Into Microbial Ecology [QIIME] split_libraries_fastq.py script was used for demultiplexing and filtering the FASTQ sequence data. Further quality filtering was performed using the USEARCH analysis tool [9]. Single unique reads were removed; following denoising, chimera removal and grouping into operational taxonomic units [OTUs] at 97% similarity and was performed using USEARCH v7 [64 bit] [Robert C. Edgar et al., 2010]. Alignment of OTUs was performed using python nearest alignment space termination [PyNAST]; a flexible tool for aligning sequences to a template alignment][10]. Assigning of taxonomic ranks was performed by using BLAST against the SILVA SSURef database release 132 [11]. For species determination, the U.S National Library of Medicine, and National centre for biotechnology Information was then accessed in order to assign species identity to those FASTA sequences with > 98% percentage homogeneity by using the BLAST® tool for bacterial 16S rRNA sequencing database [12]. We further deployed the nucleotide Basic Local Alignment Search Tool [BLASTn] taxonomy database helping to align OTU sequences with greater than 98% identity to their up-to-date species database. The functional roles of microbiota were also predicted using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States [PICRUSt] [13]. These assimilations were determined using the Kyoto Encyclopaedia for Genes and Genomes [KEGG] pathways.

Statistical analysis

Graphical outputs were achieved through the Microbiome Analyst R package [14]. Data were normalised and profiled for compositional and functional analysis. Relative abundances were converted to a percentage and log space upper and lower interquartile ranges with the variations of the medians determined between Sud, SudIrel and Irel.

Linear Discriminant Analysis Effect Size [LEfSe] analysis was utilised to identify microbiome discriminants between the three groups, while one-way Anova was employed to test for significant differences. DeSeq2 was utilised to calculate q-values using the Benjamini-Hochberg procedure correcting for p values through the control of false discovery rates [15], and sparse high throughput sequencing data [MetagenomeSeq] was approached to account for rare taxa. Alpha diversity profiled richness, evenness, and abundance estimates of microbiome communities amongst faecal samples. Beta diversity assessed relative abundances of OTUs, clustering and outliers and was formulated using Bray–Curtis matrices and visualised using unconstrained analysis; Principal coordinate analysis [PcoA]. Finally, correlation plotting was utilised to determine multivariable associations between the microbiome and the metadata group features.

Metabolomics

Faecal extraction

We analysed the semi-polar metabolite profile and short-chain fatty acid content of 40 faecal samples. All samples were processed in randomised order to reduce the introduction of systematic variations during sample preparation and analysis. Samples were extracted by first fortifying with stable isotope labelled Isoleucine and extracted with water [1:4 wt/vol] in 1.5 ml tubes. The temperature was kept below 4°C during extraction, after which tubes were vortexed for 2 minutes until a suspension was achieved. This was sonicated for 5 minutes. Samples were pre-separated by vortexing for 30 seconds, followed by centrifugation at 4000 X g for 5 min. The supernatant was transferred to a new tube and centrifuged at 16,000 X g for 10 minutes at 4°C. Finally, the supernatant was filtered through a 0.2 um spin filter and stored at minus 20°C for further analysis. A quality control sample was prepared by pooling small equal aliquots from each sample extract to obtain a representative average of the entire set.

This sample was treated and analysed at regular intervals throughout the sequence. Possible matrix effects on short-chain fatty acids for quantification were tested by spiking aliquots of the quality control sample at a minimum of two levels.

Semi-polar metabolite analytics

Sample analysis was carried out by MS-Omics [Denmark] as follows. The analysis was carried out using a Vanquish LC coupled to Orbitrap Exploris 240 MS [ThermoFisher Scientific]. An electrospray ionization interface was used as an ionisation source. Analysis was performed in polarity switching ionisation mode. The UPLC was performed using a slightly modified version of the protocol described by Catalin et al. [UPLC/MS Monitoring of Water-Soluble Vitamin Bs in Cell Culture Media in Minutes, Water Application note 2011, 720004042en]. Peak areas were extracted using Compound Discoverer 3.2 [ThermoFisher Scientific]. Identification of compounds were performed at four levels; Level 1: identification by retention times [compared against in-house authentic standards], accurate mass [with an accepted deviation of 3ppm], and MS/MS spectra, Level 2a: identification by retention times [compared against authentic in-house standards], accurate mass [with an acceptable deviation of 3ppm]. Level 2b: identification by accurate mass [with an accepted deviation of 3ppm], and MS/MS spectra, Level 3: identification by accurate mass alone [with an accepted deviation of 3ppm].

For short-chain fatty analysis, samples were acidified using hydrochloride acid, and deuterium-labelled internal standards were added. All samples were analysed in a randomised order. Analysis was performed using a high polarity column [Zebron™ ZB-FFAP, GC Cap. Column 30 m x 0.25 mm x 0.25 μm] installed in a GC [7890B, Agilent] coupled with a time-of-flight MS [Pegasus® BT, LECO]. The system was controlled by ChromaTOF® [LECO]. Raw data were converted to netCDF format using Chemstation [Agilent] before the data were imported and processed in Matlab R2014b [Mathworks, Inc.] using the PARADISe software described by [16]. The following ten compounds were analysed and quantified in umol/g: acetic acid, formic acid, propionic acid, isobutyric acid, butyric acid, isovaleric acid, pentanoic/valeric acid, 4-methylpentanoic [isocaproic] acid, hexanoic acid and heptanoic acid. Short-chain fatty acids were profiled using T-test results, $\log^2$ ratios between Irel/Sud, Sud/SudIrel and Irel/SudIrel. The p values and corrected significant testing results

[i/m] Q with Benjamini-Hochberg were reported. A critical value for each comparison was detected depending on the number of total tests [10], and ranks were utilised to order the most significant compounds. Principal component analysis, a multivariate statistic, was used to plot data into a 2d scatter plot.

5.4 Results

Nutrient, food group and FFQ data

Average mean intakes of nutrients produced by the EPIC-Norfolk FFQ and ASA24 form were recorded. We looked for non-significance between the recall method and food frequency method to detect validity of its use amongst the Sudanese population. Seventeen nutrients were found to be non-significant in mean value distributions between the EPIC-Norfolk FFQ and ASA24, while ten nutrients varied in mean values [Table 2]. This indicated that most nutrient intake [63%] could be reproducibly recorded using the modified EPIC-Norfolk FFQ. Food types consumed in lowest amounts in Sud included cereals and cereal products, fats and oil, fish and fish products, fruits and meat and meat products [Table 3]. Egg and egg products and non-alcoholic beverages were found to be highly consumed in Sud. The highest consumption of nuts and seeds was in SudIrel. The energy Kcal distributions were significantly lower [p=0.002] amongst the EPIC-Norfolk FFQ compared to the ASA24 system. Both Sudanese groups exhibited low or negligible alcohol intake.

Table 2: Mean value comparisons from food frequency questionnaires and 24-hour recall were evaluated in Table 2. Nutrients with a non-significant value comparison between mean intakes or similar in recording between the two methods, indicated validity of the EPIC-Norfolk FFQ modified for the Sudanese population.

No.	**Nutrient**	**Epic-Norfolk FFQ mean value**	**ASA-24 mean value**	**Distribution the same across two captures of intake [p value non-significant]**	**Distribution varied between two captures of intake [p value significant]**
1	Fibre	30.31	34.69	0.347	
2	Carbohydrates [total] [g]	32.69	32.31	0.936	
3	Proteins [g]	38.03	26.97		0.017
4	Fats [g]	30.66	34.34	0.428	
5	Monounsaturated fatty acids [g]	27.56	37.44		0.034
6	Polyunsaturated fatty acids [g]	21.09	43.91		<0.001

7	Saturated fatty acids [g]	34.34	30.66	0.428	
8	Calcium [mg]	41.34	23.66		<0.001
9	Iron [mg]	45.97	19.03		<0.001
10	Cholesterol [mg]	36.31	28.69	0.101	
11	Folate [mcg]	26.06	38.94		0.006
12	Potassium [mg]	39.53	25.47		0.003
13	Magnesium [mg]	35.66	29.34	0.175	
14	Sodium [mg]	30.13	34.88	0.308	
15	Niacin [mg]	31.94	33.06	0.809	
16	Vitamin A [mcg]	33.66	31.34	0.619	
17	Vitamin B2 [mg]	36.25	28.75	0.107	
18	Selenium [mcg]	23.19	41.81		<0.001
19	Vitamin B1[mg]	28.91	36.09	0.123	
20	Vitamin B12 [mcg]	42.16	22.84		<0.001
21	Vitamin B6 [mg]	36.09	28.91	0.123	
22	Vitamin C [mg]	32.94	32.06	0.851	
23	Vitamin D [mcg]	34.91	30.09	0.301	
24	Vitamin E [mg]	31.34	33.66	0.619	

25	Vitamin K [mcg]	19.53	45.47		<0.001
26	Zinc [mg]	30.41	34.59	0.368	
27	Copper [mg]	32.97	32.03	0.840	
Total				**17**	**10**

Table 3: Average food consumptions patterns amongst 11 food groups. Comparisons were evaluated between Sud, SudIrel and Irel. Seven high strengths of associations were produced indicating values are comparable for the populations questioned. Three values were of a lower strength association indicating that the variations existing, may not fully appreciate food group pattern in analysed cohort.

	Mean values			
Food consumptions	**Sud**	**SudIrel**	**Irel**	**Eta^2**
Cereal and cereal products	159g	428.7g	585.6g	0.328 [high]
Fats and oils	8.42g	24.4g	40.8g	0.354 [high]
Fish and fish products	9g	105.2g	103.8g	0.217 [high]
Eggs and egg products	48g	37.7g	31.4g	0.04 [low]
Fruits	172.8g	572.1g	442.5g	0.29 [high]
Vegetables	237.7g	421.5g	477.1g	0.175 [high]
Meat and meat products	43.5g	301.7g	314.1g	0.265 [high]
Milk and milk products	380g	440.8g	491.6g	0.013 [low]
Non- alcoholic beverages	892g	628.4g	560g	0.108 [medium]
Nuts and seeds	8g	21.1g	9.1g	0.05 [low]
Sugars, preserves and snacks	20.8g	60.8g	99g	0.311 [high]

Eta^2 associations

Table 4 summarises mean nutrient values with significance levels and eta^2 [strength of association] scores. Energy consumption was lowest in Sud [1077 kcal], highest in Irel [3442 kcal] and intermediate in SudIrel [2729 kcal], which was statistically significant [$p<0.001$] with high strength of association [eta^2=0.435].

Total carbohydrate mean values were lowest in Sud and highest in Irel with eta^2 = 0.315 [$p <0.001$]. Cholesterol mean values were similar in Irel and SudIrel [560 and 546 mg, respectively] but were lowest in Sud [390mg] [eta^2 =0.10]. Fibre mean values were the lowest in Sud [11.7g] and highest in Irel [36.1g]. Proteins were significantly lower in Sud [74.2g] compared to 184.9g in Sud and 175.5g in SudIrel with a large effect size [eta^2 = 0.29]. Total fats were low in Sud [average 66g] compared to 147 and 171g in SudIrel and Irel respectively.

Iodine mean values were lowest in Sud [72.9 mcg] and highest in SudIrel [199.4 mcg] [eta^2 =0.362] with $p< 0.001$. Calcium mean values [mg] were lowest in Sud and highest in SudIrel with eta^2 =0.041, [$p = 0.05$]. Average Iron levels were increased in SudIrel [59.3mg] and were lowest in those living in Sudan [21.1mg]. Potassium and sodium levels [mg] were both lowest in Sud [2830 and 2422mg, respectively]. Vitamin B5 [pantothenic acid] average, on the other hand, was found to be higher amongst the Sud group [1.17mg] compared to Irel [0.27mg] and SudIrel [0.53mg], respectively. Vitamins B1, B6 and B12 were all found to be low in Sud.

Microbiome findings

A total count of 11438812 reads was obtained and assigned to 2074 OTUs. *Firmicutes* was the commonest phylum in all groups comprising an average of 73% of gut microbiome composition. This was followed by *Bacteroidetes* in Irel [14%] and *Actinobacteria* in Sud [13%]. SudIrel exhibited a middle lane expression of [15%] *Bacteroidetes* and [12%] *Actinobacteria*. *Proteobacteria* were not evident in SudIrel. Other phyla in all groups included *Verrucomicrobia, Euryarchaeota, Desulfobacterota* and *Patescibacteria*.

Lachnospiraceae was the most dominant family in the study groups with Irel having the greatest abundance [41%] and Sud having the lowest abundance [29%], while SudIrel was in between [30%] [Figures 1a-1b]. Genera in this family were also low in

Sudanese cohorts and included *Lachnospira* [q=1.623e-09], *Coprococcus* [q=0.006], *Blautia* [q=5.3133e-6], *Roseburia* [q= 3.6265e-5] and *Dorea* [q=1.6031e-4] [Figures 1c-1g]. In addition, the median abundance of *Marvinbryantia* [Figure 1h] was found to be lowest in Irel, a trend unlike the other *Lachnospiraceae* genera [q= 0.002].

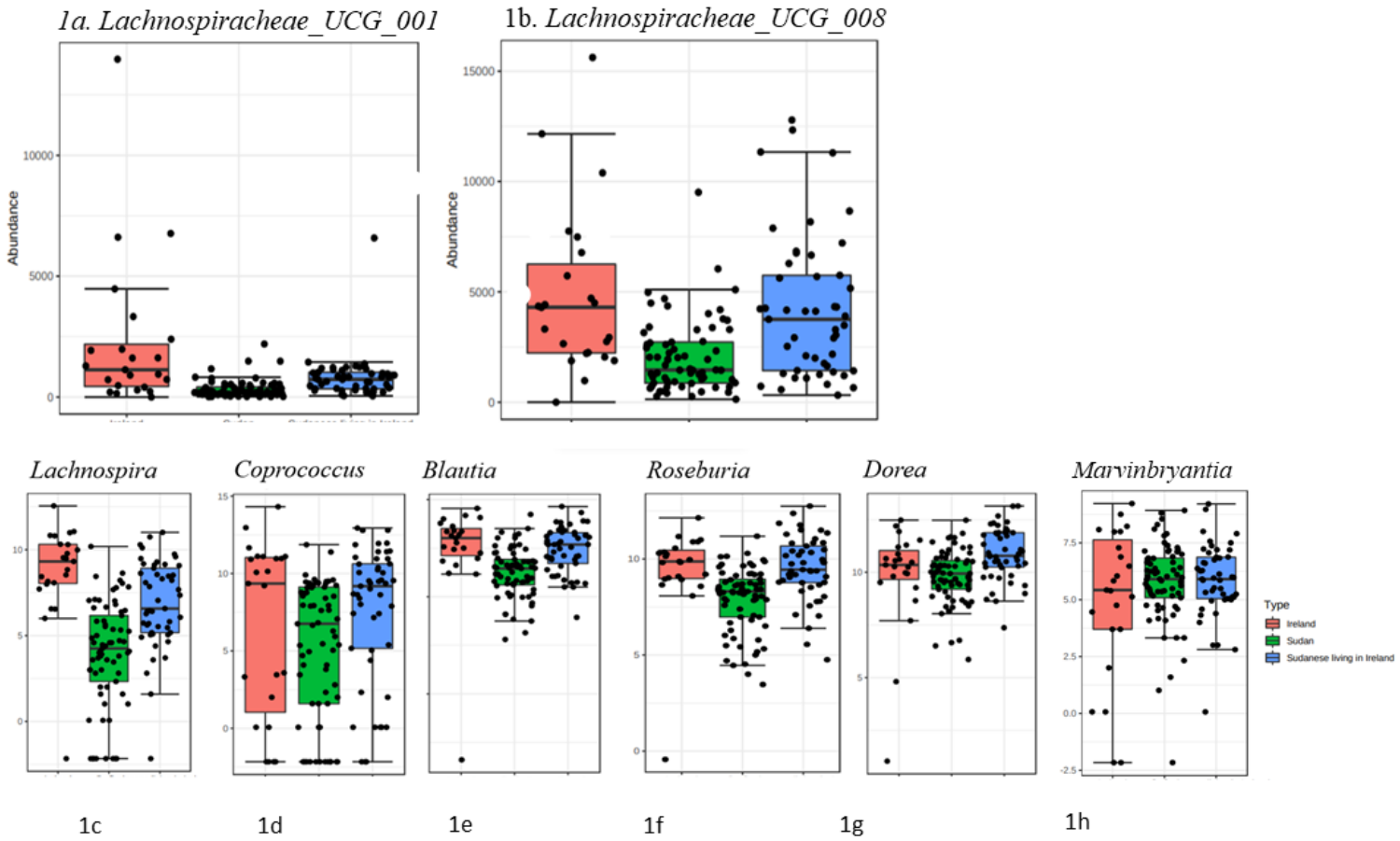
Table 4: Mean nutrient value summary with significance levels and eta^2 [strength of association] scores. A generally lower nutrient intake was observed in Sud compared to the remaining groups while in this cohort, fibre intake in Irel was relatively high compared to the population average [~20g] [120].

Nutrient	**Mean values**			**P value**	**0.01= small, 0.06 = medium & 0.14 or higher = large effect size**
	Sud	**Irel**	**SudIrel**		**Eta^2 [strength of association]**
Macronutrients					
Proteins [g]	74.2g	184.9g	175.5g	<0.001	0.297
Fibre [g]	11.7g	36.1g	32g	<0.001	0.413
Carbohydrates [g]	211g	463g	425g	<0.001	0.315
Fats [g]	66.6 g	171.2 g	147.8	<0.001	0.313
Monounsaturated fatty acids [g]	17.1 g	55.5 g	49.7 g	<0.001	0.375
Polyunsaturated fatty acids [g]	6.56g	26.8 g	24.6g	<0.001	0.345
Saturated fatty acids [g]	30.4 g	64g	52.2g	<0.001	0.241
Other nutrients					
Calcium [mg]	1326mg	1696mg	1578mg	0.05	0.041

Cholesterol [mg]	390mg	560mg	546mg	<0.001	0.100
Iron [mg]	21.1mg	46.4mg	59.3mg	<0.001	0.130
Folate [mcg]	215 mcg	477 mcg	507 mcg	< 0.001	0.337
Iodine [mcg]	72.9 mcg	238.6 mcg	199.4 mcg	<0.001	0.362
Potassium [mg]	2830 mg	6708 mg	6366 mg	<0.001	0.329
Sodium [mg]	2422 mg	6271 mg	4153 mg	<0.001	0.312
Niacin [mg]	12.6 mg	44.9 mg	47.1 mg	<0.001	0.319
Vitamin A [mcg]	657 mcg	1170 mcg	1368.3 mcg	<0.001	0.108
Vitamin C [mg]	111 mg	205.1 mg	202.6 mg	<0.001	0.199
Vitamin B2 riboflavin [mg]	2.05 mg	3.35 mg	3.34 mg	<0.001	0.177
Vitamin D [mcg]	4.77 mcg	8.74 mcg	7.78 mcg	<0.001	0.123
Vitamin B5 pantothenic acid [mg]	1.17mg	.27mg	.53mg	<0.001	0.329
Vitamin B1 thiamine [mg]	0.92 mg	3.32 mg	2.22 mg	<0.001	0.454

Vitamin B12 [mcg]	6.26 mcg	14.68 mcg	18.99 mcg	<0.001	0.281
Vitamin B6 [mg]	1.42 mg	4.28 mg	4.35mg	<0.001	0.375
Vitamin E [mg]	7.93 mg	22.3 mg	19.61 mg	<0.001	0.334
Vitamin K [mcg]	15.8g	12.8g	35.1g	<0.001	0.177
Selenium [mcg]	49.4 mcg	134.54 mcg	116.6	<0.001	0.351
Zinc [mg]	6.77 mg	19.2 mg	17.36 mg	<0.001	0.364

Figure 1:


1a. Lachnospiracheae_UCG_001
1b. Lachnospiracheae_UCG_008
Abundance
Lachnospira
Coprococcus
Blautia
Roseburia
Dorea
Marvinbryantia
Type
Ireland
Sudan
Sudanese living in Ireland
1c
1d
1e
1f
1g
1h

1a and 1b: Pink=Irel, green=Sud and blue=SudIrel. One-way Anova abundance findings. Lachnospiraceae are reduced in abundance in Sud. 1c-1h. log abundances of Lachnospiraceae associated genera show continuously marked lower medians in Sud with increased composition in SudIrel and Irel.

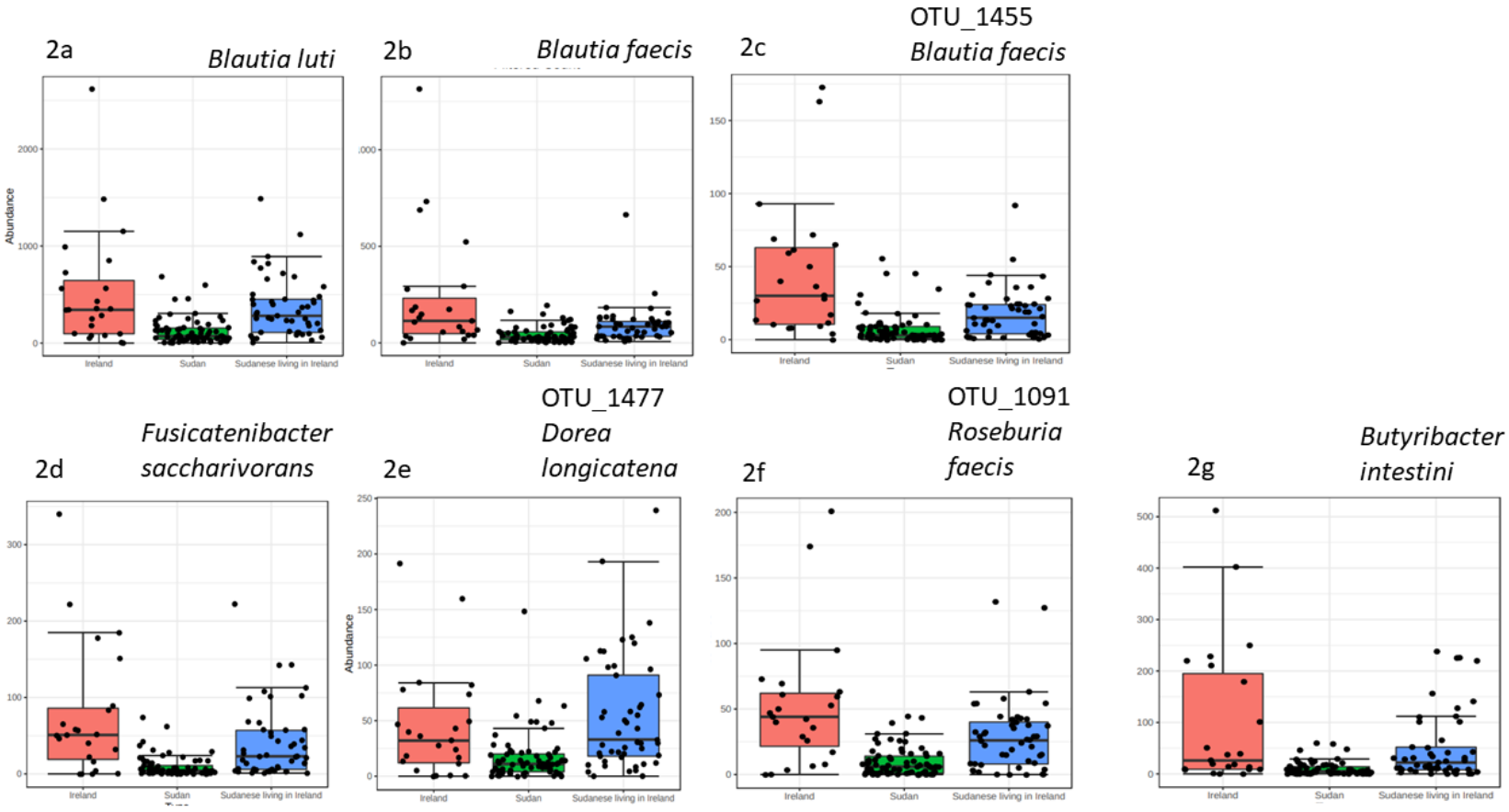
Species associated with the *Lachnospiraceae* family and found to be low in Sud included the *Blautia* species, *luti* [98% ident] [Figure 2a] and *faecis* [Figure 2b and 2c] and *Fusicatenibacter saccharivorans* [OTU_157, 96% ident], [q= 2.5397e-6] [Figure 2d]. Sequences in similarity to *Dorea longicatena [*OTU 1477, >95% ident] [Figure 2e], and *Roseburia faecis [*OTU 1091, >96% ident] [q=3.3873e-7] [Figure 2f] were also low in Sud. *Butyribacter intestini* [OTU_343, 99% ident], [q=1.2864e-5] was also found to be virtually absent in Sud but were more relatively abundant in SudIrel and Irel [Figure 2g].

Ruminococcaceae was the next most abundant family, equal in abundance amongst all groups [~ 24.5%]. *Bifidobacteriaceae* was similar in relative abundance in Sud and SudIrel [~ 8.5%] but were least abundant in Irel [5%]. *Bacteroidaceae*, on the other hand, was most enriched in Irel [10%] compared to [6%] and [3%] in SudIrel and Sud respectively. Sud and SudIrel groups also had similar abundances of *Prevotellaceae* [7%], unlike [2%] in the Irel group. *Lactobacillaceae* [3%], *Coriobacteriaceae* [4%] and *Erysipelotrichaceae* [6%] were enriched in Sud.

We utilised one-way Anova to detect significant genera variations between groups. Those microbiome compositions reduced in Irel but in high relative abundance in the other groups, included *Prevotella* [q=3.169e-06], *Alloprevotella* [q= 1.463e-02], *Megasphaera [q=4.61e-05], Holdemanella* [q=4.461e-05], *Intestimonas* [q=6.243e-03] and *Lactobacillus* [q=2.918e-09] [Figure 3]. *Bifidobacterium* was enriched in Sud and SudIrel, [q=7.023e-03], while *Weissella* was enriched in Sud [q=7.120e-07] compared to Irel. *Prevotella* was retained and enriched in SudIrel [52%], followed by Sud [41%], while the Irel group exhibited only [7%] with *Prevotella copri* [*Prevotella_9*, q = 2.7708e-9], the most abundant species. *Streptococcus* was also found to be abundant in the Sudanese [34-56%] compared to the Irel cohort [9.3%]. In this study, the Irel group had a much lesser relative abundance [2.8%] of *Bifidobacterium* than both Sudanese groups [23-73%]. *Clostridium cluster 1* was

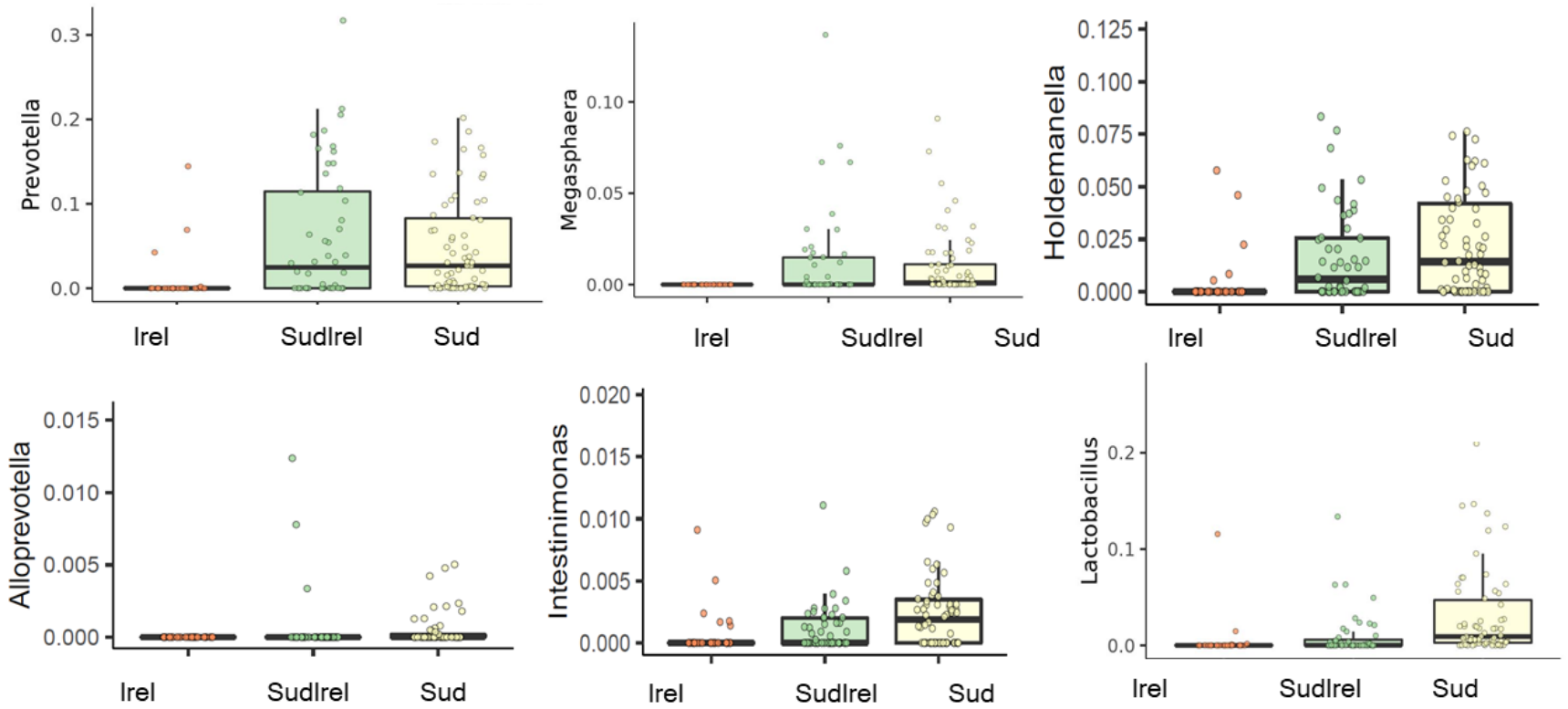
found to be in highest relative abundance in SudIrel [56%] compared to just over 20% in Sud and Irel.

Figure 2:


2a
Blautia luti
2b
Blautia faecis
2c
OTU_1455
Blautia faecis
Abundance
Ireland
Sudan
Sudanese living in Ireland
2d
Fusicatenibacter saccharivorans
2e
OTU_1477
Dorea longicatena
2f
OTU_1091
Roseburia faecis
2g
Butyribacter intestini

2: The further following of Lachnospiraceae associated species with reduced abundance in Sud. Species 2a: Blautia luti, 2b: Blautia faecis, 2d: Fusicatenibacter saccharivorans, and 2g: Butyribacter intestini shows greater than 98% identity utilising BLASTn sequence identification while 2c: OTU_1455, 2e: OTU_1477, and OTU_1091, were similar in sequence to associated species and may represent similar species.

Figure 3:



3: Depletions of genera within Irel. One-way Anova annotations between groups. Prevotella: q=3.169e-06, coefficient: 1.8e+00, Alloprevotella: q=1.463e-02, coefficient: 4.76e-01, Megasphaera: q=4.61e-05, coefficient: 1.45e=00, Intestimonas: q=6.243e-03, coefficient: 6.00e-01, Holdemanella: q= 4.461e-05, coefficient: 1.56e+00 and Lactobacillus: q=2.918e-09, coefficient 1.94e+00.

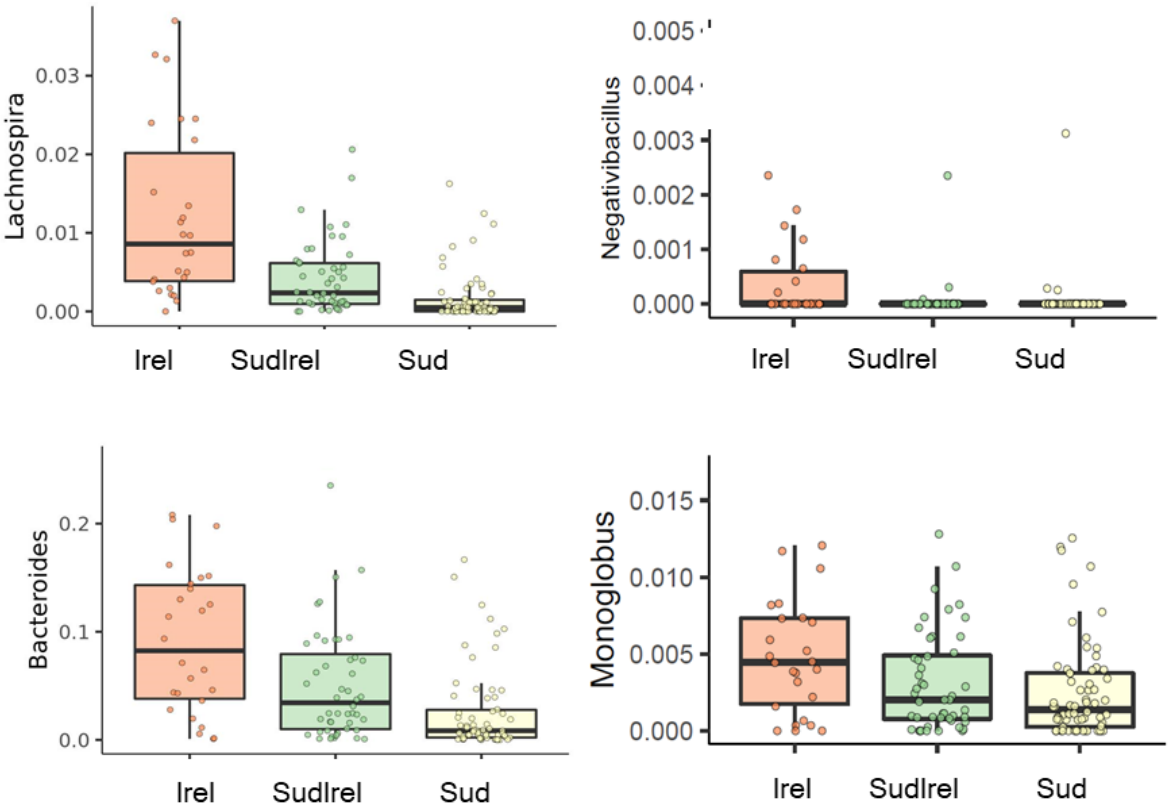
Similar sequences to *Clostridium leptum [cluster IV]* species could also be observed to be depleted in Sud [q= 3.654e-5, 94% ident]. *Ruminococcus_1* [q=1.6503e-4] was found to be least abundant in Sud. *Subdoligranulum* [q= 5.6124e-4] was found to be enriched in Irel compared to the other cohorts.

Several genera were found to be enriched in Irel, of which *Bacteroides* [q=3.056e-05], *Lachnospira* [q=1.623e-09], *Negativibacillus* [q=5.73-e-05] and *Monoglobus* [q=4.199e-02] [Figure 4] were in high relative abundance. *Bacteroides* abundance in Sud was found to be 15%, 38% in Irel and 46% in SudIrel. Low in all groups were the genera *Enterococcus* [q=1.191e-04] and *Megamonas* [q=1.132e-02]. Table 5 highlights other genera relative abundances between the three cohorts after p-adjusted corrections.

Several *Eubacterium* species were identified within the groups. *Eubacterium eligens* [q = 3.389E-6] was found to be low in Sud [99% identity] and *Eubacterium ramulus*. *Eubacterium ventrosium* [q=1.6972-e] and *Eubacterium coprostanoligenes* [q= 0.003], on the other hand, were similar in abundance amongst all groups, while *Eubacterium hallii* was the species in highest relative abundance and composed [7-9%] of the gut microbiome [Figure 5].

In SudIrel, Clostridium cluster 1 [*clostridium sensu stricto*] was most abundant [56%] compared to Sud and Irel [20%]. Similar sequences to *Clostridium leptum* [cluster IV] could be observed to be low in the Sud group. Lower relative abundances of *Faecalibacterium* could also be observed in Irel [24%], while Sud [31%] and SudIrel [44%] had higher relative abundances [q = 1.2206e-4]. *Anaerostipes* [q = 0.0016] was found to be enriched in SudIrel [50%] and Sud [32%] compared to Irel [only present at 16%].

Figure 4:


Lachnospira
0.03
0.02
0.01
0.00
Irel
SudIrel
Sud
Negativibacillus
0.005
0.004
0.003
0.002
0.001
0.000
Irel
SudIrel
Sud
Bacteroides
0.2
0.1
0.0
Irel
SudIrel
Sud
Monoglobus
0.015
0.010
0.005
0.000
Irel
SudIrel
Sud

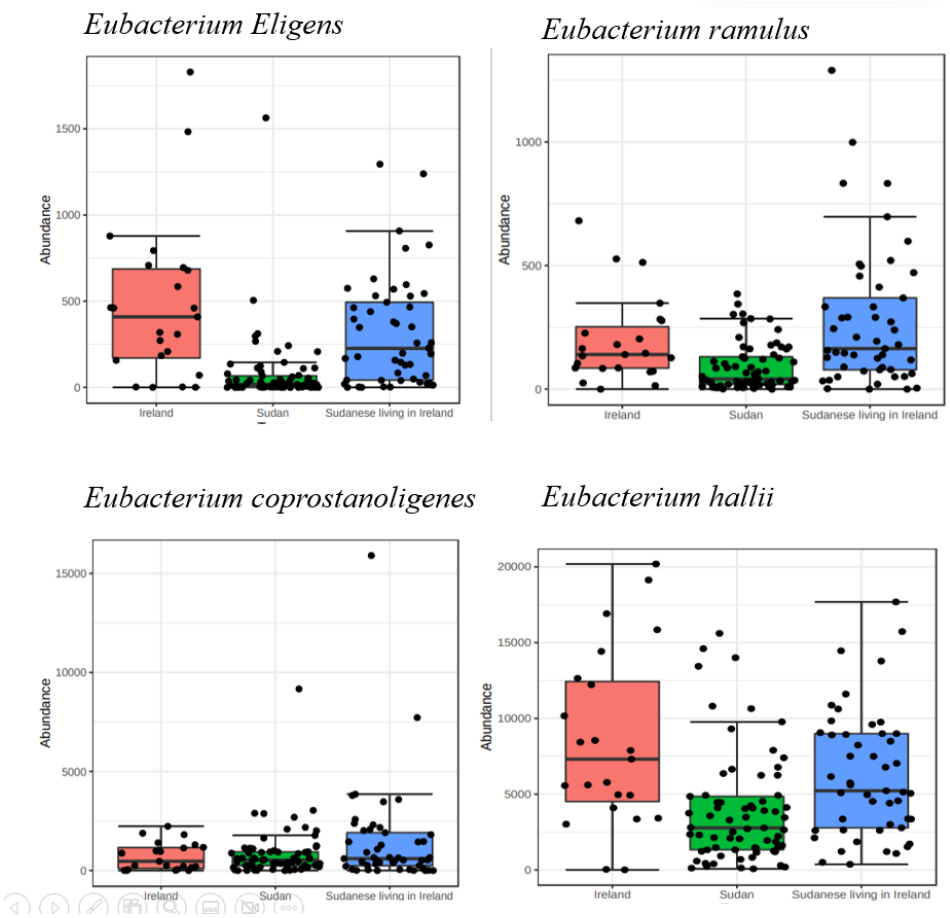
4: Enrichments in microbiome genera in Irel. Lachnospira: q=1.623e-09, coefficient 1.3e+00, Bacteroides: q=3.056e-05, coefficient: 9.05e-01, Negativibacillus: q=5.730e-05, coefficient: 4.40e-01 and Monoglobus; q=4.199e-02, coefficient 4.77e-01

*Table 5: Genera with significantly associated p- adjusted values in the three cohorts, Sud, SudIrel and Irel. Seven genera were associated with Sud and SudIrel. Sud exhibited the greatest increased stand-alone genera significance [seven genera] that were found to be uniquely either high or low in Sud. *Alistipes finegoldii [99% ident] was the species identified and found to be depleted in the Sud group with an increased abundance in SudIrel with highest interquartile ranges found in Irel.*

Genus	**Sud**	**SudIrel**	**Irel**
Slackia	High [q=6.243e-03]		
Oribacterium	High [q=1.753e-02]		
Methanobrevibacter	High [q=2.249e-02]		
Senegalimassilia	High [q=8.396e-03]		
Catenibacterium	High [q=1.229e-02]		
Fournierella	High [q=3.218e-02]		
Mitsuokella	High [q=3.937e-02]		
Butyricimonas	High [q=4.169e-02]		
Sutterella	Low [q=1.266e-02]		
Parasutterella		High [q=1.451e-02]	
Gordonibacter			High [q=1.443e-04]
Enterorhabdus	High [q=3.056e-05]		

Klebsiella	High [q=3.9067-e]		
Subdoligranulum	Low [q=5.6124e-4]		
Ruminiclostridium_5			High [q=0.002]
Odoribacter	Low [q=1.372e-5]		
*Alistipes**	Low [q=1.3425e]		

Figures 5:



5: Eubacterium species identified in this study included eligens, ramulus, coprostanoligenes and hallii. Eubacterium hallii was the most enriched species, albeit still lower in abundance in Sud. Eubacterium coporostanoligenes was similar in abundance throughout the cohorts.

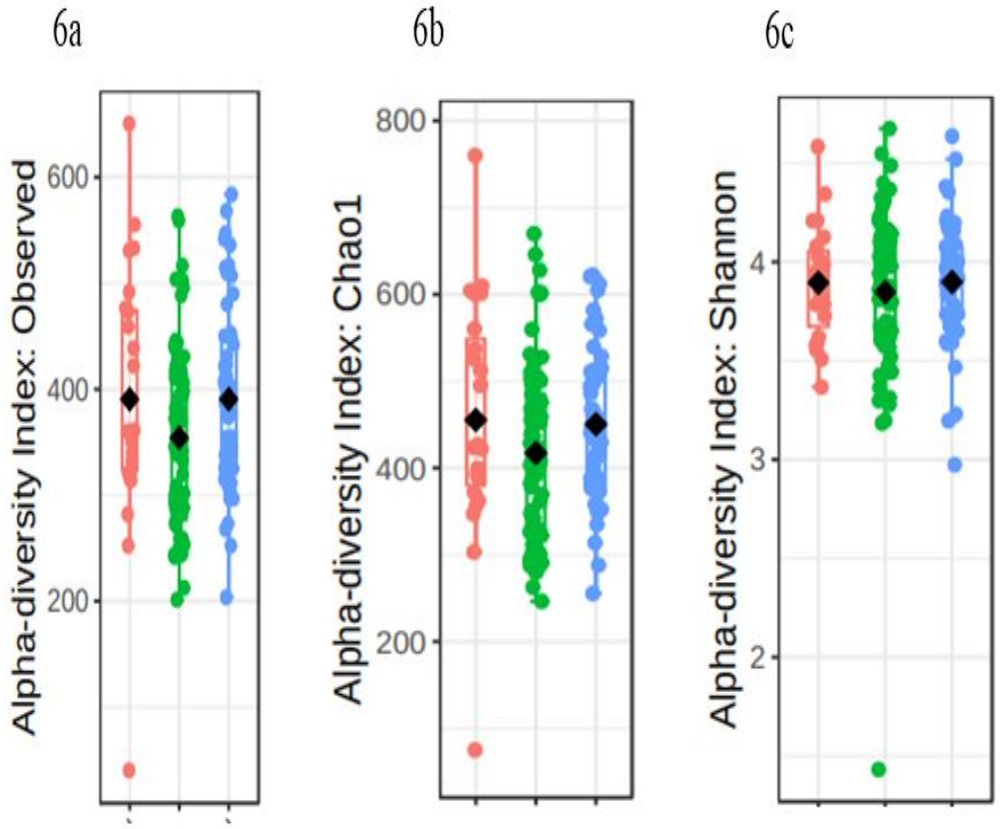
Akkermansia, on the other hand was found to be nearly doubled in SudIrel [43%] compared to the Irel and Sud groups [~27%], while *Collinsella* was four times more abundant in Sud and SudIrel [44%] compared to Irel [11%] [q=2.766e-5]. *Holdemanella* [q=461e-05] was enriched in Sud and SudIrel and *Dialister* was also abundant in the Sudanese groups [41%] compared to 17% in Irel [q = 1.347e-12]. However, *Dialister invisus* [OTU_28, 99.35% ident] was absent from the Sud group compared to those living in Ireland. The genus *Klebsiella* was only abundant in Sud [q=3.9067-e]. *Methanobrevibacter* [q=2.249e-02] was also found to be significantly abundant in Sud.

Bilophila [q=4.4426e-4], *Odoribacter* [q= 1.372e-5] and *Alistipes* [q=1.3425e] were found to be least abundant in Sud with *Alistipes finegoldii* [OTU 164, 99% ident], the species identified. A similar trend for *Alistipes timonensis* [OTU 1697, 97% ident] was also identified. *Ruminiclostridium_5* was low in Sud and SudIrel compared to Irel [q =0.004], and *Ruminococcus_1* was found to be least abundant in Sud [q= 1.6503e-4] with the species, *Ruminococcus faecis* [OTU 977, >96% ident] [q=3.65011e-5] identified. *Parasutterella* was also found to be low in Sud [q=1.6348e-4].

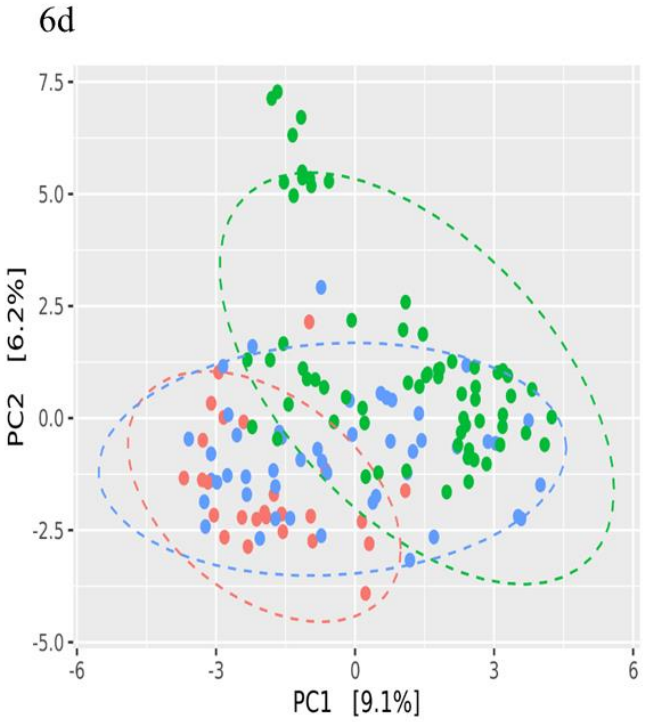
Alpha and Beta diversity

Alpha diversities [alpha diversity index] were found not to be significantly different between groups but were found to be high. Observed species or richness was found to be >200 variety indices [p= 0.08] [Figure 6a]. Chao's index measurements were found to be between 200-800 [p= 0.15], further appreciating richness including rare species [Figure 6b] while Shannon diversity index or evenness, [p= 0.74] was also found to be high [3-4.5] [Figure 6c][17]. β diversity highlighted a clear separated distinction between Irel and Sud groups, with SudIrel 'sandwiched' in between the other groups. β diversity was analysed through principal co-ordinate analysis using unweighted UniFrac distance and was found to be significant [p< 0.001] [Figure 6d].

Figures 6a-6c:



*6: Non-significant **α** variations were generally observed. Pink=Ireland, green=Sudan and Blue=Sudanese living in Ireland. 6a: Richness [observed], P=0.08. 6b: Chao1, p=0.15, and 6c: Shannon diversity, p=0.74. However, a high **α** diversity was evident in all models. Chao1 was used to take into consideration any low abundance taxa and the independency of likelihood of presence of taxa in the gut microbiome community using Poisson distribution. The line inside the boxes represents the median while each box represents the upper and lower interquartile range with circles being outliers.*



6d. Elliptical PCoA graph of β diversity on the other hand was significantly varied between groups [p<0.001] with 'sandwiching' of SudIrel in relation to the remaining study cohorts

Nutrient-genera correlations

Correlation plotting highlighted several significant nutrient-genera associations. This included alcohol and the genus *Parabacteroides* in SudIrel [q=1.31e-11], Galactose and *Coprococcus* in Irel [q=1.71e-06] and *Monoglobus* in SudIrel [q=0.006]. In Sud, the family *Lachnospiraceae* was positively correlated with iodine [q=0.004] [Figure 7a]. The low fibre intake associated with Sud was significantly correlated with *Eubacterium eligens* [q=0.04]. *Erysipelotrichaceae* was positively associated with Sud and lactose [q=0.005] [Figure 7b]. *Megamonas* was found to be the most positively correlated with key nutrients and specifically with the cohort Sud. It was significantly associated with calcium [q=0.001], magnesium [q=0.001], phosphorus [q=0.0002], potassium [q=0.02], protein [q=0.004], vitamin B2 [q=0.002], vitamin B5 [q=0.01], vitamin E [q=0.03], total fats [q=0.006] and saturated fats [q=0.005] [Figure 7a-7c].

Vitamin B5 was positively correlated with three different genera depending on cohort and included, *Monoglobus* for SudIrel [q=0.01], *Megamonas* for Sud and *Coprococcus* for Irel [q=0.01]. *Collinsella* was positively correlated with the cohort Sud and Vitamin D [q=0.04]. In contrast, zinc was positively correlated with four distinct genera that included *Ruminococcus* in Irel [q=0.04], *Phascolarctobacterium* in SudIrel [q=0.01] and *Lactobacillus* [q=0.04] and *Collinsella* [q=0.04] in Sud [Figure 7c].

Figure 7a:

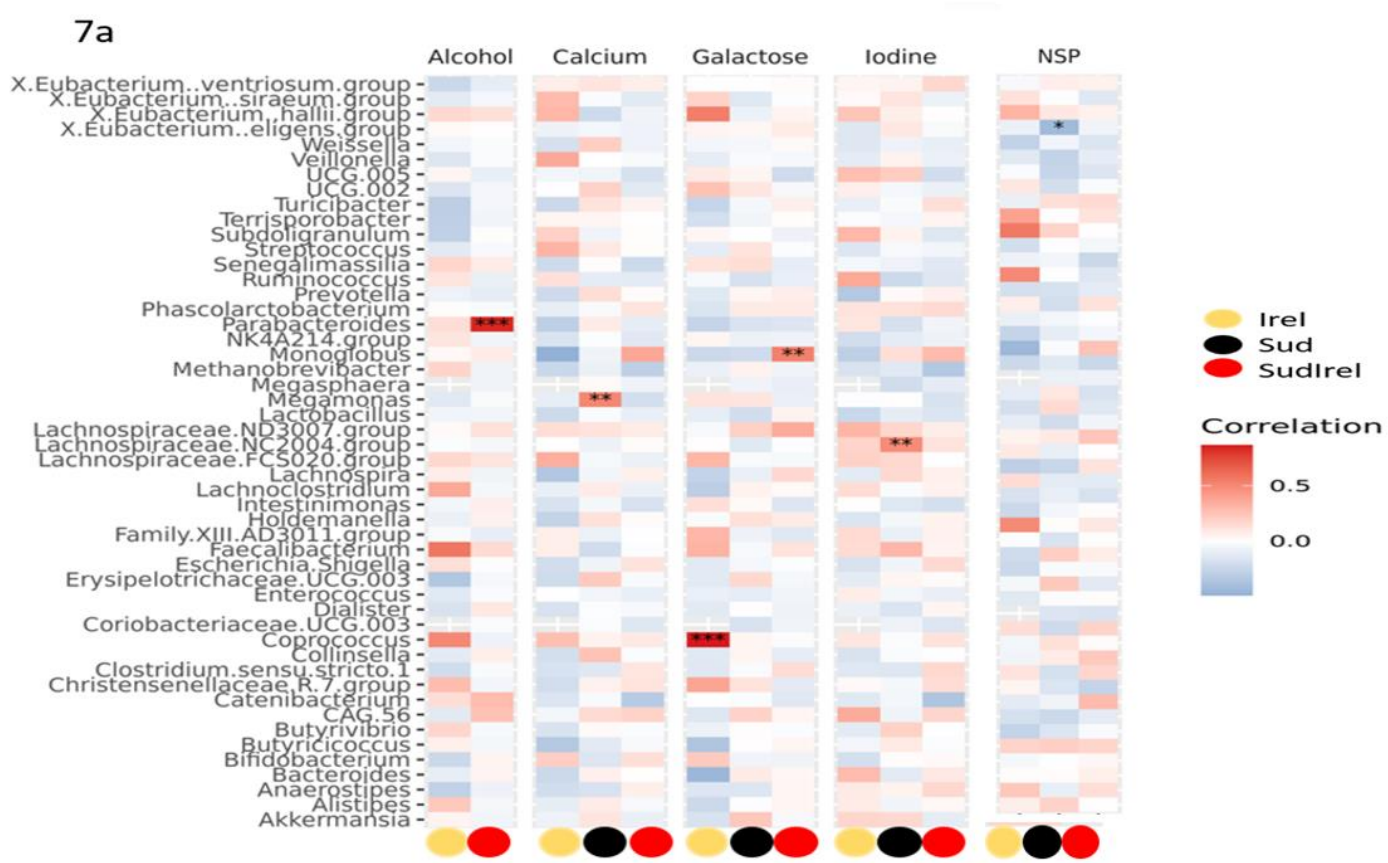

7a
Alcohol
Calcium
Galactose
Iodine
NSP
X.Eubacterium..ventriosum.group
X.Eubacterium..siraeum.group
X.Eubacterium..hallii.group
X.Eubacterium..eligens.group
Weissella
Veillonella
UCG.005
UCG.002
Turicibacter
Terrisporobacter
Subdoligranulum
Streptococcus
Senegalimassilia
Ruminococcus
Prevotella
Phascolarctobacterium
Parabacteroides
NK4A214.group
Monoglobus
Methanobrevibacter
Megasphaera
Megamonas
Lactobacillus
Lachnospiraceae.ND3007.group
Lachnospiraceae.NC2004.group
Lachnospiraceae.FCS020.group
Lachnospira
Lachnoclostridium
Intestinimonas
Holdemanella
Family.XIII.AD3011.group
Faecalibacterium
Escherichia.Shigella
Erysipelotrichaceae.UCG.003
Enterococcus
Dialister
Coriobacteriaceae.UCG.003
Coprococcus
Collinsella
Clostridium.sensu.stricto.1
Christensenellaceae.R.7.group
Catenibacterium
CAG.56
Butyrivibrio
Butyricicoccus
Bifidobacterium
Bacteroides
Anaerostipes
Alistipes
Akkermansia
Irel
Sud
SudIrel
Correlation
0.5
0.0

Figure 7b:

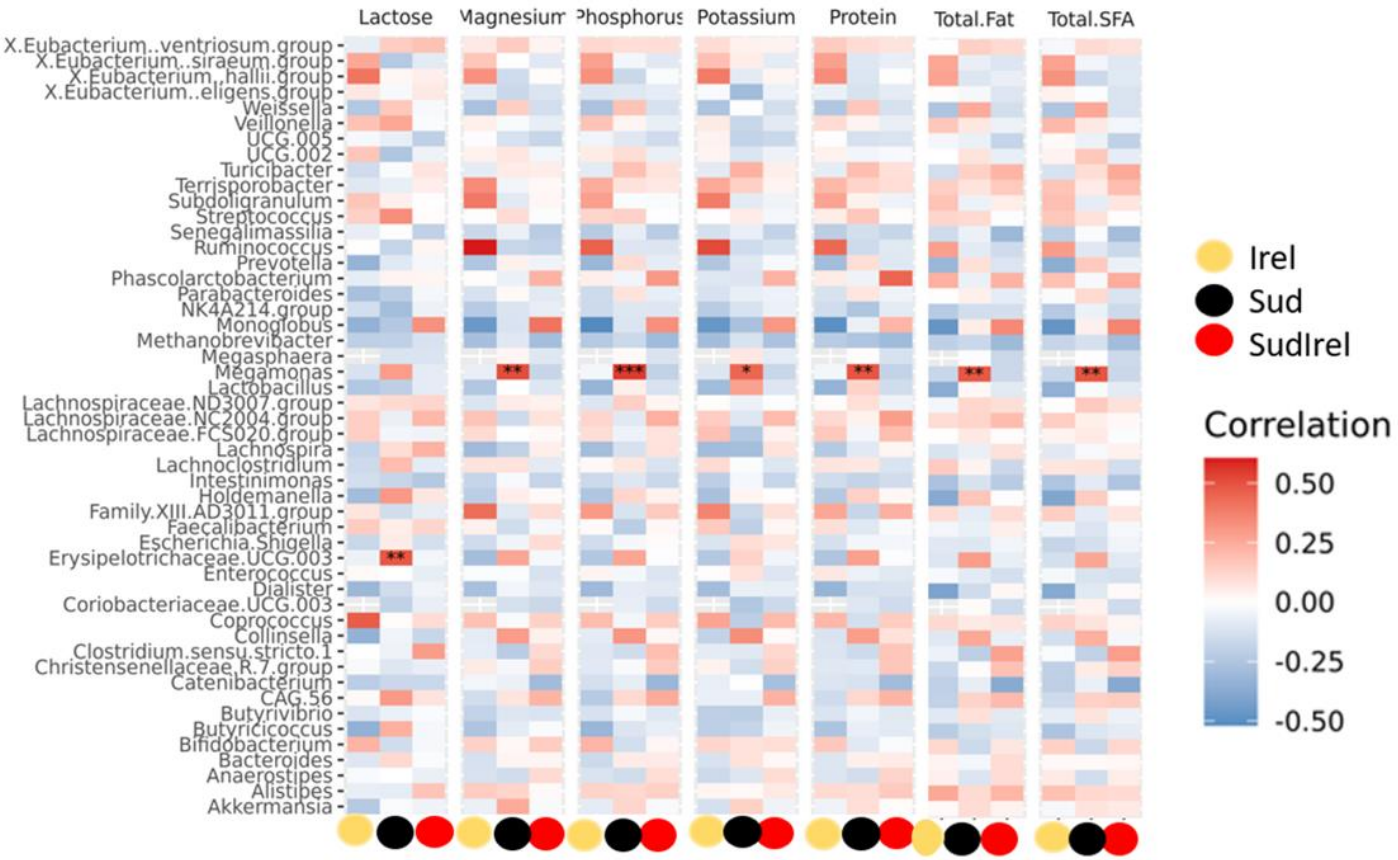
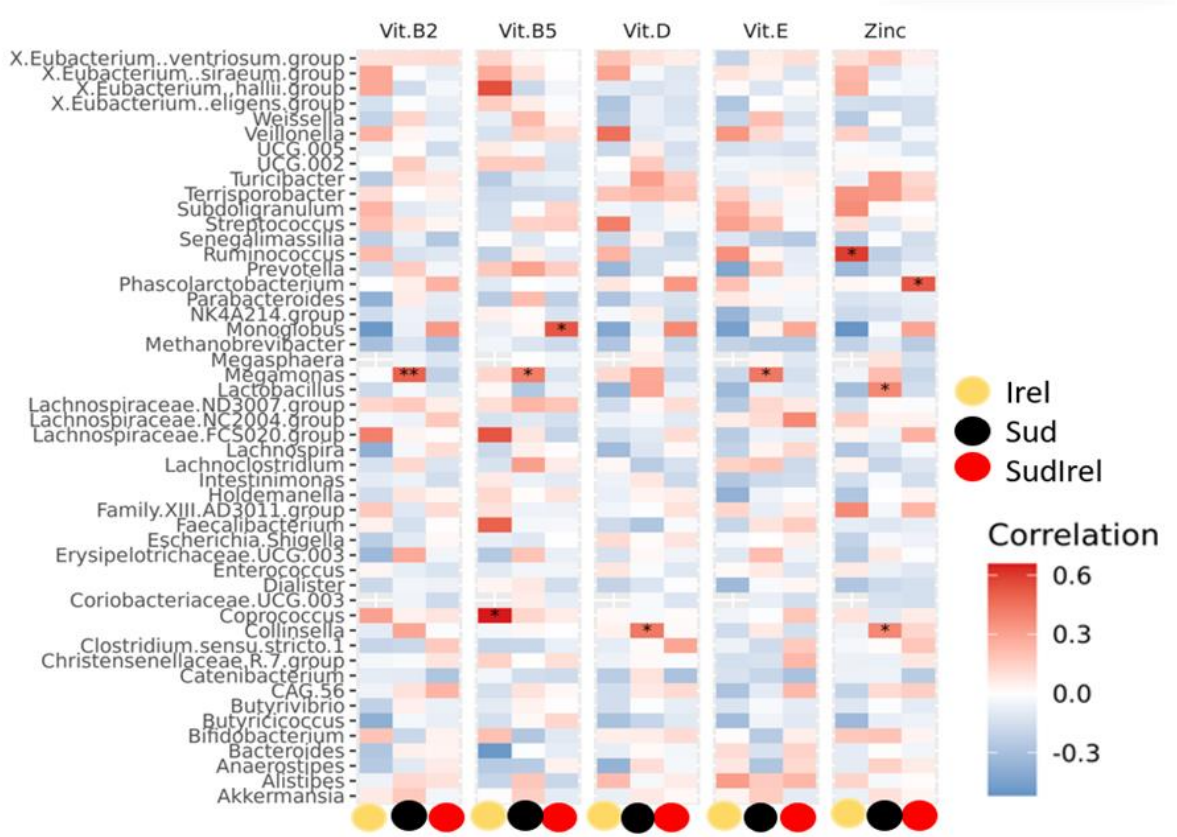

Lactose
Magnesium
Phosphorus
Potassium
Protein
Total.Fat
Total.SFA
X.Eubacterium..ventriosum.group
X.Eubacterium..siraeum.group
X.Eubacterium..hallii.group
X.Eubacterium..eligens.group
Weissella
Veillonella
UCG.005
UCG.002
Turicibacter
Terrisporobacter
Subdoligranulum
Streptococcus
Senegalimassilia
Ruminococcus
Prevotella
Phascolarctobacterium
Parabacteroides
NK4A214.group
Monoglobus
Methanobrevibacter
Megasphaera
Megamonas
Lactobacillus
Lachnospiraceae.ND3007.group
Lachnospiraceae.NC2004.group
Lachnospiraceae.FCS020.group
Lachnospira
Lachnoclostridium
Intestinimonas
Holdemanella
Family.XIII.AD3011.group
Faecalibacterium
Escherichia.Shigella
Erysipelotrichaceae.UCG.003
Enterococcus
Dialister
Coriobacteriaceae.UCG.003
Coprococcus
Collinsella
Clostridium.sensu.stricto.1
Christensenellaceae.R.7.group
Catenibacterium
CAG.56
Butyrivibrio
Butyricicoccus
Bifidobacterium
Bacteroides
Anaerostipes
Alistipes
Akkermansia
Irel
Sud
SudIrel
Correlation
0.50
0.25
0.00
-0.25
-0.50

Figure 7c:


Vit.B2
Vit.B5
Vit.D
Vit.E
Zinc
X.Eubacterium..ventriosum.group
X.Eubacterium..siraeum.group
X.Eubacterium..hallii.group
X.Eubacterium..eligens.group
Weissella
Veillonella
UCG.005
UCG.002
Turicibacter
Terrisporobacter
Subdoligranulum
Streptococcus
Senegalimassilia
Ruminococcus
Prevotella
Phascolarctobacterium
Parabacteroides
NK4A214.group
Monoglobus
Methanobrevibacter
Megasphaera
Megamonas
Lactobacillus
Lachnospiraceae.ND3007.group
Lachnospiraceae.NC2004.group
Lachnospiraceae.FCS020.group
Lachnospira
Lachnoclostridium
Intestinimonas
Holdemanella
Family.XIII.AD3011.group
Faecalibacterium
Escherichia.Shigella
Erysipelotrichaceae.UCG.003
Enterococcus
Dialister
Coriobacteriaceae.UCG.003
Coprococcus
Collinsella
Clostridium.sensu.stricto.1
Christensenellaceae.R.7.group
Catenibacterium
CAG.56
Butyrivibrio
Butyricicoccus
Bifidobacterium
Bacteroides
Anaerostipes
Alistipes
Akkermansia
Irel
Sud
SudIrel
Correlation
0.6
0.3
0.0
-0.3

7. Nutrient genera correlation plotting of top 50 genera in Sud, SudIrel and Irel. Although Lachnospiraceae was found to be decreased in Sud, they were positively associated with iodine, known to be endemically deficient in Sudan [7a]. Megamonas was the most positively associated nutrient-genus group in Sud [7a-7c]. The low fibre intake in Sud may also be associated with the low abundance of Eubacterium eligens [7a].

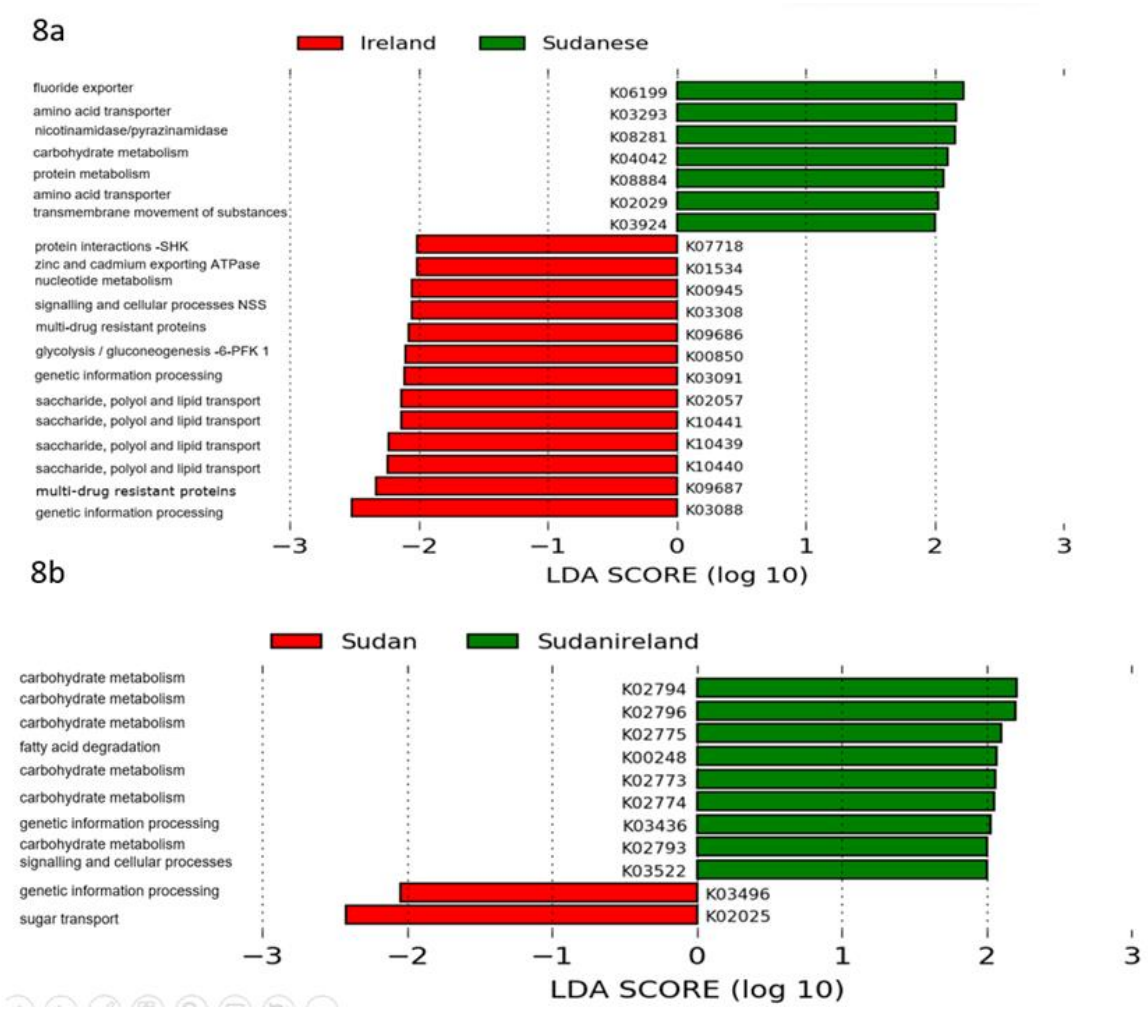
Phylogenetic Investigation of Communities by Reconstruction of Unobserved States [PICRUSt]

LEFSe plotting highlighted important differentiations in the KEGG orthology pathways between the Sudanese [Sud and SudIrel] and Irel groups [Figure 8a]. The most prominent pathways in the Sudanese were K06199 [fluoride exporter], K03293 [amino acid transporter], K08281 [nicotinamidase/pyrazinamidase], K04042 [carbohydrate metabolism - bifunctional UDP-N-acetylglucosamine pyrophosphorylase / glucosamine-1-phosphate N-acetyltransferase], K08884 [metabolism of protein - serine/threonine protein kinase], K02029 [amino acid transport] and K03924 [metabolism – MoxR-like ATPase involved in the transmembrane movement of substances].

In Irel, prominent pathways were K03088 [genetic information processing – RNA polymerase sigma – 70 factor], K09686, K09687, K10440, K10439, K10441 and K02057 [saccharide, polyol, and lipid transport – ribose transport system].

We further distinguished specific variations between the Sud and SudIrel groups [Figure 8b]. SudIrel exhibited prominent KEGG pathways related to carbohydrate metabolism [K02793, K02794, K02796, K02773, K02774, K02775] and fatty acid degradation [K00248]. Others included K03436 [genetic information processing] and K03522 [signalling and cellular processes]. In Sud, two pathways were of important variation from SudIrel including K03496 [genetic information processing associated with cytoskeleton proteins] and K02025 [sugar transport system permease protein].

Figures 8a - 8b:



PICRUSt data 8a: Comparisons between Irel, Sud and SudIrel, highlighted the Sudanese having abundant amino acid exportation, fluoride [K06199], and nicotinamidase/pyrazinamidase pathways [K08281], than Irel with multiple saccharide, polyol and lipid transport pathways and multidrug resistance KEGG orthology pathways [K09687 and K09686]. 8b: Comparisons between those living in Sudan [Sud] and those Sudanese living in Ireland [SudIrel]. Those in the SudIrel group exhibited multiple carbohydrate metabolism pathways while those in Sud harboured sugar transport [K02025], and genetic information processing [K03496], as discriminant pathways.

Metabolomic findings

A total of 4177 compounds were detected in the samples. Hereof, 98 metabolites could be identified [level 1]. Further, 83 compounds could be matched to library compounds over their accurate mass and retention time [level 2a]. Additionally, 153 compounds were annotated by accurate mass and matched to MS/MS fragmentation reference spectra [level 2b]. 1222 compounds were annotated by exact mass, isotope pattern and reference count [level 3]. Finally, 2621 compounds were distinguished from the background but remain unknown. The score plot from the principal co-ordinate analysis [PCA] model calculated on the compounds annotated on levels 1 and 2a in the reduced dataset can be found in Figure 9a.

Metabolomic compound variations between Irel, SudIrel and Sud

We found 27 metabolites significantly varied between Irel and Sud, 21 between Sud and SudIrel and none between Irel and SudIrel. These are presented in Table 6.

Table 6: Metabolites with significant variations comparing Irel and Sud and Sud and SudIrel. No significant variations were found between Irel and SudIrel.

Compound	**Irel/Sud [p value]**	**Sud/SudIrel [p value]**
Proline	0.003	0.001
Threonine		0.002
Adipic acid	0.04	0.005
Sebacic acid		0.007
Urocanic acid	0.01	0.008

N-Acetylneuraminic acid		0.01
Serine	0.01	0.01
Dodocanedioic acid	0.008	0.01
Tyrosine		0.01
Glycine	0.01	0.01
Azelaic acid		0.02
Glutamic acid	0.02	
4-Guanidinobutyric acid	0.001	0.03
Isoleucine	0.01	0.03
2-Methylbutyrylglycine	0.009	0.03
2-Aminoadipic acid		0.04
Shikimic acid		0.04
Lysine	0.02	0.04
Sugar alcohol		0.04
4-hydroxy-3-methoxymandelic acid		0.05
Xanthine	0.001	
Cotinine	0.013	
Aspartic acid	0.005	
Uracil	0.002	
Thymine	0.01	
Citrulline	0.01	
Ornithine	0.02	
Phenylalanine	0.02	
Hypoxanthine	0.013	
dinosine	0.01	
Phosphocholine	0.02	
Cytosine	0.02	
dUridine	0.009	
3-Methylbutanoic acid	0.004	
Thymidine	0.01	

N, N Hexamethylene bis[acetamide]	0.03	

The amino acids proline, threonine, serine, glycine, Isoleucine, lysine, leucine, thymine, glutamine, alanine, methionine, citrulline, histidine, ornithine, and glutamic acid and the aromatic amino acids, phenylalanine, tyrosine, and tryptophan, were all elevated in SudIrel. Arginine was found to be high in Sud, theanine low in Irel, and taurine was low in Irel and Sud. Tryptophan metabolites included Kynurenine and Kynurenic acid and were found to be high in Sud and SudIrel, while tryptamine was high in SudIrel.

Those acids found to be high in SudIrel were adipic, uric, sebacic, 2 hydroxyisocaproic, 2-hydroxyphenylacetic, hippuric, 4,8-dihydroxyquinoline-2-carboxylic, pyroglutamic, urocanic, acetylmuramic, suberic, aspartic, gluconic, mucic, malic, traumatic, and alpha aminobutyric acids. Those metabolites high in SudIrel included n-acetyl-ornithine, xanthine, n-acetylmannosamine, cotinine, maltotriose, phenyl-ac-gln-oh-2, n-acetyl-5-hydroxytryptamine, cytosine and pyridoxine. Bile acids and secondary metabolites included cholic acid, a and b-muricholic, deoxycholic, hyodeoxycholic and were found to be high in Sud.

Threonic [sugar acid], diaminopimelic, glyceric and 5-aminovaleric were also found to be high in Sud. The acids glutaric, 3,4-dihydroxybenzoic, and pantothenic acids and the metabolites, allopurinol, uridine, inosine, n-acetyl-methionine, xanthosine, 3-methylhistidine and methyl-3-indoleacetate were low in Sud and Irel. The acids, n-acetyl-neuraminic [an antioxidant and bacterial metabolite], dodocanedioic, azelaic, 2-aminoadipic, 3-methyladipic and shikimic acid were elevated in all groups. Phosphocholine was increased in the Sud group compared to Irel [$\log^2$ ratio=5.3] and SudIrel [$\log^2$ ratio=0.8]. Choline was similar between the groups.

Quinic, succinic [carbohydrate fermentation], β-d-glucopyranuronic, n-acetyl-glutamic, 3,4-dihydroxyphenylacetic, 2-oxoglutaric, methanesulfonic, 2-oxo-3-phenylpropanoic and 4-hydroxy-3-methoxymandelic acids were high in both Sud and SudIrel, while the metabolites, carnitine, glycyl-valine, caffeine, theobromine, hexose, theophylline, 2 – deoxyglucose, creatine, artemisinin, galactosamine and 4-

hydroxyproline were also high in both Sud and SudIrel. The acids a-muricholic, gallic, cis-aconitic, 2,5-dihydroxybenzoic, deoxycholic, hyodeoxycholic and homovanillic acid, and the metabolites, anserine, carnosine, leucyl-alanine, and n-acetyl-tyrosine were high in Sud compared to Irel but low when compared to SudIrel.

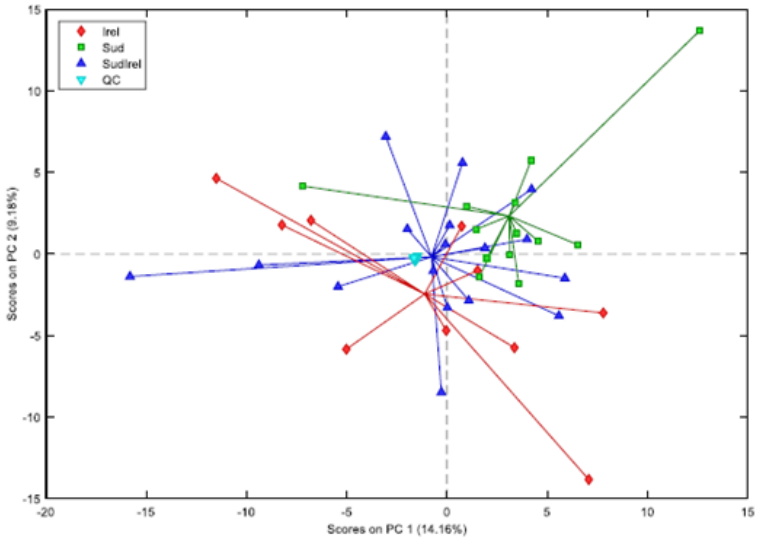
The acids n-acetyl-aspartic, 3-methylbutanoic, 2-hydroxybutyric, fumaric and picolinic acids, and metabolites, 4-acetamidophenol, trigonelline, glycyl-glycine, 4-vinylguaiacol, glycyl-leucine, d-uridine, n-acetyl-phenylalanine, GMP, thymidine, 5-methyltryptophan and N, N′-hexamethylene bis[acetamide] were all low in Irel. N8-acetylspermidine was found to be high in Sud and SudIrel. Trimethylamine–N-Oxide was higher in logfold change between Sud and SudIrel [log^2 ratio = 1.5] compared to Irel and SudIrel [log^2 ratio= 0.2]. 4 vinylphenol was high in Sud and SudIrel compared to Irel with log^2 ratios of 4 and 1.9, respectively.

Short-chain fatty acid results

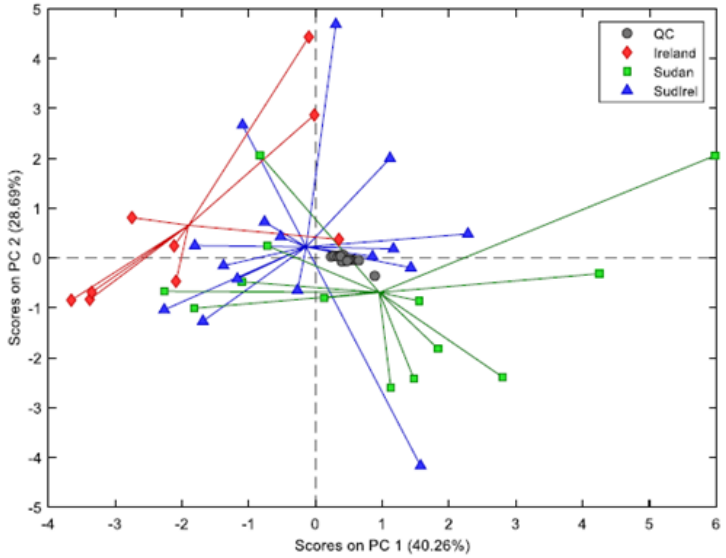
Isobutyric [log^2 ratio = -0.8], isovaleric [log^2 ratio = -1], valeric [log^2 ratio = -1.1], propionic [log^2 ratio = -0.8] and heptanoic [log^2 ratio = -0.9] acids were lower in Irel compared to Sud and SudIrel. Isobutyric [p=0.0], isovaleric [0.002] and valeric [p=0.03] acids, were found to be significantly varied between Irel and Sud groups. Isocaproic acid was low in both Irel [log^2 ratio = -0.7] and Sud [log^2 ratio = -0.5] compared to SudIrel. Formic [log^2 ratio = 0.2] and butyric acids [log^2 ratio = 0.3] were found to be higher in Sud, with formic acid being significant [p=0.03]. Hexanoic acid was highest in SudIrel [log^2 ratio = 0.9]. A score plot of the relative concentrations of ten short-chain fatty acids is presented in Figure 9b.

Figures 9a - 9b:

9a



9b



Green squares = Sud, dark blue triangles = SudIrel and red diamonds = Irel. Quality control samples are indicated by light blue reversed triangles. 9a: Score plot from a PCA model calculated on the compounds annotated on levels 1 and 2a in the reduced dataset. In this plot, the quality control samples are tightly grouped indicating that the analytical variance [represented by the spread of the quality control samples] is greatly exceeded by the biological variance [represented by the spread of the samples]. The quality control samples are located close to the centre of the plot indicating that matrix effects are not an issue with these samples. 9b: Score plot of the relative concentrations of ten short chain fatty acids. Isobutyric [p=0.0], isovaleric [p=0.002] and valeric [p=0.03] acids, were found to be significantly varied between Irel and Sud groups. 'Sandwiching' effect of SudIrel is continued in metabolomic findings where expression of functional activity resembles both traditional and resculpted microbiome gut parameters.

5.5 Discussion

Here, we report on how gut microbiome, metabolome and dietary patterns of the Sudanese population relate to traditional [Sud] and westernised [SudIrel] settings in comparison with Irish controls and how the modulatory effects through habitat and diet may offer a significant variation on either susceptibility or protection against the development and progression of several disorders and nutrient profiles.

Dietary intakes were significantly varied, with five and two food groups found to be consumed at lowest and highest rates in Sud, respectively, compared to the remaining groups. A low energy intake [kcal] was associated with Sud, i.e., living in Sudan, compared to Irel and SudIrel, i.e., living in Ireland. Carbohydrates were found to be consumed less in Sud, including fibre, while cholesterol, fat and protein levels were found to be similar among those living in Ireland [Irel and SudIrel]. Other nutrients such as calcium and iron were increased in SudIrel while vitamin B5 was highest in mean intake in Sud. These findings highlight that those living in Sudan have a different dietary landscape that may be predominantly healthier in aspects and less nourished in others compared to those who migrate to westernised environments. While African diets tend to promote a higher fibre intake, we found low fibre intake in Sud, likely associated with the urbanisation of our study cohort. Indeed, gut microbiome profiles amongst Africans have been shown to skew towards the favourability of developing type two diabetes [18], although consumption of sorghum and millet-based foods amongst the Sudanese can promote a healthier gut lining and improved microbiota-immune tolerance [19].

The migration of people to westernised countries is said to improve health, although this effect is found to normalise and even weaken over time [20]. Reports suggest that physical activity is often low in migratory groups [21], for example, further promoting the development of non-communicable disease. In North Africans, including the Sudanese, the risk of developing diabetes after migration is high while the risk of developing stroke and coronary heart disease development varies in the literature, with both a reduced [22] and an increased [23] risk reported. In Sudan, there is now an ever-

increasing burden of non-communicable diseases including higher rates of obesity, type two diabetes, renal disease and cancer [24]. In a recent large-scale national survey of 7,722 participants, it was found that nearly one in three Sudanese had more than three or more risk factors for developing of non-communicable disease [25].

Among the gut microbiota, *Firmicutes* was the most relatively abundant phylum in all groups, *Bacteroidetes* were high in Irel, and *Actinobacteria* were found to be high in both Sudanese cohorts, agreeing with the literature [26]. *Proteobacteria* were not evident in SudIrel and have been shown to be the most abundant phylum in an Egyptian cohort [27]. When further appreciating metagenomic findings, we have determined that traditional gut microbiome may remain so for various genera [*Collinsella*]. In contrast other genera become reduced [*Lactobacillus*] or increased [*Bacteroides*] in abundance in relation to migration. In a study by Vangay et al. [2018], gut microbiome westernisation is thought to begin at around 6-9 months in those relocating to the United States [28]. In the SudIrel group, the gut microbiome remained traditional for *Prevotella* and *Bifidobacterium* and was resculpted for *Bacteroides*, most likely in response to the higher fat and carbohydrate intake in the diet after settling into a westernised culture [29].

The study cohort, SudIrel, predominantly exhibited a microbiome [Figure 6d] and metabolome [Figure 9] marked by a 'sandwiched' composition between the other two groups, Sud and Irel, highlighting that migration has an important role in the sculpting of the natural existence of metagenomic findings, and indeed carries a functional role in contributing to human adaptability [30]. After moving to a westernised environment such as Ireland, the gut microbiome of the Sudanese becomes modified, likely following the changes in lifestyle and diet that occur upon migration but also in health, environment, and stress.

Interestingly, *Prevotella* was retained and even enriched in SudIrel [52%]. Other studies have highlighted a loss of *Prevotella* after migration [28] [31], and have primarily linked *Prevotella* abundance with the gut microbiome of those living in rural Africa [32]. *Prevotella* abundance is likely to increase with long-term fibre intake [diets rich in plant-based foods] but is lower in those with animal-based dominated diets [29, 33]. Furthermore, the consumption of traditional foods that are of African origin has been associated with abundance in *Prevotella* [34] [35]. This would indicate

that gut microbiome memory may be retained after moving to Ireland due to the continuation of Sudanese traditional food consumption even outside the traditional setting. However, the increase of *Prevotella* in SudIrel may also suggest that even though this genus may once have been in harmony with the remaining gut microbiome, transfer to westernised environments could create 'a toppling effect', allowing for these bacteria to exist in a presumptuous state. *Alloprevotella* was found only in the Sud and SudIrel groups, a genus that may decrease the levels of triglycerides and may allow for a reduced risk of cardiovascular disease development [36]. *Alloprevotella* abundance has also been associated with the decreased incidence of non-alcoholic fatty liver disease [37].

Furthermore, modern Sudanese gut microbiome was enriched in the species *Prevotella copri* [*Prevotella_9*, q = 2.7708e-9], a multi-clade genus with four distinct groups recognised in the African microbiome compared to only one clade amongst westernised populations [38]. An enrichment of *Prevotella* in the SudIrel group may suggest that differing *Prevotella copri* clades are capable of utilising numerous dietary polysaccharides [39] due to a varied diet intermixed with both African and European identity, which may not have been accessible before migration. Research suggests that *Prevotella* vary in genomic capability, whereby westernised abundance is associated with drug metabolism while non-westernised abundance is related to carbohydrate metabolism [31]. *Prevotella* is both negatively and positively associated with health. Its abundance is particularly linked to the development of inflammatory diseases such as rheumatoid arthritis due to its association with T-helper 1 and T-helper 17 immune cell responses and the release of inflammatory cytokines that include interleukin 6 and 23 [40]. Increased abundance of *Prevotella* has further been associated with the development of insulin resistance and metabolic syndrome [26].

Several studies have developed new understandings of the African gut microbiome landscape. Rubel et al. [2020] noted that the top three gut bacterial genera amongst a Cameroonian cohort were *Prevotella, Bacteroides* and *Faecalibacterium* [32], while in Tanzania, the gut microbiome of Hadza hunter-gatherers was found to be dominant in *Prevotella, Treponema, unclassified Bacteroidetes, Clostridiales* and *Oscillibacter* [41]. These hunter-gatherers showed depleted levels of *Bifidobacterium,* unlike in the Sudanese population, but a reduced level of *Lachnospiraceae* [*Roseburia, Blautia* and

Dorea], like our findings. In Burkino Faso, *Prevotella* was also found to be abundant [42]. In another study in healthy Ghanians, *Subdoligranulum* and *Escherichia-Shigella* were enriched as part of the normal Ghanian gut microbiome however this was not the case in our Sudanese cohort [43], where instead, *Subdoligranulum* was more enriched in the Irel group. *Bifidobacterium* was low in a study carried out in Papa New Guinea, while *Slackia* and *Propionibacterium* [both not detected in the Sudanese], were high. *Lactobacillus* and *Streptococcus* were also high in these studies, similar to our findings [44]. Other studies have noted an increased abundance of *Peptostreptococcus* and *Desulfovibrio* amongst African gut microbiome [26] as well as *Treponema*; however these were not evident in our study group [32]. Sulphate- reducing bacteria are noted to be high among Africans due to increased saccharolytic fermentation [45].

In SudIrel, *Bacteroides* enrichment was found to be tripled [46%], compared to Sud [15%]. *Bacteroides* is known to be associated with westernised diets and the changes in dietary consumption have considerable effects on *Bacteroides* composition amongst the Sudanese living in this environment. It appears that the more distinguished *Bacteroides/ Prevotella* abundances in the Irel and Sud group were seen to be lost in SudIrel, with merging of microbiome gradiences. This may be associated with a complex array of abnormal enterotype patterns, altering states of susceptibilities and protections against diseases that could be unique to this cohort [35].

We further found probiotic bacterial predisposition amongst the Sudanese. *Bifidobacterium* [q=7.023e-03], *Lactobacillus* [q=2.918e-09] and *Weissella* [q=7.120e-07], were found to be enriched amongst the Sudanese compared to Irel. These lactic acid bacteria generally become prominent early in life through breastfeeding, a practice more commonly established in Sudan and other African countries [46-48] [49] but diets also play a role where such genera are commonly associated with the regular consumption of African fermented foods [47].

In this study, Irel had a much lesser abundance [2.8%] of *Bifidobacterium* compared to both Sudanese groups [23-73%], a finding in contrast to other studies from Europe [32] In comparison, *Bifidobacterium* has been found in greater abundance in Japanese adults [50, 51]. Sudanese diets often contain traditional leafy vegetables and peanuts that have been shown to enrich *Lactobacillus* and *Bifidobacterium* [2, 52]. Furthermore, *Bifidobacterium* abundance is inversely correlated with reduced alcohol

consumption [53], a trait strongly evident in the Sudanese population, irrespective of migration. Lactate produced by *Bifidobacterium* may be linked with insulin resistance [32]. *Bifidobacterium* abundance has been shown to protect against diarrhoea, while gut colonisation with *Lactobacillus* and *Bifidobacterium* may offer protection from malarial disease [54].

We found a diminishing abundance of *Lactobacillus* amongst SudIrel compared to Sud, likely due to the replacement of *Lactobacillus*-rich foods in the diets of SudIrel, with more protein, fat, and carbohydrate enriched foods. In addition, the reduced fat intake highlighted in the dietary analysis of Sud may further promote *Lactobacillus* abundance [55], while the higher iron levels observed in the diets of SudIrel, are associated with lower *Lactobacillus* abundance in the gut [56].

Lactobacillus abundance has been positively associated with improved gut integrity and transepithelial resistance of the proximal colon, but several *Lactobacillus* species may also be reservoirs of antimicrobial resistance gene transfer [46] [55]. While *Lactobacillus* genera were enriched in Sud, certain species were not. *Lactobacillus rogosae* was found to be absent in the Sud group. This would suggest that *Lactobacillus* strains also differ in geographic origin and are likely affected by local dietary or probiotic consumption. Overall, these lactic acid bacteria can produce neurotransmitters including gamma aminobutyric acid by *Bifidobacterium*, and dopamine, serotonin and histamine by *Lactobacillus* [54].

Weissella, enriched in Sud, improves the nutritional quality of sorghum-based foods, a vital stable in Sudan [57]. Its high presence in Sud may indicate a common source of food-based products, which is absent from Irish consumption. It may also offer protection against diseases that include irritable bowel syndrome, although it is known to be an opportunistic pathogen [58]. Abundances of *Streptococcus* in the Sudanese [34-56%] compared to the Irel cohort [9.3%] was also evident. *Streptococcus* plays a role in reducing inflammatory gut activity and in the reduction of uremic toxins. This suggests a natural protection against chronic kidney disease in Sudanese cohorts [59]. Probiotics with *Streptococcus thermophilus* has been shown to play a role in preventing the advancement of chronic kidney disease.

Butyrate [butyric acid] production can occur by *Clostridium cluster IV* and *Clostridium cluster XIVa* which includes the *Clostridiales* order and *Lachnospiraceae*

family [60]. *Lachnospiraceae* generally account for 50% abundance of gut microbiome families ; however, in this study, they comprised [29%] of the Sud microbiome composition [61]. While butyric acid production was found to be non-significant between groups, the *Lachnospiraceae* family and several of its genera were consistently reduced in Sud, with median abundances trending upwards in SudIrel i.e. after migration [Figures 1a-1h]. Diets high in meats and cholesterol have been associated with reduced abundances in *Lachnospiraceae* [62], while diets high in omega-3 polyunsaturated fatty acids and pectin are associated with abundance in this genus [63]. A reduced *Lachnospiraceae* abundance in the Sudanese may predispose to the development of *Clostridium difficile* infection, obesity, and type two diabetes [26, 64].

Dorea [Figure 1g] was highest in SudIrel [50%], followed by 32% in Sud and 16% in Irel. This genus has been found to be low in those who develop food allergy as well as Kawasaki disease, a form of vasculitis in children [65, 66]. *Blautia* [Figure 1e] and several species [Figure 2a-2c] were also low in the Sudanese population. *Blautia* has beneficial anti-inflammatory properties and is vital to metabolising undigested carbohydrates. Increases in *Blautia* on the other hand, has been linked to diseases that include the development of renal dysfunction, non-alcoholic fatty liver disease, diabetes, and multiple sclerosis. The genus *Marvinbryantia*, however was significantly lower in the Irel group, a trend unlike the remaining *Lachnospiraceae* genera [Figure 1h]. Interestingly, this may be related to the reduced alcohol consumption amongst the Sudanese as *Marvinbryantia* abundance decreases with alcohol consumption [67]. A reduced abundance of short-chain fatty acid-producing bacteria amongst the Sudanese population may be a factor in allowing for an environment for *Klebsiella* to thrive, a unique finding amongst the Sud group [68]. *Klebsiella* and *Prevotella* combinatory abundance amongst Sud may predispose to hypertensive states [69]. The genera *Prevotella, Ruminococcus* and *Holdemanella,* all abundant in the Sudanese were also found to be high in Tibetans with coronary artery disease [70].

In this study, we found an increased abundance of *Clostridium* amongst SudIrel, which may be related to a higher fibre intake upon migration [71], however, this may also predispose this study cohort to atopy states such as asthma and allergy [72]. *Clostridium leptum* was found to be increased in SudIrel and depleted in Sud, a genus

associated with the development of IBD. Several *Eubacterium* species detected in this study show cholesterol-reducing properties and were found to be low in Sud [62] [Figure 5]. *Ruminiclostridium* was commonly low in Sud and SudIrel compared to Irel [q =0.004], which may be associated with pro-inflammatory diets [73]. Low abundance of *Ruminiclostridium* has been positively correlated with kidney stone development, while it is also considered a part of a microbiome related to healthy sleep patterns [74]. Minimum counts of *Oscillibacter* [q=0.001] was however evident in all groups, a feature suggestive of healthy kidney function [75].

Participants from Irel [24%] and Sud [31%] had the lowest abundances of *Faecalibacterium* compared to those in SudIrel [44%] [q = 1.2206e-4]. Low levels of *Faecalibacterium* in the gut microbiome may also be associated with developing IBD such as Crohn's disease and ulcerative colitis. Our data suggest that *Faecalibacterium* abundance increases after migrating to westernised environments which may impose good protection against gut inflammation in those Sudanese living in Ireland [SudIrel] while those in the Irel and Sud cohort may continue to be susceptible. The species, *Faecalibacterium prausnitzii*, is a strong butyrate producer with anti-inflammatory properties.

Using correlation plotting, we further associated the Sudanese gut microbiome composition to nutrient profiles. Of interest, is the significant association between reduced *Lachnospiraceae* found in Sud and the endemically low iodine levels reported in Sudan [76]. Normal thyroid function has been associated with abundances in *Lachnospiraceae* [77]. Since iodine mean values were significantly lower in Sud [72.9 mcg], $p < 0.001$, it may be an important future consideration when evaluating causation for endemic thyroid and targeting preventative programs [78].

Megamonas was found to be the most positively correlated nutrient-genus, reported to be abundant in healthy individuals, where it harbours many fermentative capabilities towards non-digestible carbohydrates [79]. Sorghum rich foods that consumed regularly in Sudan contain phenolic compounds [80] that may relate to the positive associations with many Sudanese associated gut microbial genera such as *Prevotella*, *Megamonas* and *Bifidobacterium* [81]. In Sud, positive correlations with Vitamin D and *Collinsella* were seen, a vitamin highlighted to increase the abundance of this genus in studies [82], while zinc and *Phascolarctobacterium* have been associated

together in the literature through animal studies [83]. Zinc bioabsorption may also be improved with the abundance of *Lactobacillus* in Sud [84]. These findings, therefore, relate to specific gut microbiome compositional properties in the Sudanese in the maintenance of health and extraction of nutrients in the body.

Our study highlights that the Sudanese population have reduced fibre intake, which is also highlighted by the abundance of genera composition that are likely to thrive in a low-fibre setting. Some of these include *Dialister* [85], *Bifidobacterium* and *Streptococcus* [85]. *Holdemanella*, was significantly abundant in both Sudanese cohorts and is associated with the increased consumption of white rather than red meat [86]. At the same time, *Collinsella* has been shown to be enriched in diets high in triglycerides, cholesterol, red meat, processed foods and in those diets with low fibre intake [26, 85, 87, 88]. *Collinsella* enrichment amongst the Sudanese may be related to the progressive inflammation of rheumatoid arthritis, the development of non-alcoholic fatty liver disease, obesity, and type two diabetes [89-91]. A high abundance of *Akkermansia* amongst those migrating to Ireland [SudIrel] could be related to the higher caloric intake [85] in this cohort and may predispose to the development of type two diabetes [92].

Dialister [from the family *Veillonellaceae*] was abundant in the Sudanese compared to Irel [q = 1.347e-12] group. The species, *Dialister invisus*, however was found to be depleted in Sud in particular. Although often metagenomically abundant, this species has minimal transcription processes in the gut and is thought to originate from the oral cavity [93]. It is, however associated with the development of IBD and type one diabetes [94], while it is also implicated in Crohn's disease patients [95]. Absence of *Dialister invisus* in the Sud group, therefore, and the presence of it in the SudIrel group highlights that the migratory movements of people between African and European continents can predispose to microbiome changes that may allow for the increased susceptibility to certain diseases.

Another feature of Sud gut microbiome was the depleted counts of *Odoribacter [Bacteroidetes]* compared to SudIrel and Irel. *Odoribacter* is part of a healthy microbiome and produce numerous short-chain fatty acids that play a role in exerting gut homeostasis and immunomodulatory activity. We found an increase, however of *Odoribacter* in the SudIrel group, highlighting that those living in Ireland show

increased gains in *Odoribacter* with migratory movement, which may highlight protection in this population against Crohn's disease and severe forms of ulcerative colitis however, in other studies, abundance of *Odoribacter* is associated with reduced blood pressure.

Alistipes was also depleted in Sud, including the species, *Alistipes finegoldii* and *Alistipes timonensis*. A reduced protein intake in Sud, may be a factor in this due to *Alistipes* having a rich presence of putrefaction pathways involved in protein metabolism. However, *Alistipes finegoldii* has been implicated as a pro-inflammatory genus; thus, the Sud cohort could have a more beneficial association with the depletion of this genus, particularly the lessened risk of developing colorectal cancer, IBD and kidney disease [75]. In other studies, a low presence of this species has been associated with chronic fatigue syndrome, depression [96] and the progression of liver diseases [cirrhosis and hepatic encephalopathy] [97], autoimmune disease and atrial fibrillation [98].

Furthermore, the stability and balance of a healthy gut are affected by methane producers such as *Methanobrevibacter,* which was found to be abundant in Sud. African populations have shown increased methanogen abundance in relation to diets with low animal food consumption and caloric intake [99], likely in response to more efficient nutrient extraction, however, studies have shown both abundance [100] and depletions of this genus in relation to obesity [101]. Depletion of this genus has also been associated with IBD [102].

PICRUSt data was utilised to infer the functional contribution of the microbiota from the three cohorts, and, incorporated with LEfSe plotting, proved subsequent variations between the Sudanese and Irish cohorts [Figures 8a-8b]. The Sudanese had active pathways for nicotinate and nicotinamide metabolism [K08281], a pathway associated with abundances in the genus *Faecalibacterium* [103]. This pathway may be involved in the degradation of neuroprotective molecules that protect against Parkinson's disease [104] and is also an active pathway in depression [103]. The activity of the fluoride exporter, K06199, was discriminant for the Sudanese populations and may highlight fluoride tolerance but also increased exposure to polluted environments [105].

In Irel, prominent activated pathways included genetic information processing and the ABC transporters, sugar transport permease proteins that have been associated with health [106]. On the other hand, multiple drug resistance ontologies [K09687 and K09686] were prominent in Irel, a feature that could highlight increased antibiotic use in westernised environments. PICRUSt data also indicated that pathways involved in extra cytoplasmic stress responses [K03088] were discriminant of this cohort [107].

Variations between the Sud and SudIrel groups included prominent KEGG pathways related to sugar transport, carbohydrate metabolism and fatty acid degradation in SudIrel, highlighting that changes in the diet of those that migrate to Ireland may give rise to more active orthology pathways associated with nourishment breakdown, while in Sud, pathways that included, K03496 [genetic information processing associated with cytoskeleton proteins] and K02025 [sugar transport system permease protein] differed and are associated with healthy gut characteristics [108].

We analysed gut metabolites to better understand diet, microbiome, and host relations with several distinct patterns appreciated within the metabolomic findings. Many amino acids had greater log fold change in SudIrel compared to the remaining two groups. An important bidirectional link is present between amino acids and the gut microbiome. Amino acid breakdown often relates to the promotion of a healthy gut microbiome and is involved in many host functions including immunity. The microbiome of Sud was abundant in microbiota such *as Klebsiella* and *Prevotella* that are involved in amino acid utilisation compared to Irel and SudIrel, thus giving rise to a reduced amino acid portfolio in this cohort [109, 110].

Genera and species that promote branched-chain amino acids [isoleucine and leucine] include, *Akkermansia*, *Bifidobacterium* and *Prevotella copri*, all higher in abundance across the Sudanese groups but have been associated with obesity and type two diabetes [111].

Tryptophan derivatives, Kynurenine and Kynurenic acid, were found to be high in Sud and SudIrel, while tryptamine was high in SudIrel. Arginine was high in Sud, which may relate to a protective effect against bacterial virulence proteins. Putrescine, produced from the catabolism of arginine, was higher in Sud than Irel [q=0.02]. This soybean protein toxin derivative may damage the intestinal wall [110] and is likely

associated with the staple soybean diet of Sudan. Median levels of theanine were found to be lowest in Irel and highest in the Sudanese, likely related to black tea consumption habits of the latter cohort [112].

Sugar-alcohol [polyols] was high in both Sud and SudIrel compared to Irel [log2 ratio 3.6 and 2.7, respectively]. Sugar-alcohol includes sorbitol, mannitol and xylitol [113] and may indicate increased consumption of processed foods after migration. However they have been shown to have limited effect on raising blood glucose and insulin and may have positive health effects, particularly in diabetic control [114]. N8-acetylspermidine is a biomarker for ischaemic cardiomyocyte death and cardiac dysfunction and was found to be high amongst the Sudanese groups compared to Irel. This may indicate an increased susceptibility towards cardiovascular disease [115]. Trimethylamine–N-Oxide was also high in both Sudanese groups. Increases in Trimethylamine–N-Oxide levels are associated with the increased risk of cardiovascular disease [116], development of atherosclerosis [117], kidney failure and obesity [111].

Furthermore, metabolites of the kynurenine pathway involved in metabolic disease were prominent in Sud [kynurenic acid] [111]. 4- coumaric acid, a phenolic acid found abundantly in plants, was found to be high in Sud compared to Irel, a metabolite with antioxidant properties [118]. The vinyl derivative, 4 vinylphenol, was high in Sud and SudIrel, which has been shown to have a significant inhibitory effect on gram-negative food-borne bacteria [119].

Short-chain fatty acids are end products of microbial fermentation that are involved in healthy mucosal integrity, glucose and lipid metabolism, and the regulation of inflammation as well as immune responses. Overall, short-chain fatty acids were considerably lower in the Irel group compared to Sud and SudIrel. This may have a protective effect on the Sudanese population against metabolic disorders. Butyric acid is a histone deacetylase inhibitor related to the epigenetic regulation that protects against tumour growth [110].

Conclusion

Dietary variations between Africa and Europe are related to the sourcing of foods, food content, preparation, storage, and food frequency. Such factors in turn, affect the abundances and depletions in microbiome compositions, which can significantly impact many aspects of health. The Sudanese living in Ireland [SudIrel] were a unique cohort bundled by the unique and continuous trend of 'sandwiched' positioning between the other two groups reflected in their microbiome and metabolomic expressions. Variations in the metabolome were most distant between Sud and Irel and negligible between SudIrel and Irel, indicating that bacterial function trails a similar pattern in accordance with geographical habitat.

In SudIrel, a gain of westernised microbiota such as *Bacteroides* and loss of African-related heritage genera such as *Lactobacillus* and *Bifidobacterium* further indicates how migrational resculpting of the microbiota composition occurs with varying impacts on health. For example, the SudIrel group may have an increased risk of developing diseases, an affect that might only take place after loss of traditional habitat. However, on the contrary, loss of abundance in some genera such as *Klebsiella* [which was only found in Sud, and seemed to be lost in SudIrel], may harbour a protective affect against cardiovascular disease, which may only take place after migration.

This study serves as the first platform for the fundamental understanding of the microbiome changes and responsive facets that occur after population movements in those from an African lineage. It serves to be a cornerstone for developing new research into focused migrational-associated diseases and advancements of public health interventions to non-communicable diseases that may be targeted with a scientific microbial strategy. It may be that assessment of the microbiome in those who migrate becomes a necessary step in evaluating disease risk and development as well as future management and recording of progressive outcomes concerning diet and microbiome composition.

5.6 References

1. Allali I, Abotsi RE, Tow LA, Thabane L, Zar HJ, Mulder NM, Nicol MP. 2021. Human microbiota research in Africa: a systematic review reveals gaps and priorities for future research. Microbiome 9:241-241.
2. Brewster R, Tamburini FB, Asiimwe E, Oduaran O, Hazelhurst S, Bhatt AS. 2019. Surveying Gut Microbiome Research in Africans: Toward Improved Diversity and Representation. Trends in Microbiology 27:824-835.
3. Khalil S. 2020. Cardiovascular disease in Sudan: Past, Present and Future Prospects.
4. Awadalla H, Noor SK, Elmadhoun WM, Almobarak AO, Elmak NE, Abdelaziz SI, Sulaiman AA, Ahmed MH. 2017. Diabetes complications in Sudanese individuals with type 2 diabetes: Overlooked problems in sub-Saharan Africa? Diabetes & Metabolic Syndrome: Clinical Research & Reviews 11:S1047-S1051.
5. Kellermayer R, Zilbauer M. 2020. The Gut Microbiome and the Triple Environmental Hit Concept of Inflammatory Bowel Disease Pathogenesis. Journal of Pediatric Gastroenterology and Nutrition 71.
6. Hurwitz EE, Simon M, Vinta SR, Zehm CF, Shabot SM, Minhajuddin A, Abouleish AE. 2017. Adding Examples to the ASA-Physical Status Classification Improves Correct Assignment to Patients. Anesthesiology 126:614-622.
7. Mulligan AA, Luben RN, Bhaniani A, Parry-Smith DJ, O'Connor L, Khawaja AP, Forouhi NG, Khaw K-T. 2014. A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability. BMJ Open 4:e004503.
8. Magoč T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957-2963.
9. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-1.
10. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26:266-7.

11. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590-6.
12. McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20-5.
13. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814-21.
14. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. 2017. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Research 45:W180-W188.
15. Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B [Methodological] 57:289-300.
16. Johnsen L, Skou P, Khakimov B, Bro R. 2017. Gas chromatography mass spectrometry data processing made easy. Journal of Chromatography A 1503.
17. Thukral A. 2017. A review on measurement of Alpha diversity in biology. Agricultural Research Journal 54:1.
18. Doumatey AP, Adeyemo A, Zhou J, Lei L, Adebamowo SN, Adebamowo C, Rotimi CN. 2020. Gut microbiome profiles are associated with type 2 diabetes in urban Africans. Frontiers in cellular and infection microbiology:63.
19. Isibor P, Akinduti P, Aworunse O, Oyewale J, Oshamika O, Ugboko H, Taiwo O, Ahuekwe E, Obafemi Y, Onibokun E, Oziegbe O, Oniha M, Olopade B, Atolagbe O, Adekeye B, Ajiboye I, Bello O, Popoola J, Ayanda O, Akinnola O, Olasehinde G, Eni A, Nwinyi O, Omonhinmin C, Oranusi S, Obembe O. 2021. Significance of African Diets in Biotherapeutic Modulation of the Gut Microbiome. Bioinformatics and Biology Insights 15:11779322211012697.
20. Lassetter JH, Callister LC. 2009. The impact of migration on the health of voluntary migrants in western societies: a review of the literature. Journal of transcultural nursing 20:93-104.

21. Caperchione CM, Kolt GS, Mummery WK. 2009. Physical activity in culturally and linguistically diverse migrant groups to Western Society. Sports medicine 39:167-177.
22. Agyemang C, van den Born B-J. 2019. Non-communicable diseases in migrants: an expert review. Journal of travel medicine 26:tay107.
23. Bo A, Zinckernagel L, Krasnik A, Petersen JH, Norredam M. 2015. Coronary heart disease incidence among non-Western immigrants compared to Danish-born people: effect of country of birth, migrant status, and income. European journal of preventive cardiology 22:1281-1289.
24. Charani E, Cunnington AJ, Yousif AH, Ahmed MS, Ahmed AE, Babiker S, Badri S, Buytaert W, Crawford MA, Elbashir MI. 2019. In transition: current health challenges and priorities in Sudan. BMJ global health 4:e001723.
25. Pengpid S, Peltzer K. 2022. Prevalence and correlates of multiple non-communicable diseases risk factors among male and female adults in Sudan: results of the first national STEPS survey in 2016. African Health Sciences 22:728-735.
26. Doumatey AP, Adeyemo A, Zhou J, Lei L, Adebamowo SN, Adebamowo C, Rotimi CN. 2020. Gut Microbiome Profiles Are Associated With Type 2 Diabetes in Urban Africans. Frontiers in Cellular and Infection Microbiology 10.
27. Mobeen F, Sharma V, Tulika P. 2018. Enterotype Variations of the Healthy Human Gut Microbiome in Different Geographical Regions. Bioinformation 14:560-573.
28. Vangay P, Johnson AJ, Ward TL, Al-Ghalith GA, Shields-Cutler RR, Hillmann BM, Lucas SK, Beura LK, Thompson EA, Till LM, Batres R, Paw B, Pergament SL, Saenyakul P, Xiong M, Kim AD, Kim G, Masopust D, Martens EC, Angkurawaranon C, McGready R, Kashyap PC, Culhane-Pera KA, Knights D. 2018. US Immigration Westernizes the Human Gut Microbiome. Cell 175:962-972.e10.
29. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559-63.

30. Amato KR, Jeyakumar T, Poinar H, Gros P. 2019. Shifting climates, foods, and diseases: the human microbiome through evolution. Bioessays 41:1900034.
31. Shanahan F, Ghosh TS, O'Toole PW. 2021. The healthy microbiome—what is the definition of a healthy gut microbiome? Gastroenterology 160:483-494.
32. Rubel MA, Abbas A, Taylor LJ, Connell A, Tanes C, Bittinger K, Ndze VN, Fonsah JY, Ngwang E, Essiane A, Fokunang C, Njamnshi AK, Bushman FD, Tishkoff SA. 2020. Lifestyle and the presence of helminths is associated with gut microbiome composition in Cameroonians. Genome Biology 21:122.
33. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science 334:105-8.
34. Power SE, O'Toole PW, Stanton C, Ross RP, Fitzgerald GF. 2014. Intestinal microbiota, diet and health. British Journal of Nutrition 111:387-402.
35. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, Mele MC. 2019. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. Microorganisms 7:14.
36. Kelly TN, Bazzano LA, Ajami NJ, He H, Zhao J, Petrosino JF, Correa A, He J. 2016. Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. Circulation Research 119:956-964.
37. Joo E-J, Cheong HS, Kwon M-J, Sohn W, Kim H-N, Cho YK. 2021. Relationship between gut microbiome diversity and hepatitis B viral load in patients with chronic hepatitis B. Gut Pathogens 13:65.
38. Tett A, Huang KD, Asnicar F, Fehlner-Peach H, Pasolli E, Karcher N, Armanini F, Manghi P, Bonham K, Zolfo M, De Filippis F, Magnabosco C, Bonneau R, Lusingu J, Amuasi J, Reinhard K, Rattei T, Boulund F, Engstrand L, Zink A, Collado MC, Littman DR, Eibach D, Ercolini D, Rota-Stabelli O, Huttenhower C, Maixner F, Segata N. 2019. The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations. Cell Host Microbe 26:666-679.e7.

39. Fehlner-Peach H, Magnabosco C, Raghavan V, Scher JU, Tett A, Cox LM, Gottsegen C, Watters A, Wiltshire-Gordon JD, Segata N, Bonneau R, Littman DR. 2019. Distinct Polysaccharide Utilization Profiles of Human Intestinal Prevotella copri Isolates. Cell Host Microbe 26:680-690.e5.
40. Larsen JM. 2017. The immune response to Prevotella bacteria in chronic inflammatory disease. Immunology 151:363-374.
41. Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, Turroni S, Biagi E, Peano C, Severgnini M, Fiori J, Gotti R, De Bellis G, Luiselli D, Brigidi P, Mabulla A, Marlowe F, Henry AG, Crittenden AN. 2014. Gut microbiome of the Hadza hunter-gatherers. Nat Commun 5:3654.
42. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet J-B, Massart S, Collini S, Pieraccini G, Lionetti P. 2010. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proceedings of the National Academy of Sciences 107:14691 - 14696.
43. Parbie PK, Mizutani T, Ishizaka A, Kawana-Tachikawa A, Runtuwene LR, Seki S, Abana CZ, Kushitor D, Bonney EY, Ofori SB, Uematsu S, Imoto S, Kimura Y, Kiyono H, Ishikawa K, Ampofo WK, Matano T. 2021. Fecal Microbiome Composition in Healthy Adults in Ghana. Jpn J Infect Dis 74:42-47.
44. Martínez I, Stegen JC, Maldonado-Gómez MX, Eren AM, Siba PM, Greenhill AR, Walter J. 2015. The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. Cell Rep 11:527-38.
45. Ou J, Carbonero F, Zoetendal EG, DeLany JP, Wang M, Newton K, Gaskins HR, O'Keefe SJD. 2013. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. The American journal of clinical nutrition 98:111-120.
46. Ouoba LII, Lei V, Jensen LB. 2008. Resistance of potential probiotic lactic acid bacteria and bifidobacteria of African and European origin to antimicrobials: Determination and transferability of the resistance genes to other bacteria. International Journal of Food Microbiology 121:217-224.
47. Adimpong DB, Nielsen DS, Sørensen KI, Derkx PMF, Jespersen L. 2012. Genotypic characterization and safety assessment of lactic acid bacteria from indigenous African fermented food products. BMC Microbiology 12:75.

48. Grzeskowiak L, Collado MC, Mangani C, Maleta K, Laitinen K, Ashorn P, Isolauri E, Salminen S. 2012. Distinct Gut Microbiota in Southeastern African and Northern European Infants. Journal of Pediatric Gastroenterology and Nutrition 54.
49. Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A. 2012. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. Am J Clin Nutr 96:544-51.
50. Kato K, Ishida S, Tanaka M, Mitsuyama E, Xiao J-z, Odamaki T. 2018. Association between functional lactase variants and a high abundance of Bifidobacterium in the gut of healthy Japanese people. PLOS ONE 13:e0206189.
51. Odamaki T, Kato K, Sugahara H, Hashikura N, Takahashi S, Xiao J-Z, Abe F, Osawa R. 2016. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. BMC microbiology 16:90-90.
52. Kassim MA, Baijnath H, Odhav B. 2014. Effect of traditional leafy vegetables on the growth of lactobacilli and bifidobacteria. International Journal of Food Sciences and Nutrition 65:977-980.
53. Latorre-Pérez A, Hernández M, Iglesias JR, Morán J, Pascual J, Porcar M, Vilanova C, Collado L. 2021. The Spanish gut microbiome reveals links between microorganisms and Mediterranean diet. Scientific Reports 11:21602.
54. Dordević D, Jančíková S, Vítězová M, Kushkevych I. 2021. Hydrogen sulfide toxicity in the gut environment: Meta-analysis of sulfate-reducing and lactic acid bacteria in inflammatory processes. Journal of Advanced Research 27:55-69.
55. Lam YY, Ha CWY, Campbell CR, Mitchell AJ, Dinudom A, Oscarsson J, Cook DI, Hunt NH, Caterson ID, Holmes AJ, Storlien LH. 2012. Increased Gut Permeability and Microbiota Change Associate with Mesenteric Fat Inflammation and Metabolic Dysfunction in Diet-Induced Obese Mice. PLOS ONE 7:e34233.
56. Zimmermann MB, Chassard C, Rohner F, N'Goran EK, Nindjin C, Dostal A, Utzinger J, Ghattas H, Lacroix C, Hurrell RF. 2010. The effects of iron fortification on the gut microbiota in African children: a randomized controlled trial in Côte d'Ivoire. The American Journal of Clinical Nutrition 92:1406-1415.

57. Olojede AO, Sanni AI, Banwo K, Michael T. 2022. Improvement of Texture, Nutritional Qualities, and Consumers’ Perceptions of Sorghum-Based Sourdough Bread Made with Pediococcus pentosaceus and Weissella confusa Strains. Fermentation 8:32.
58. Fusco V, Quero GM, Cho G-S, Kabisch J, Meske D, Neve H, Bockelmann W, Franz CMAP. 2015. The genus Weissella: taxonomy, ecology and biotechnological potential. Frontiers in Microbiology 6.
59. Vitetta L, Llewellyn H, Oldfield D. 2019. Gut Dysbiosis and the Intestinal Microbiome: Streptococcus thermophilus a Key Probiotic for Reducing Uremia. Microorganisms 7.
60. Pichler MJ, Yamada C, Shuoker B, Alvarez-Silva C, Gotoh A, Leth ML, Schoof E, Katoh T, Sakanaka M, Katayama T, Jin C, Karlsson NG, Arumugam M, Fushinobu S, Abou Hachem M. 2020. Butyrate producing colonic Clostridiales metabolise human milk oligosaccharides and cross feed on mucin via conserved pathways. Nature Communications 11:3285.
61. Biddle A, Stewart L, Blanchard J, Leschine S. 2013. Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. Diversity 5:627-640.
62. Vacca M, Celano G, Calabrese FM, Portincasa P, Gobbetti M, De Angelis M. 2020. The Controversial Role of Human Gut Lachnospiraceae. Microorganisms 8:573.
63. Cotta M, Forster R. 2006. The Family Lachnospiraceae, Including the Genera Butyrivibrio, Lachnospira and Roseburia, p 1002-1021. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E [ed], The Prokaryotes: Volume 4: Bacteria: Firmicutes, Cyanobacteria doi:10.1007/0-387-30744-3_35. Springer US, New York, NY.
64. Sagheddu V, Patrone V, Miragoli F, Puglisi E, Morelli L. 2016. Infant Early Gut Colonization by Lachnospiraceae: High Frequency of Ruminococcus gnavus. Frontiers in pediatrics 4:57-57.
65. Shen J, Ding Y, Yang Z, Zhang X, Zhao M. 2020. Effects of changes on gut microbiota in children with acute Kawasaki disease. PeerJ 8:e9698-e9698.
66. Savage JH, Lee-Sarwar KA, Sordillo J, Bunyavanich S, Zhou Y, O'Connor G, Sandel M, Bacharier LB, Zeiger R, Sodergren E, Weinstock GM, Gold DR, Weiss ST, Litonjua AA. 2018. A prospective microbiome-wide association

study of food sensitization and food allergy in early childhood. Allergy 73:145-152.

67. Yang F, Wei J-d, Lu Y-f, Sun Y-l, Wang Q, Zhang R-l. 2019. Galacto-oligosaccharides modulate gut microbiota dysbiosis and intestinal permeability in rats with alcohol withdrawal syndrome. Journal of Functional Foods 60:103423.

68. Vornhagen J, Bassis CM, Ramakrishnan S, Hein R, Mason S, Bergman Y, Sunshine N, Fan Y, Holmes CL, Timp W, Schatz MC, Young VB, Simner PJ, Bachman MA. 2021. A plasmid locus associated with Klebsiella clinical infections encodes a microbiome-dependent gut fitness factor. PLOS Pathogens 17:e1009537.

69. Li J, Zhao F, Wang Y, Chen J, Tao J, Tian G, Wu S, Liu W, Cui Q, Geng B, Zhang W, Weldon R, Auguste K, Yang L, Liu X, Chen L, Yang X, Zhu B, Cai J. 2017. Gut microbiota dysbiosis contributes to the development of hypertension. Microbiome 5:14.

70. Ma Y, Zhu L, Ma Z, Gao Z, Wei Y, Shen Y, Li L, Liu X, Ren M. 2021. Distinguishing feature of gut microbiota in Tibetan highland coronary artery disease patients and its link with diet. Sci Rep 11:18486.

71. Walter J, Ley R. 2011. The Human Gut Microbiome: Ecology and Recent Evolutionary Changes. Annual Review of Microbiology 65:411-429.

72. Guo P, Zhang K, Ma X, He P. 2020. Clostridium species as probiotics: potentials and challenges. Journal of Animal Science and Biotechnology 11:24.

73. Zheng J, Hoffman KL, Chen J-S, Shivappa N, Sood A, Browman GJ, Dirba DD, Hanash S, Wei P, Hebert JR, Petrosino JF, Schembre SM, Daniel CR. 2020. Dietary inflammatory potential in relation to the gut microbiome: results from a cross-sectional study. British Journal of Nutrition 124:931-942.

74. Thompson RS, Gaffney M, Hopkins S, Kelley T, Gonzalez A, Bowers SJ, Vitaterna MH, Turek FW, Foxx CL, Lowry CA, Vargas F, Dorrestein PC, Wright KP, Jr., Knight R, Fleshner M. 2021. Ruminiclostridium 5, Parabacteroides distasonis, and bile acid profile are modulated by prebiotic diet and associate with facilitated sleep/clock realignment after chronic disruption of rhythms. Brain Behav Immun 97:150-166.

75. Kim JE, Kim HE, Park JI, Cho H, Kwak MJ, Kim BY, Yang SH, Lee JP, Kim DK, Joo KW, Kim YS, Kim BS, Lee H. 2020. The Association between Gut Microbiota and Uremia of Chronic Kidney Disease. Microorganisms 8.
76. Medani AMM, Elnour AA, Saeed AM. 2011. Endemic goitre in the Sudan despite long-standing programmes for the control of iodine deficiency disorders. Bulletin of the World Health Organization 89:121-126.
77. Cayres LCdF, de Salis LVV, Rodrigues GSP, Lengert AvH, Biondi APC, Sargentini LDB, Brisotti JL, Gomes E, de Oliveira GLV. 2021. Detection of alterations in the gut microbiota and intestinal permeability in patients with Hashimoto thyroiditis. Frontiers in Immunology 12:579140.
78. Bani I. 2007. Accelerating progress on salt iodisation in Sudan: time for action. Disasters 31:S139-S149.
79. Shimizu J, Kubota T, Takada E, Takai K, Fujiwara N, Arimitsu N, Ueda Y, Wakisaka S, Suzuki T, Suzuki N. 2019. Relative abundance of Megamonas hypermegale and Butyrivibrio species decreased in the intestine and its possible association with the T cell aberration by metabolite alteration in patients with Behcet's disease [210 characters]. Clinical rheumatology 38:1437-1445.
80. Xu J, Wang W, Zhao Y. 2021. Phenolic Compounds in Whole Grain Sorghum and Their Health Benefits. Foods 10.
81. Martín-Núñez GM, Cornejo-Pareja I, Coin-Aragüez L, Roca-Rodríguez MdM, Muñoz-Garach A, Clemente-Postigo M, Cardona F, Moreno-Indias I, Tinahones FJ. 2019. H. pylori eradication with antibiotic treatment causes changes in glucose homeostasis related to modifications in the gut microbiota. PLoS One 14:e0213548.
82. Pham VT, Fehlbaum S, Seifert N, Richard N, Bruins MJ, Sybesma W, Rehman A, Steinert RE. 2021. Effects of colon-targeted vitamins on the composition and metabolic activity of the human gut microbiome–a pilot study. Gut microbes 13:1875774.
83. Pereira AM, Maia MR, Pinna C, Biagi G, Matos E, Segundo MA, Fonseca AJ, Cabrita AR. 2021. Effects of Zinc Source and Enzyme Addition on the Fecal Microbiota of Dogs. Frontiers in microbiology 12.

84. Leonardi A, Zanoni S, De Lucia M, Amaretti A, Raimondi S, Rossi M. 2013. Zinc uptake by lactic acid bacteria. International Scholarly Research Notices 2013.
85. Gomez-Arango LF, Barrett HL, Wilkinson SA, Callaway LK, McIntyre HD, Morrison M, Dekker Nitert M. 2018. Low dietary fiber intake increases Collinsella abundance in the gut microbiota of overweight and obese pregnant women. Gut Microbes 9:189-201.
86. Lin Q, Lun J, Zhang J, He X, Gong Z, Gao X, Cao H. 2021. [Gut microbiome composition in pre-adolescent children with different meat consumption patterns]. Nan Fang Yi Ke Da Xue Xue Bao 41:1801-1088.
87. Astbury S, Atallah E, Vijay A, Aithal GP, Grove JI, Valdes AM. 2020. Lower gut microbiome diversity and higher abundance of proinflammatory genus Collinsella are associated with biopsy-proven nonalcoholic steatohepatitis. Gut Microbes 11:569-580.
88. Olendzki B, Bucci V, Cawley C, Maserati R, McManus M, Olednzki E, Madziar C, Chiang D, Ward DV, Pellish R. 2022. Dietary manipulation of the gut microbiome in inflammatory bowel disease patients: Pilot study. Gut microbes 14:2046244.
89. Companys J, Gosalbes MJ, Pla-Pagà L, Calderón-Pérez L, Llauradó E, Pedret A, Valls RM, Jiménez-Hernández N, Sandoval-Ramirez BA, Del Bas JM. 2021. Gut microbiota profile and its association with clinical variables and dietary intake in overweight/obese and lean subjects: a cross-sectional study. Nutrients 13:2032.
90. Ruiz-Limón P, Mena-Vázquez N, Moreno-Indias I, Manrique-Arija S, Lisbona-Montañez JM, Cano-García L, Tinahones FJ, Fernández-Nebro A. 2022. Collinsella is associated with cumulative inflammatory burden in an established rheumatoid arthritis cohort. Biomedicine & Pharmacotherapy 153:113518.
91. Cancello R, Turroni S, Rampelli S, Cattaldo S, Candela M, Cattani L, Mai S, Vietti R, Scacchi M, Brigidi P. 2019. Effect of short-term dietary intervention and probiotic mix supplementation on the gut microbiota of elderly obese women. Nutrients 11:3011.
92. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai

Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, et al. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490:55-60.

93. Rashidi A, Ebadi M, Weisdorf DJ, Costalonga M, Staley C. 2021. No evidence for colonization of oral bacteria in the distal gut in healthy adults. Proceedings of the National Academy of Sciences 118:e2114152118.

94. Schirmer M, Franzosa EA, Lloyd-Price J, McIver LJ, Schwager R, Poon TW, Ananthakrishnan AN, Andrews E, Barron G, Lake K, Prasad M, Sauk J, Stevens B, Wilson RG, Braun J, Denson LA, Kugathasan S, McGovern DPB, Vlamakis H, Xavier RJ, Huttenhower C. 2018. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat Microbiol 3:337-346.

95. Joossens M, Huys G, Cnockaert M, De Preter V, Verbeke K, Rutgeerts P, Vandamme P, Vermeire S. 2011. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. Gut 60:631-637.

96. Frémont M, Coomans D, Massart S, De Meirleir K. 2013. High-throughput 16S rRNA gene sequencing reveals alterations of intestinal microbiota in myalgic encephalomyelitis/chronic fatigue syndrome patients. Anaerobe 22:50-56.

97. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. 2020. The Genus Alistipes: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. Front Immunol 11:906.

98. Zuo K, Li J, Li K, Hu C, Gao Y, Chen M, Hu R, Liu Y, Chi H, Wang H, Qin Y, Liu X, Li S, Cai J, Zhong J, Yang X. 2019. Disordered gut microbiota and alterations in metabolic patterns are associated with atrial fibrillation. Gigascience 8.

99. Armougom F, Henry M, Vialettes B, Raccah D, Raoult D. 2009. Monitoring Bacterial Community of Human Gut Microbiota Reveals an Increase in Lactobacillus in Obese Patients and Methanogens in Anorexic Patients. PLOS ONE 4:e7125.

100. Maya-Lucas O, Murugesan S, Nirmalkar K, Alcaraz LD, Hoyo-Vadillo C, Pizano-Zárate ML, García-Mena J. 2019. The gut microbiome of Mexican children affected by obesity. Anaerobe 55:11-23.

101. Crovesy L, Masterson D, Rosado EL. 2020. Profile of the gut microbiota of adults with obesity: a systematic review. European Journal of Clinical Nutrition 74:1251-1262.
102. Conlon MA, Bird AR. 2014. The impact of diet and lifestyle on gut microbiota and human health. Nutrients 7:17-44.
103. Dong Z, Shen X, Hao Y, Li J, Li H, Xu H, Yin L, Kuang W. 2021. Gut Microbiome: A Potential Indicator for Differential Diagnosis of Major Depressive Disorder and General Anxiety Disorder. Frontiers in Psychiatry 12.
104. Wallen ZD, Demirkan A, Twa G, Cohen G, Dean MN, Standaert DG, Sampson TR, Payami H. 2022. Metagenomics of Parkinson's disease implicates the gut microbiome in multiple disease mechanisms. Nature Communications 13:6958.
105. Mandal M, Mandal S. 2022. Cross-biome metagenomic analyses of the impact of pollutants on taxonomic and functional diversity of bacterial communities from different geographical regions. Gene Reports 29:101690.
106. Guard BC, Barr JW, Reddivari L, Klemashevich C, Jayaraman A, Steiner JM, Vanamala J, Suchodolski JS. 2015. Characterization of microbial dysbiosis and metabolomic changes in dogs with acute diarrhea. PloS one 10:e0127259.
107. Sung JJY, Coker OO, Chu E, Szeto CH, Luk STY, Lau HCH, Yu J. 2020. Gastric microbes associated with gastric inflammation, atrophy and intestinal metaplasia 1 year after Helicobacter pylori eradication. Gut 69:1572-1580.
108. Han M, Yang K, Yang P, Zhong C, Chen C, Wang S, Lu Q, Ning K. 2020. Stratification of athletes' gut microbiota: the multifaceted hubs associated with dietary factors, physical characteristics and performance. Gut Microbes 12:1842991.
109. Ma N, Ma X. 2019. Dietary Amino Acids and the Gut-Microbiome-Immune Axis: Physiological Metabolism and Therapeutic Prospects. Compr Rev Food Sci Food Saf 18:221-242.
110. Zhao J, Zhang X, Liu H, Brown MA, Qiao S. 2019. Dietary Protein and Gut Microbiota Composition and Function. Curr Protein Pept Sci 20:145-154.
111. Agus A, Clément K, Sokol H. 2021. Gut microbiota-derived metabolites as central regulators in metabolic disorders. Gut 70:1174-1182.

112. Xu W, Kong Y, Zhang T, Gong Z, Xiao W. 2022. L-Theanine regulates lipid metabolism by modulating gut microbiota and bile acid metabolism. J Sci Food Agric doi:10.1002/jsfa.12222.
113. Hattori K, Akiyama M, Seki N, Yakabe K, Hase K, Kim YG. 2021. Gut Microbiota Prevents Sugar Alcohol-Induced Diarrhea. Nutrients 13.
114. Msomi NZ, Erukainure OL, Islam MS. 2021. Suitability of sugar alcohols as antidiabetic supplements: A review. J Food Drug Anal 29:1-14.
115. Nayak A, Liu C, Mehta A, Ko YA, Tahhan AS, Dhindsa DS, Uppal K, Jones DP, Butler J, Morris AA, Quyyumi AA. 2020. N8-Acetylspermidine: A Polyamine Biomarker in Ischemic Cardiomyopathy With Reduced Ejection Fraction. J Am Heart Assoc 9:e016055.
116. Schoeler M, Caesar R. 2019. Dietary lipids, gut microbiota and lipid metabolism. Rev Endocr Metab Disord 20:461-472.
117. Oliphant K, Allen-Vercoe E. 2019. Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health. Microbiome 7:91.
118. Luceri C, Guglielmi F, Lodovici M, Giannini L, Messerini L, Dolara P. 2004. Plant phenolic 4-coumaric acid protects against intestinal inflammation in rats. Scand J Gastroenterol 39:1128-33.
119. Licandro-Seraut H, Roussel C, Perpetuini G, Gervais P, Cavin JF. 2013. Sensitivity to vinyl phenol derivatives produced by phenolic acid decarboxylase activity in Escherichia coli and several food-borne Gram-negative species. Appl Microbiol Biotechnol 97:7853-64.
120. Galvin M, Kiely M, Harrington K, Robson P, Moore R, Flynn A. 2001. The North/South Ireland food consumption survey: the dietary fibre intake of Irish adults. Public Health Nutrition 4:1061-1068.

General Discussion

North Sudan is the third largest country in Africa and is of low to middle-income status, with 47% of the population [~41 million] predominantly living in poverty. In 2011, the separation of south and north Sudan further led to the downfall of the northern Sudanese economy, with the loss of 75% of oil resources and almost half of the country's budget revenue.

The country now faces many social and economic insecurities that, in recent years have contributed incrementally to the incidence and intensity of poverty which are measurements of the global multidimensional poverty index [MDPI] [1]. 14% of adults and 22% of children in Sudan are now categorised as being in an MDPI setting [2], measured by deprivations in health, education and living standards. Food consumption in particular, accounts for 23% for the poverty factors in Sudan [2]. Furthermore, Sudan has a young population [median age = 19.6 years] and a total life expectancy estimated at around 62-66 years, compared to a life expectancy of 82.1 years in Ireland.

The Sudabiome project aimed for the first time to venture onto new land through one of the most modern research advancements [metagenomics] for the understanding of a multi-ethnic society. Indeed, metagenomic science is becoming an indispensable discipline, continuously informing us about the individual and broader interactions that the microbiome has with health and disease. Another aim of the Sudabiome was to generate robust datasets useful in studying the complexity, variability, and effects of microbial communities in humans and the Sudanese environment. Optimisation of this information was further enhanced by comparability at an international level, such as with the Irish population. Ireland is a leading country in microbiome science, helping to deepen profoundly the impact such research can offer to both populations. The justification of this thesis was thus the integration of tradition and science with urgency to apprehend how the microbiome relates to key areas in Sudan's population, culture, and diet, as well as evaluating Sudanese human health and disease-associated portfolios.

The project was first paralleled with the most suitable United Nations sustainable development goals [SDGs]. It was found to contribute to goals 3 [good health and well-being], 11 [sustainable cities and communities], 12 [responsible consumption and production] and 17 [partnership for the goals] [Figure 1]. This was an essential undertaking as orientating Sudabiome research aims with the SDGs and subsequently, the produced results helped achieve a solid anchor for future research pathways in an organised manner.

It was imperative to then achieve an updated global review [Chapter 1] on the oral microbiome's role in the development of or protection against oral squamous cell carcinoma [OSCC], and how it may take part in contributing to the better diagnosis as well as treatment of this disease [3]. It was appreciated that ever since Virchow in 1863 linked 'bacteria, inflammation and cancer' together, research has made advanced contributions in the interrelatedness of these three areas. It was found that the oral microbiome may both attenuate or help drive the chronic inflammation that often precedes OSCC development and may also have a role in influencing host cell responses to cancer. *Fusobacteria,* in our review were found to have the highest number of active genes in OSCC sites and was also found to be significantly enriched in the buccal microbiome of Toombak users through our research analysis. Whether there is an unknown connection between the known aggression of buccal OSCC [4] tumours and increased *Fusobacteria* in buccal mucosal sites, is an important consideration. On the other hand, we detected genera, such as *Leptotrichia* amongst Toombak users [high-risk of developing OSCC], in contrast to review findings that they are associated with health.

We further illustrated how bacteria might become part of the tumour environment either before or after tumour development and considering that view; we found a differing microbial scenario in those that develop OSCC and are non-users of Toombak [*Lactobacillus, Collinsella* and *Flavobacterium*] and those who do develop OSCC in the context of Toombak use [*Corynebacterium_1, Mogibacterium, Enterococcus, Empedobacter Stenotrophomonas* and *Schlegelella*]. We also found similar results with *Rothia* species that were found to be abundant in premalignant lesions from other studies as well as in our own samples [Chapter 4].

Figure 1:

The Sudabiome in relation to SDGs and multi-target in depth categorisation

Further SDG 3 targets:

- Implement the WHO Framework Convention on Tobacco Control
- Improve early warning systems for global health risks

Further SDG 11 targets:

- Protect the world's cultural and natural heritage
- Strong national and regional development planning

Further SDG 12 targets:

- Support developing countries scientific and technological capacity for sustainable consumption and production
- Sustainable management and use of natural resources

Further SDG 17 targets:

- Mobilise financial resources for developing countries
- Knowledge sharing and cooperation for access to science, technology and innovation

It is estimated that 6-10 million Toombak users exist in Sudan, predominantly males. The use of the smokeless tobacco, Toombak, has been arduously linked to numerous systemic conditions [Table 1], which include but are not limited to, the development of hypertension, oral and pharyngeal cancer, and other forms of cancer, while orally, its use is attributed to the development of Xerostomia [dry mouth] and periodontal disease. Toombak is said to be responsible for up to 81% of oral cancers in Sudan, while the country has the second highest incidence rate of the disease in the Middle East [5-9]. Unfortunately, incidence of oral cancer in Sudan is only expected to double by 2030 in Sudan [10]. The sites of placement of Toombak dip have also been strongly correlated with oral tumour location [11], and thus we achieved a comprehensive community profiling of the oral cavity microbiome of Toombak users at hotspot locations throughout the inner linings of the mouth.

Table 1: Oral and systemic implications of Toombak use

Oral	Halitosis [bad breath] Tooth and gum abrasion Dry mouth Periodontal disease Oral and other head and neck cancers
Systemic	Gastrointestinal disturbances Nutritional deficiencies Nausea Peptic ulcer development Chronic inflammation Diabetes Cardiovascular Chronic hypertension/ increased heart rate Increased mortality after myocardial infarction Reproductive concerns Psychological effects Cancer [pancreas, stomach, liver] Alzheimer's disease development

Currently, the Toombak is produced haphazardly, with the final product exposed to widespread contamination and long-term unregulated upkeep and storage [often under heat]. Furthermore, the selling of Toombak, particularly to the young of age [under 18] needs to be addressed. The findings of the ultra-structural composition of Toombak that included elevated levels of iron, a rich abundance of *Corynebacterium casei* and *Staphylococcus gallinarum*, *Aspergillus* species, *heterocaryoticus* and *austwickii* [12], and potently high levels of tobacco-specific nitrosamines [TSNAs] [1270-4830 mg/kg], acetaldehyde [> 600 mg/kg], and formaldehyde [> 7mg/kg], all helped to attend to one of the most problematic areas in the culture of Sudan, aiming to achieve a driving force of change that may combat these findings in the future.

The genus, *Corynebacterium_1* for example, was noted to comprise the majority of Toombak microbiome composition and was found to be significantly enriched in the oral cavity [buccal inner cheek, the floor of the mouth and saliva] of Toombak users compared to non-users, while also playing a part in the Toombak-OSCC microbiome. This highlights that further research in the field of whole genome sequencing may be relevant to ascribing why *Corynebacterium_1* may have such a triad of connection, which has been highlighted from this project. In addition, *Staphylococcus*, common to the Toombak microbiome and found to be enriched throughout the oral mucosa of Toombak users, may further predispose this group to an increased prevalence of oral and systemic infections as well as antimicrobial resistance which may not normally occur in other cohorts of the Sudanese population.

By understanding the ultra-structural composition of Toombak [Chapters 2A and 2B], structured initiatives in tobacco farming and production, industry, and research, can be looked upon, to improve quality and safety of the final Toombak product sold to consumers. Farming initiatives include disinfection of tobacco leaf, use of other *Nicotiana* plant species other than *Rustica*, and finding alternatives to soil with a heavy metal content. During the production of Toombak, improved sanitation and ventilation during the fermentation process and use of refrigeration during storage are necessary incorporations to reduce bacterial, fungal and TSNA content. Meanwhile, industry initiatives include the introduction of wide-scale pasteurisation and irradiation methods [13], while research initiatives include plant genetic and molecular

approaches to lower TSNA precursors in tobacco plant as well as whole genome sequencing that targets the understanding of bacterial species specifically contributing to the toxic products in Toombak [14].

Furthermore, it appears that Toombak use significantly impinges on long-term cortisol production in the body [Chapter 2C], allowing new contribution to the understanding of how use of this smokeless tobacco alters systemic health. The findings of enhanced addiction [>85% of Toombak users indicating high scores in psycho-dependency FTND-ST questionnaire] and cortisol blunting [mean hair cortisol concentrations found to be 9.7 pg/ml in Toombak users compared to 19.4 pg/ml in non-users], could pave the path for future studies to address related surgical outcomes in Toombak users and quality of life parameters of this cohort. Other clinical initiatives include the development of oral microbiome biomarkers which can aid in the early detection of cancerous lesions, and oral cancers that are more aggressive or have an increased chance of metastasis and recurrence, in Toombak users. Socially, the translation of current scientific findings in an understandable way for Sudanese Toombak users may contribute to modifying social responses against Toombak use, while new public health measures could be made active for the successful rehabilitation after quitting, due to this research.

Chapter 3 aimed to merge the SDG aims of this project with the evaluation of traditional fermented foods known to enrich the health of Sudanese communities. 16s rRNA sequencing and ITS fungal sequencing were achieved on the largest subset of indigenous fermented foods from Sudan to date, offering invaluable insight into how the microorganisms in these foods contribute to affordable food manufacture, and are a source of nutrient refinement, energy production and food preservation. We found first-time sequenced foods such as ‘Kawal’ to be enriched in *Bacillus, Lactobacillus, Pediococcus, Citrobacter* and *Bifidobacterium* as well as harbouring an abundant source of fungal species. However, it was also noted that several pathogenic bacteria and fungi reside in these foods, which may contribute to illness and deprivation of health [*Escherichia-Shigella* found abundantly in the plant-based foods and *Wohlfahrtiimonas* in animal-based foods], while also contributing to inferior food quality and early undesired spoilage.

This research has observed the modern Sudanese gut microbiome in two landscapes, the traditional and migratory, highlighting metagenomic aspects that harbour complex translations to health [Chapter 5]. Furthermore, comparing the Sudanese population in a dual breakdown based on migration as well as to the Irish population was a novel undertaking. Evaluating how migration to Ireland can sculpt the gut microbiota widens the understanding of how geographical, dietary, cultural, social, and other factors come into play in distinguishing the systemic health of a population. We found that the microbiome of the Sudanese living in Ireland undergoes significant changes after migration with both loss and gains in microbial abundance while also attempting to maintain position in several 'traditional' microbiome characteristics. These microbial edits offer vital insights into how susceptibilities and protections to human disease play out in the Sudanese population and can help expand upon future research that aims to survey, prevent, manage, or treat any condition in this cohort.

Furthermore, we modified a validated food frequency questionnaire [EPIC-Norfolk FFQ] to suit the diets of the Sudanese population, through the addition of twenty-four Sudanese- related foods, and concluded that nutrient intake could be monitored effectively through this method amongst the Sudanese and Irish. Nutrient-genera correlations provided insight into how the microbiome relates to nutrient outcomes where the genus *Megamonas* was the most positively associated genus in Sud and several nutrients including calcium, protein, and fat intakes. It is imperative, therefore, to continue the journey on Sudanese food patterns that can further reflect on bioactive metabolic pathways and host- microbe cross talk.

Expanding upon the findings from Sudanese gut metagenomics, we found a microbiome composition that may predispose and protect against specific diseases. Obesity in the Sudanese, could be associated with abundances of *Collinsella* [a genus strongly associated with obesity], *Lactobacillus* [reported to harbour both anti [15] and pro-obesity [16] effects] and *Bifidobacterium* [primarily linked to a protective effect against the disease [17]]. In this study, the reduced relative abundance of genera that produce short-chain fatty acids may however highlight a reduced predisposition to obesity amongst the Sudanese [18].

The reduced abundance of *Lachnospiraceae* in those living in Sudan in particular, may be either positively [19] or negatively [20] associated with the development of type two diabetes. Furthermore, reduced *Lachnospiraceae* amongst the Sudanese may be linked with the development and progression of inflammatory bowel disease [21] although, it may also offer a protective effect against liver diseases such as primary sclerosing cholangitis and non-alcoholic fatty liver disease [22]. *Collinsella* enrichment in Sudan, may further contribute to the development of non-alcoholic fatty liver disease [22] through the formation of oxobile acids that increase intestinal permeability, while at the same time the *Alistipes* depletion associated with Sudanese guts may allow the progression of liver diseases such as cirrhosis and hepatic encephalopathy [23] and chronic kidney damage [24]. The reduced abundance of *Ruminiclostridium* in the Sudanese may further be associated with the occurrence of nephrolithiasis [25].

Prevotella and *Collinsella*, both found to be abundant in the Sudanese population, are also known to be abundant in those with type two diabetes [22], while *Akkermansia* [abundant in the Sudanese living in Ireland], has been shown to be both enriched [26] but also depleted [19] concerning the development of type two diabetes. In addition, *Prevotella* abundance may protect against the development of Inflammatory bowel disease [27] and colon cancer [28], while enrichment with *Collinsella* may infer the opposite association [29]. Furthermore, the increased abundance of *Collinsella,* evident in the Sudanese gut microbiome, is positively correlated with the development of rheumatoid arthritis, likely due to human leukocyte antigen mimicry through similarities with the DRB1*0401 allele. At the same time, *Lactobacillus* has been shown to be reduced [30] and increased in those with this disease [31]. The Sudanese may have increased protection from the development of multiple sclerosis due to the abundance of *Collinsella* [32] [33] and reduced levels of *Blautia* and *Dorea* [20] reported in this study. At the same time, *Collinsella* abundance has also been associated with reduced mortality rates from Covid-19 [34].

We further found several microbial distinctions that may predispose the Sudanese towards the development of Crohn's disease and recurring states of ulcerative colitis, which included the decreased abundance of the genera, *Roseburia, Coprococcus, Dorea*, *Blautia*, *Ruminococcus* and *Odoribacter* in those living in Sudan [20, 35, 36].

Prevotella abundance may further predispose to irritable bowel syndrome [37], while on the contrary, *Parasutterella*, which was found to be low in those living in Sudan has been shown to be abundant amongst those with irritable bowel syndrome [38].

Further genera increased in abundance in both Sudanese populations and also associated with the development of rheumatoid arthritis included the abundance of the phylum *Actinobacteria* and the genera *Streptococcus*, *Prevotella* and *Alloprevotella* [30] [39]. Abundances of *Dialister* and *Klebsiella*, were also evident in the Sudanese and have been strongly associated with the development of spondylarthritis [40].

Furthermore, several species were seen to become enriched after migration, a trend that may have surprising conclusions on the health of the Sudanese. Our findings would suggest that those living in Sudan for example, are the least at risk of developing type two diabetes compared to those who live in Ireland [41]. Species that may promote type two diabetes includes *Butyribacter intestini*, a butyric acid and vitamin B6 producer, which was found to be absent in those living in Sudan but elevated in both populations living in Ireland.

The Sudanese may also have protection against depression in part due to their gut microbiome composition. The high abundance of *Dialister* found in this cohort may offer protection against this disease [42]. High abundances of *Klebsiella*, *Prevotella* and *Streptococcus* also seen in the Sudanese, are however associated with depressive states [43] [44].

We endeavour to utilise these research data as a stepping platform for targeting new areas such as the screening and diagnosis of non-communicable diseases, the prevention and early detection of oral cancer and the improvement of child and senior citizen quality of life. Other facets of this research include building relations between institutions and developing vital capacity for establishing metagenomic infrastructure in Sudan so that further research can be more independently obtained.

In Sudan, there has often been poor research implementation due to economic instability and continuous shortage of research expenditure budget. However, it is essential to propound that microbiome science and economic value go hand in hand, where knowledge of the microbiome in a population and its environment can substantially help improve economic statuses at a personal and national level.

Furthermore, through this research we acknowledge a wide array of interplay between diet, host physiology, endemic disease, and socioeconomic concerns that have never been questioned before. Therefore, data from this research is influential to the Sudanese population but will contribute to the evolving Irish microbiome space, adding novel and interesting insights that can challenge and flourish current knowledge.

Through the Sudabiome, a new commitment to research is employed for the improvement of the health of a population. The Sudabiome harbours the future intention that metagenomics in Sudan rapidly moves from an offside position to a central component in the prevention, screening, diagnosis, and treatment of diseases in the Sudanese population, while Table 2 highlights branching research topics of interest from this project. Of importance for example, is the development of a Sudanese microbiome biobank from Sudanese fermented foods as well as disease-associated states from the mouth and gastrointestinal tract that may be of interest in the early screening of diseases as well as management orientations. The studies presented here aim to formalise a concrete future towards the prevention, management, and treatment of chronic disease to find ways to promote life expectancy for many of the impoverished population of Sudan. However, through this project, Irish research can flourish through new understandings of the microbiome from varied genetic backgrounds, dietary feedback, and new data correlations.

Table 2: Future research projects that take precedence from current results chapters of Sudabiome project

Sudabiome area of research	**Differing types of future pathways that can be undertaken**
• The ultrastructural metabolomic and metagenomic characterisation of the Sudanese smokeless tobacco Toombak • The Mycobiome of Toombak • Sudanese smokeless tobacco [Toombak] users harbour significantly altered long-term cortisol body production	1) Analysing differing stages of Toombak production to fully understand the stages of microbial entry 2) Whole genome sequencing of Toombak for species and strain analysis 3) Metatranscriptomics for the understanding of active DNA replication within Toombak 4) Longitudinal studies analysing cortisol production from prior and early Toombak use as well as after cessation recovery levels of hormone
The fermented foods of Sudan through a metagenomic lens	1) Economic nutrition – Impact of fermented food consumption to be further targeted in order to highlight their impact on the gut microbiome. Longitudinal follow up of 'first time' consumers of these foods can provide invaluable knowledge on the effects of these foods on git microbiome alterations 2) 'Sudanese Golden Foods' project and Sudanese nutrition online database profiling – first time online database of biochemical and phytochemical analysis of the unique foods of Sudan,

	including fermented foods for the accessibility of national and international use in research, nutrition science and medicine 3) Sudan biobank – whole genome sequencing can offer a vast opportunity to discover novel species and fungi that can harbour many potential scientific opportunities 4) Modern fermentabilistics of Sudan – transforming current tradition into modernised industrial production such as 'Kawal' into a palatable probiotic supplement or 'Garis' camel milk into a pasteurised yet rich source of safe nourishment 5) Nutritional epidemiology at a large scale to pattern micro and macro nutrient deficiency traits amongst various cohorts of susceptible age groups and in differing locations, all the while characterising gut microbiome with critical topics in Sudan such as malnutrition and basic need requirements 6) Drinking water microbiome
Altered oral microbiome in Sudanese smokeless tobacco users carries a newly emerging risk of squamous cell carcinoma development and progression	1) Pre-oral microbiome assessment prior to Toombak use with longitudinal follow up on microbiome alterations after Toombak use

	2) Questioning susceptibility taxonomics for oral cancer, progression, aggression, and recurrence
Traditional memory and reshaping of Sudanese gut metagenomics in relation to dietary impact and migratory movements to westernised countries	1) Inclusion of other African neighbouring countries as a first-time initiative for the building of an African gut microbiome databank. 2) Sudanese demographic population gut microbiome mapping for disease control and prevention – a distinct pathway for African human microbiome mapping and disease correlations through biomarkers. 3) Microbiome biomarker distinctions in a wide range of urban and rural [village, nomadic and tribal] settings in Sudan 4) Anti-microbial resistance – analysing genetic antibiotic resistance profiles in commensal gut bacteria that may be key drivers in the spread of antibiotic resistance genes

References

1. Alkire S, Kanagaratnam U, Suppa N. 2022. The global multidimensional poverty index [mpi] 2022 country results and methodological note.
2. Ballon P, Duclos J-Y. 2015. Multidimensional poverty in Sudan and south Sudan.
3. Sami A, Elimairi I, Stanton C, Ross RP, Ryan CA. 2020. The Role of the Microbiome in Oral Squamous Cell Carcinoma with Insight into the Microbiome-Treatment Axis. Int J Mol Sci 21.
4. Lubek JE, Dyalram D, Perera EH, Liu X, Ord RA. 2013. A retrospective analysis of squamous carcinoma of the buccal mucosa: an aggressive subsite within the oral cavity. Journal of oral and maxillofacial surgery 71:1126-1131.
5. Idris AM, Ibrahim SO, Vasstrand EN, Johannessen AC, Lillehaug JR, Magnusson B, Wallström M, Hirsch JM, Nilsen R. 1998. The Swedish snus and the Sudanese toombak: are they different?, p 558-66, Oral Oncol, vol 34, England.
6. Critchley JA, Unal B. 2003. Health effects associated with smokeless tobacco: a systematic review. Thorax 58:435-43.
7. Sunanda P, Panda B, Dash C, Ray PK, Padhy RN, Routray P. 2014. Prevalence of abnormal spermatozoa in tobacco chewing sub-fertile males. J Hum Reprod Sci 7:136-42.
8. Elbeshir EI, Abeen HA, Idris AM, Abbas K. 1989. Snuff dipping and oral cancer in Sudan: A retrospective study. British Journal of Oral and Maxillofacial Surgery 27:243-248.
9. Boffetta P, Hecht S, Gray N, Gupta P, Straif K. 2008. Smokeless tobacco and cancer. Lancet Oncol 9:667-75.
10. Kujan O, Farah CS, Johnson NW. 2017. Oral and oropharyngeal cancer in the Middle East and North Africa: Incidence, mortality, trends, and gaps in public databases as presented to the Global Oral Cancer Forum. Translational Research in Oral Oncology 2:2057178X17698480.
11. Hassanin AA, Idris AM. 2017. Attribution of oral cancer in the Sudan to Toombak dipping. Translational Research in Oral Oncology 2:2057178X16685729.

12. Sami A, Elimairi I, Patangia D, Watkins C, Ryan CA, Ross RP, Stanton C. 2021. The ultra-structural, metabolomic and metagenomic characterisation of the sudanese smokeless tobacco 'Toombak'. Toxicol Rep 8:1498-1512.
13. Warke RG, Kamat AS, Kamat MY. 1999. Irradiation of chewable tobacco mixes for improvement in microbiological quality. Journal of food protection 62:678-681.
14. Pickworth WB. 2020. Smokeless Tobacco Products: Characteristics, Usage, Health Effects, and Regulatory Implications. Elsevier.
15. Kim S, Huang E, Park S, Holzapfel W, Lim S-D. 2018. Physiological characteristics and anti-obesity effect of Lactobacillus plantarum K10. Korean journal for food science of animal resources 38:554.
16. Million M, Maraninchi M, Henry M, Armougom F, Richet H, Carrieri P, Valero R, Raccah D, Vialettes B, Raoult D. 2012. Obesity-associated gut microbiota is enriched in Lactobacillus reuteri and depleted in Bifidobacterium animalis and Methanobrevibacter smithii. International journal of obesity 36:817-825.
17. Da Silva CC, Monteil MA, Davis EM. 2020. Overweight and obesity in children are associated with an abundance of Firmicutes and reduction of Bifidobacterium in their gastrointestinal microbiota. Childhood Obesity 16:204-210.
18. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, Mele MC. 2019. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. Microorganisms 7:14.
19. Brunkwall L, Orho-Melander M. 2017. The gut microbiome as a target for prevention and treatment of hyperglycaemia in type 2 diabetes: from current human evidence to future possibilities. Diabetologia 60:943-951.
20. Vacca M, Celano G, Calabrese FM, Portincasa P, Gobbetti M, De Angelis M. 2020. The Controversial Role of Human Gut Lachnospiraceae. Microorganisms 8:573.
21. Qiu Z, Yang H, Rong L, Ding W, Chen J, Zhong L. 2017. Targeted metagenome based analyses show gut microbial diversity of inflammatory bowel disease patients. Indian journal of microbiology 57:307-315.

22. Astbury S, Atallah E, Vijay A, Aithal GP, Grove JI, Valdes AM. 2020. Lower gut microbiome diversity and higher abundance of proinflammatory genus Collinsella are associated with biopsy-proven nonalcoholic steatohepatitis. Gut Microbes 11:569-580.
23. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. 2020. The Genus Alistipes: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. Front Immunol 11:906.
24. Kim JE, Kim HE, Park JI, Cho H, Kwak MJ, Kim BY, Yang SH, Lee JP, Kim DK, Joo KW, Kim YS, Kim BS, Lee H. 2020. The Association between Gut Microbiota and Uremia of Chronic Kidney Disease. Microorganisms 8.
25. Tang R, Jiang Y, Tan A, Ye J, Xian X, Xie Y, Wang Q, Yao Z, Mo Z. 2018. 16S rRNA gene sequencing reveals altered composition of gut microbiota in individuals with kidney stones. Urolithiasis 46:503-514.
26. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, et al. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490:55-60.
27. Mobeen F, Sharma V, Tulika P. 2018. Enterotype Variations of the Healthy Human Gut Microbiome in Different Geographical Regions. Bioinformation 14:560-573.
28. Ou J, Carbonero F, Zoetendal EG, DeLany JP, Wang M, Newton K, Gaskins HR, O'Keefe SJD. 2013. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. The American journal of clinical nutrition 98:111-120.
29. Rehman A, Lepage P, Nolte A, Hellmig S, Schreiber S, Ott SJ. 2010. Transcriptional activity of the dominant gut mucosal microbiota in chronic inflammatory bowel disease patients. Journal of medical microbiology 59:1114-1122.
30. Sun Y, Chen Q, Lin P, Xu R, He D, Ji W, Bian Y, Shen Y, Li Q, Liu C, Dong K, Tang Y-W, Pei Z, Yang L, Lu H, Guo X, Xiao L. 2019. Characteristics of

Gut Microbiota in Patients With Rheumatoid Arthritis in Shanghai, China. Frontiers in cellular and infection microbiology 9:369-369.

31. Liu X, Zou Q, Zeng B, Fang Y. 2013. Analysis of Fecal Lactobacillus Community Structure in Patients with Early Rheumatoid Arthritis. Current microbiology 67.
32. Bernstein CN, Forbes JD. 2017. Gut microbiome in inflammatory bowel disease and other chronic immune-mediated inflammatory diseases. Inflammatory intestinal diseases 2:116-123.
33. Olendzki B, Bucci V, Cawley C, Maserati R, McManus M, Olednzki E, Madziar C, Chiang D, Ward DV, Pellish R. 2022. Dietary manipulation of the gut microbiome in inflammatory bowel disease patients: Pilot study. Gut microbes 14:2046244.
34. Hirayama M, Nishiwaki H, Hamaguchi T, Ito M, Ueyama J, Maeda T, Kashihara K, Tsuboi Y, Ohno K. 2021. Intestinal Collinsella may mitigate infection and exacerbation of COVID-19 by producing ursodeoxycholate. PloS one 16:e0260451-e0260451.
35. Biddle A, Stewart L, Blanchard J, Leschine S. 2013. Untangling the Genetic Basis of Fibrolytic Specialization by Lachnospiraceae and Ruminococcaceae in Diverse Gut Communities. Diversity 5:627-640.
36. Hiippala K, Barreto G, Burrello C, Diaz-Basabe A, Suutarinen M, Kainulainen V, Bowers JR, Lemmer D, Engelthaler DM, Eklund KK, Facciotti F, Satokari R. 2020. Novel Odoribacter splanchnicus Strain and Its Outer Membrane Vesicles Exert Immunoregulatory Effects in vitro. Frontiers in Microbiology 11.
37. Su T, Liu R, Lee A, Long Y, Du L, Lai S, Chen X, Wang L, Si J, Owyang C, Chen S. 2018. Altered Intestinal Microbiota with Increased Abundance of Prevotella Is Associated with High Risk of Diarrhea-Predominant Irritable Bowel Syndrome. Gastroenterology Research and Practice 2018:6961783.
38. Chen YJ, Wu H, Wu SD, Lu N, Wang YT, Liu HN, Dong L, Liu TT, Shen XZ. 2018. Parasutterella, in association with irritable bowel syndrome and intestinal chronic inflammation. J Gastroenterol Hepatol 33:1844-1852.

39. Chen J, Wright K, Davis JM, Jeraldo P, Marietta EV, Murray J, Nelson H, Matteson EL, Taneja V. 2016. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. Genome medicine 8:43-43.
40. Cock IE, van Vuuren SF. 2015. The potential of selected South African plants with anti-Klebsiella activity for the treatment and prevention of Ankylosing spondylitis. Inflammopharmacology 23:21-35.
41. Zou Y, Xue W, Lin X, Lv M, Luo G, Dai Y, Sun H, Liu SW, Sun CH, Hu T, Xiao L. 2021. Butyribacter intestini gen. nov., sp. nov., a butyric acid-producing bacterium of the family Lachnospiraceae isolated from human faeces, and reclassification of Acetivibrio ethanolgignens as Acetanaerobacter ethanolgignens gen. nov., comb. nov. Syst Appl Microbiol 44:126201.
42. Valles-Colomer M, Falony G, Darzi Y, Tigchelaar EF, Wang J, Tito RY, Schiweck C, Kurilshikov A, Joossens M, Wijmenga C, Claes S, Van Oudenhove L, Zhernakova A, Vieira-Silva S, Raes J. 2019. The neuroactive potential of the human gut microbiota in quality of life and depression. Nature Microbiology 4:623-632.
43. Barandouzi ZA, Starkweather AR, Henderson WA, Gyamfi A, Cong XS. 2020. Altered Composition of Gut Microbiota in Depression: A Systematic Review. Front Psychiatry 11:541.
44. Lin P, Ding B, Feng C, Yin S, Zhang T, Qi X, Lv H, Guo X, Dong K, Zhu Y, Li Q. 2017. Prevotella and Klebsiella proportions in fecal microbial communities are potential characteristic parameters for patients with major depressive disorder. Journal of Affective Disorders 207:300-304.

Appendices

Appendix 1

International Journal of
Molecular Sciences

MDPI

Review

The Role of the Microbiome in Oral Squamous Cell Carcinoma with Insight into the Microbiome–Treatment Axis

Amel Sami [1,2], Imad Elimairi [2,*], Catherine Stanton [1,3], R. Paul Ross [1] and C. Anthony Ryan [4]

1 APC Microbiome Ireland, School of Microbiology, University College Cork, Cork T12 YN60, Ireland; amelsami1@hotmail.com (A.S.); catherine.stanton@teagasc.ie (C.S.); p.ross@ucc.ie (R.P.R.)
2 Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, National Ribat University, Nile Street, Khartoum 1111, Sudan
3 Teagasc Food Research Centre, Moorepark, Fermoy, Cork P61 C996, Ireland
4 Department of Paediatrics and Child Health, University College Cork, Cork T12 DFK4, Ireland; tonyryan007@gmail.com
* Correspondence: elimairi.imad@gmail.com

Received: 30 August 2020; Accepted: 12 October 2020; Published: 29 October 2020

check for updates

Abstract: Oral squamous cell carcinoma (OSCC) is one of the leading presentations of head and neck cancer (HNC). The first part of this review will describe the highlights of the oral microbiome in health and normal development while demonstrating how both the oral and gut microbiome can map OSCC development, progression, treatment and the potential side effects associated with its management. We then scope the dynamics of the various microorganisms of the oral cavity, including bacteria, mycoplasma, fungi, archaea and viruses, and describe the characteristic roles they may play in OSCC development. We also highlight how the human immunodeficiency viruses (HIV) may impinge on the host microbiome and increase the burden of oral premalignant lesions and OSCC in patients with HIV. Finally, we summarise current insights into the microbiome–treatment axis pertaining to OSCC, and show how the microbiome is affected by radiotherapy, chemotherapy, immunotherapy and also how these therapies are affected by the state of the microbiome, potentially determining the success or failure of some of these treatments.

Keywords: oral microbiome; gut microbiome; oral squamous cell carcinoma; head and neck cancer; microbial

1. Introduction

In the Global Cancer report 2018, head and neck cancer is thought to be the eighth most commonly occurring cancer in the world [1], with oral squamous cell carcinoma being the most common presentation. According to The International Classification of Diseases for Oncology, World Health Organisation, oral cavity cancer includes the locations of the inner lip (C00.3–C00.9), the tongue, excluding lingual tonsil (C02), the gum (C03), the floor of the mouth (C04), the palate (C05), and other or unspecified parts of the mouth (C06) [2].

Although, in some parts of the world, there is a decreasing trend of oral squamous cell carcinoma (substituted, in part, by the increasing trend of human papilloma virus-associated oropharyngeal cancer) [3], incidence of the disease is very much still relevant all over the world. In Africa, oral squamous cell carcinoma statistics are highly underreported, but, in a review by Faggons et al., 2015, the smokeless tobacco products Toombak and Kola nut, used in Sudan and Western African countries, respectively, were the cause of the greatest number of head and neck cancer cases on the continent [4]. Meanwhile, South East and Central Asia accounts for almost half of oral squamous cell carcinoma

Appendix 2

Toxicology Reports 8 (2021) 1498–1512

Contents lists available at ScienceDirect

Toxicology Reports

journal homepage: www.elsevier.com/locate/toxrep

The ultra-structural, metabolomic and metagenomic characterisation of the sudanese smokeless tobacco 'Toombak'

Amel Sami [a,c], Imad Elimairi [c], Dhrati Patangia [a], Claire Watkins [a], C. Anthony Ryan [d], R. Paul Ross [a], Catherine Stanton [a,b,*]

[a] APC Microbiome Institute, University College Cork, Cork, T12 YN60, Ireland
[b] Teagasc Food Research Centre, Moorepark, Fermoy, Cork, P61 C996, Ireland
[c] Department of Oral and Maxillofacial Surgery and Oral Medicine, Faculty of Dentistry, National Ribat University, Nile Street, Khartoum, 1111, Sudan
[d] Department of Paediatrics and Child Health, University College Cork, Cork, T12 DFK4, Ireland

ARTICLE INFO

Handling Editor: Aristidis Tsatsakis

Keywords:
Toombak
Smokeless tobacco
Sudan
pH
Microbiome
Metabolome
Heavy metal
Scanning electron microscopy
Composition

ABSTRACT

Toombak is a smokeless tobacco produced from the *Nicotiana rustica* tobacco plant from Sudan. Pre-prepared and ready to buy Toombak samples were analysed using mass spectrometry (heavy metals), gas and liquid chromatography (metabolomics), 16S rRNA metagenomic sequencing (microbiome) and Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt), scanning electron microscopy with energy dispersive X-ray spectroscopy (SEM-EDX) and pH analysis.

Chromium, cobalt, and copper were high in the pre-prepared form of Toombak while iron, tobacco specific nitrosamines (TSNAs), formaldehyde and acetaldehyde were high in both types. *Firmicutes* and *Actinobacteria* dominated Toombak. Samples of ready to buy Toombak showed inter-variational differences depending on place of purchase. We found *Virgibacillus* were increased in the pre-prepared form while *Corynebacterium casei*, *Atopococcus tabaci*, *Atopobium suicloacalis*, *Oceanobacillus chironomi* and *Staphylococcus gallinarum* were the most abundant species in the ready to buy forms. PICRUSt analysis highlighted increased activity of metal transport systems in the ready to buy samples as well as an antibiotic transport system. SEM-EDX highlighted large non-homogenous, irregular particles with increased sodium, while pH of samples was in the alkaline range.

The final composition of Toombak is affected by its method of preparation and the end product has the potential to impart many negative consequences on the health of its users. TSNA levels observed in Toombak were some of the highest in the world while the micro-environment of Toombak supports a distinct microbiota profile.

1. Introduction

Smokeless tobacco refers to the use of unburnt or pyrolysis free tobacco, differing around the world in tobacco plant source, nicotine content and additives. Products are often applied orally (chewed, sucked, or placed passively to the gingivae) and include forms such as chewing tobacco, moist forms, dry snuff, and snus, while fine powdered types are administered nasally. It is thought that more than 300 million people use a form of smokeless tobacco worldwide, in over 70 countries that include Southeast Asia, many regions of Africa, the Middle East, Europe (Scandinavian countries), the Americas and Russia [1]. In Sudan, there are an estimated 4–10 million users of Toombak; a moist smokeless tobacco used primarily by males. Idris et al. (1998) concluded that 60 % of Sudanese households contain at least one family member with Toombak use, while it has been estimated that Toombak prevalence of use amongst adolescents, young adults and those above 60 years of age is; 34 %, 32 % and 47 % respectively [2].

Toombak is produced from the *Solanaceae* species, *Nicotiana rustica*, a plant with characteristic yellow flowers, that contains up to nine times more nicotine compared to *Nicotiana tabacum*, that is more widely utilised in the production of tobacco products worldwide [3]. *Nicotiana rustica* is also used to make smokeless tobacco products in Turkey (Maras powder) [4], South America (Rapé) [5], Vietnam (Thuốc lào) and Russia (Makhorka). Plantation begins around November in western Sudan, leaves are cultivated from February to March, and air curing occurs during April and May. During this latter stage, temperatures can reach 45 °C-60 °C. Leaves change from yellow to brown in a process of natural or 'compost' like fermentation and the Toombak is then further prepared

* Corresponding author at: APC Microbiome Institute and Teagasc Food Research Centre, Moorepark, Fermoy, Cork, P61 C996, Ireland.
E-mail address: Catherine.stanton@teagasc.ie (C. Stanton).

https://doi.org/10.1016/j.toxrep.2021.07.008
Received 31 March 2021; Received in revised form 8 June 2021; Accepted 7 July 2021

Appendix 3

The threat of Toombak in Sudan

Researchers from Ireland and Sudan have partnered up to shed light on the health implications of one of the world's most perilous types of smokeless tobacco.

Toombak is a type of popular smokeless tobacco sold by local vendors in Sudan. Most commonly used by men, it's placed inside the mouth using the fingertips and replaced up to several times a day. Whilst it's never swallowed, the long-term use of Toombak can lead to many serious and life-threatening issues, such as gum disease, oral cancer, cardiovascular disease, prostate cancer and infertility.

Toombak is produced by fermenting the leaves of the *Nicotiana rustica* tobacco plant, in a process that can take months. *Nicotiana rustica* is used because of its rich nicotine content, and numerous additives are added along the way. Sudanese people have found a way to increase the kick even more by adding sodium bicarbonate to the mix.

Sodium bicarbonate creates an alkaline environment that allows for high absorption rates of nicotine through the mouth lining. This makes the habit extremely addictive, and users find it exceedingly difficult – if not impossible – to quit.

Over the last three years, researchers from Teagasc, research centre APC Microbiome Ireland – based in University College Cork (UCC) – and National Ribat University Sudan have come together to explore in more detail the chemical and microbial (microorganism characteristic)

Appendix 4

Steroids 193 (2023) 109189

Contents lists available at ScienceDirect

Steroids

journal homepage: www.elsevier.com/locate/steroids

Sudanese Toombak smokeless tobacco users harbour significantly altered long–term cortisol body production

Amel Sami [a,c], Imad Elimairi [c], C. Anthony Ryan [d], R. Paul Ross [a], Catherine Stanton [a,b,*]

[a] APC Microbiome Ireland, School of Microbiology, University College Cork, Cork T12 YN60, Europe
[b] Teagasc Food Research Centre, Moorepark, Fermoy, Cork P61 C996, Ireland
[c] Department of Oral and Maxillofacial Surgery and Oral Medicine, Faculty of Dentistry, National Ribat University, Nile Street, Khartoum 1111, Sudan
[d] Department of Paediatrics and Child Health, University College Cork, Cork T12 DFK4, Europe

ABSTRACT

The Sudanese, in particular its male population, are known to utilise a smokeless tobacco product (Toombak) which is placed in the oral cavity and can be replaced several times a day. Toombak has been shown to harm human health and is highly addictive. The effect on body cortisol response over a retrospective period in users of this product has not been previously explored. In addition, psycho-dependency scores of Toombak users have not been analysed.

In this study, 37 male subjects, age 18–45 years were recruited, of which 18 were non-users of Toombak and 19 were Toombak users. One hair sample was collected from each user and non-user of Toombak. Each hair sample (n=37) was placed in a pre-prepared long piece of foil with two labels on either side marked: 'scalp-side' and 'distant-side'. Cortisol was extracted by mincing 10 mg of 'scalp-side' hair, not exceeding 3 cm, with methanol addition, incubation, and sonication. Cortisol was measured using the enzyme-linked immunosorbent assay kit (Enzo Life Sciences, UK). The amount of hair cortisol in the samples was determined using spectrophotometry at wavelength 405 nm measured in pg/ml and visualised with a four parametric logistic curve. Toombak users were further asked to complete the Fagerstrom Test for Nicotine Dependence-Smokeless Tobacco questionnaire (FTND-ST) comprising of six questions. Scores of > 5 indicated a significant dependence, while a score of < 4 marked low to moderate dependence.

The mean concentration of hair cortisol in Toombak users (9.7 pg/ml) was significantly lower (p=0.023) compared to non-users (19.4 pg/ml), with total concentrations ranging from 2.1 to 55.6 pg/ml. FTND–ST scores ranged from 4 to 9, with high levels of psycho-dependency (score > 5) and nicotine tolerance found in 85 % of Toombak users. Cortisol body release in Sudanese smokeless tobacco users was found to be significantly altered. While low cortisol levels do lead to anxiolytic effects, in the long-term, this can allow for increased susceptibility to low cortisol-associated diseases.

1. Introduction

Cortisol is a glucocorticoid hormone that is secreted from the adrenal cortex. It is vital in gluconeogenesis and in mediating immunity and reducing inflammation [1]. Cortisol release occurs in a circadian pattern and is regulated by the hypothalamus pituitary adrenal axis. Corticotropin-releasing hormone is released from the paraventricular nucleus of the hypothalamus after which adrenocorticotropic hormone is secreted from the pituitary gland. This mechanism allows the secretion of cortisol to occur through a negative feedback loop whereby in the field of psychoneuroendocrinology [3].

Hair cortisol analysis is a non-invasive method of adrenal gland function in the body [4]. It reflects well, the hypothalamus pituitary adrenal axis [5] and average blood cortisol levels over a broad retrospective period [6]. Utilising hair for cortisol hormone analysis also offers the advantage of the measurements being unaffected by the pulsatile release of cortisol and the psychological and physical mediators on cortisol release such as exercise, trauma, and nicotine usage [7,8].

In healthy individuals, hair cortisol levels have been found to range between 17.7 and 152.7 pg/mg [9], 12–162 pg/ml [10] and 2–51.63

Appendix 5

THE BOOLEAN
SNAPSHOTS OF DOCTORAL RESEARCH AT UCC

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Sudanese Fermented Foods

Amel Sami[1,3], Imad Elimairi[3], Paul Ross[1], Marmar Elsiddig[4], Yasmeen Elyass[4], Heyam Salih[4], and Catherine Stanton[1,2,*]

[1]*APC Microbiome Ireland, School of Microbiology, University College Cork, Cork, T12 YN60, Ireland*

[2]*Teagasc Food Research Centre, Moorepark, Fermoy, Cork, P61 C996, Ireland*

[3]*Department of Oral and Maxillofacial Surgery and Oral Medicine, Faculty of Dentistry, National Ribat University, Khartoum, 1111, Sudan*

[4]*University of Khartoum, Sudan*

[*]*Corresponding author: catherine.stanton@teagasc.ie*

Abstract

Fermented foods of Sudan are a great source of affordable daily nutrition for many families and age groups all over the country. These foods include a diverse categorisation of starting ingredients and incur traditional methodologies of production which have been preserved for centuries. We used next generation sequencing (16S rRNA and 18S rRNA) to analyse 44 Sudanese Fermented foods in five major categories food types; sorghum, plant, meat, fish and dairy. Samples were collected in Khartoum, Sudan and analysed in Cork, Ireland. We found

Appendix 5

First oral microbiome characterisation amongst the Sudanese population and how use of the smokeless tobacco; Toombak can influence the comrade of oral microorganism life

Amel Sami, Imad Elmairi, R. Paul Ross, C Anthony Ryan, and Catherine Stanton

APC microbiome institute Ireland, University College Cork, Teagasc research centre, Moorepark, Fermoy, Co. Cork and National Ribat University, Khartoum Sudan

Background

In this study we elucidate how oral mucosal interfaces and biofilm of hard tooth surfaces (plaque) are affected by the smokeless tobacco; Toombak use amongst the Sudanese population.

Materials and methods

We achieved 16S rRna metagenomic sequencing of 73 pooled saliva, 71 plaque and 272 mucosal swabs using Puritan Diagnostic Swabs (Figure 1) from four locations in 78 participants (47 Toombak and 31 non-users) aged between 20-70 years. Two oral keratinised (dorsum tongue, palate) and two non- keratinised locations (floor of the mouth, buccal mucosa) were chosen.

Figure 1

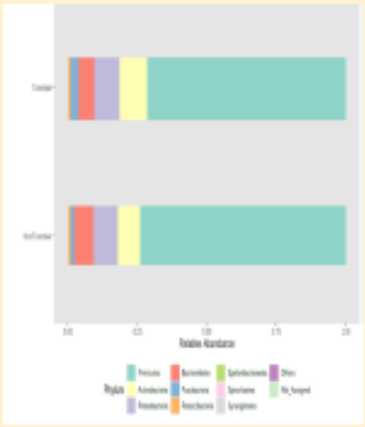


Figure 2

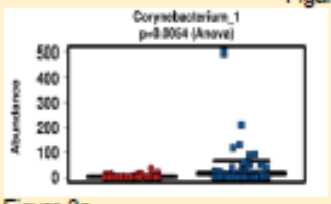


Figure 3a

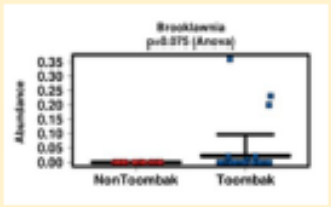


Figure 3b

Figure 4

Results and discussion

Streptobacillus, Shuttleworthia and Eubacterium yuri were unique to non-user saliva while Corynebacterium_1, Alysiella and others were found in Toombak placers. There were increased abundances of *Fusobacteria* (Figure 2), *Patescibacteria* and *Spirochaetia*, in saliva of Toombak users with higher counts of *Rothia* and *Streptococcus*. *Prevotella* were promoted in non-users and *Staphylococcus* and *Corynebacterium_1 (Figure 3a)* in Toombak users. In plaque, *Comamonas, Brooklawnia (Figure 3b)* and *Peptostreptococcus* distinguished those that dip Toombak with non-users having a more robust plaque biofilm protection. Palatal-user and tongue microbiome were lowest in alpha diversity while B diversity was significant only in relation to mucosal habitat (Figure 4) rather than Toombak use. Keratinised and non- keratinised mucosal locations followed a similar species abundance with *Prevotella*, *Veillonella* and *Lachnoanaerobaculum* more abundant in the former while *Gemella* and *Haemophilus* were more profound in the latter.

Conclusion

The oral habitats shift considerably after Toombak use allowing for bacteria that are not normal residents of the oral cavity such as *Staphylococcus*. Furthermore, those associated with normal oral health are reduced amongst users that include *Prevotella*.

Appendix 6

First time use of the *European Perspective Investigation into Cancer (EP-IC) - Norfolk* food frequency questionnaire on the Sudanese population, retrieving key nutrient intake and dietary consumption patterns

Amel Sami, Imad Elimairi, R. Paul Ross, C Anthony Ryan, and Catherine Stanton

APC microbiome institute Ireland, University College Cork, Teagasc research centre, Moorepark, Fermoy, Co. Cork and National Ribat University, Khartoum Sudan

Background: Limited studies exist analysing Sudanese population nutrient intakes. It is reported that about 16.9% of Sudanese adults are obese while 30.7% of Sudanese women aged 15-49 are anaemic[1]. Abdalla et al 2013 concluded that Sudanese rural households often consume pulses and rarely consume meat, chicken, and eggs with low intakes of fruit and vegetables[2]. Food frequency questionnaires (FFQs) are utilised to collect dietary data from a context specific food list estimating usual diets and consumption patterns throughout the food groups. Furthermore, FFQs also help to analyse specific nutrient intakes in each population. The EPIC- Norfolk FFQ was developed in 1988 and is a validated qualitative and quantitative questionnaire for the United Kingdom population[3]. This study served to utilise for the first time, the EPIC – Norfolk questionnaire as a useful, approachable, and updated method of food recording amongst the Sudanese population. Key nutritional variations were sought between Sudanese living in Sudan (Sud), Sudanese who migrate abroad to Ireland (Sudirel) as well as Irish controls (Irel).

Materials and methods: This study set out to achieve two main goals.

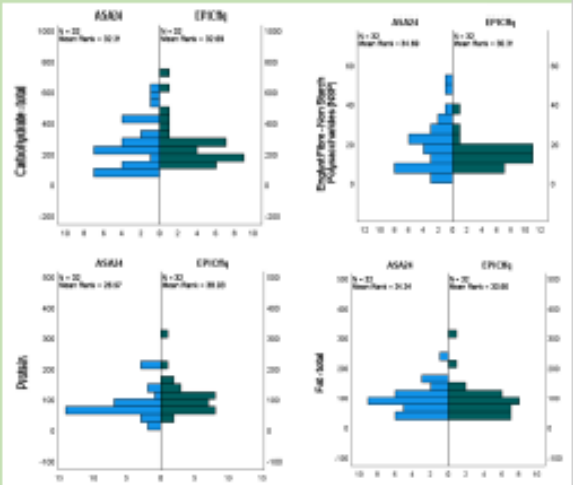
First goal	Second goal
The sublayer validation of the EPIC- NORFOLK questionnaire on the Sudanese population using ASA-24 recall online system	**The comparison of mean nutrient intakes between the Irish, Sudanese living in Ireland and Sudanese population**

Participants were asked on 230 baseline foods and a further 23 new foods were added to the current EPIC-Norfolk FFQ specific to Sudanese diets (figure 1). These results were compared with the online Automated Self-Administered Recall System (ASA24) on 32 subjects (figure 2). A further 141 FFQs were assessed for mean nutrient intake comparing three groups (Irish, Sudanese living in Ireland and Sudanese) using the FETA tool, version 6 (CAMB/PQ/6/1205) (https://www.epic-norfolk.org.uk/). Macronutrients and Micronutrients were compared for average mean values, eta squared strength of association and independent samples Mann Whitney U test for significance as well as non-significance (to assess validity) between systems for nutrient recording.

Figure 1: EPIC-Norfolk questionnaire

Fig 2: ASA24 recall system

Results: The EPIC Norfolk questionnaire was accurate in determining 17 nutrients (carbohydrates, fibre, fats and saturated fats, cholesterol, magnesium, sodium, niacin, Vitamin A, B2, B1, B6, C, D and E, zinc, and copper). Average intake values were similar with the ASA-24 for these nutrients allowing for reproducibility and sublayer reliability amongst the Sudanese subjects .10 nutrients varied in reporting between the two systems (protein, calcium, iron, folate, potassium, selenium, Vitamin B12 and K and mono and polyunsaturated fatty acids)(figure 3). Mean energy Kcal distributions were lower amongst the EPIC Norfolk FFQ compared to the ASA24 recall.



Non significant variation between 24 hour recall and FFQ indicating reliability in FFQ system in determining macronutrients: carbohydrates (p=0.936), fibre = (p=0.347) and fat (p=0.428). Protein (p=0.017) was varied in recording and could indicate overestimation through FFQ.

Figure 3: Sublayer validation of the EPIC-Norfolk questionnaire on Sudanese nutrient retrieval

When comparing three diets; Sudan, Sudanese living in Ireland and Ireland, the Sudanese population were consistently lower in mean average intakes (eta^2 medium to high strength of association >0.06) (figure 4). While vitamins A, C, D and B2, calcium, cholesterol and iron were still significantly lower in Sud, the eta squared highlighted a relatively lower strength of association with less variation between groups compared to other nutrients. Vitamin B5 or pantothenic acid average on the other hand was highest amongst the Sud group (1.17mg) compared to 0.27 and 0.53 mg in Irel and Sudirel respectively.

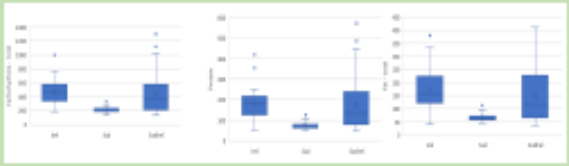


Figure 4: Consistently lower mean values amongst nutrient intake in the Sud group

Discussion and Conclusion: The FETA and EPIC-Norfolk questionnaire is a validated tool that can be utilised with ease on the Sudanese population. Mean nutrient intakes were generally lower but consistent throughout those living in Sudan, while average nutrient intake between Irel and Sudirel were similar throughout. This would indicate that population variation and dietary consumption patterns may concur this effect. Vitamin B5 may be higher amongst the Sudanese population due to the consumption of sorghum related products which are high in this vitamin. Due to economic strains, the Sudanese living in Sudan may have reductions in dietary averages particularly in higher financially valued products (animal protein) which may explain the reduced intake in protein macro group. Understanding of nutrient qualities in a population is key for further evaluation of microbiome – diet correlations and microbiome studies on the Sudanese.

Appendix 7 [Oral presentation]

Title: Gut microbiome traditional memory and new re-shaping portfolio in relation to migratory movements of the Sudanese population to westernised environments

Authors: Amel Sami, Imad Elimairi, Catherine Stanton, R Paul Ross and CA Ryan

Multiomics data integration, health and society

Background: Migratory movement of Africans to westernised countries can re-shape and impact on traditional gut microbiome memory. This study highlights preservations and merging in gut microbiome between the Sudanese and the Irish populations.

Material and Methods. 141 faecal samples were collected and frozen from three major cohorts aged 6-70 yrs. These were 25 Irish/ Caucasian participants living in Ireland [Irel], 49 Sudanese participants migrated to Ireland [SudIrel] and 67 participants living in Sudan, Africa [Sud]. DNA extraction and 16SrRNA sequencing was achieved. Microbiome Analyst was used for statistical analysis utilising Metagenomeseq, one-way Anova FDR adjusted p values, Lefse and alpha and beta diversity.

Results and Discussion: Sud and Irel groups gut microbiome differed significantly in B diversity [p value < 0.001] and SudIrel predominantly exhibited a 'middle lane' or or sandwiched status between Sud and Irel. *Faecalibacterium prausnitzii, Eubacterium rectale and Bifidobacterium adolescentis were prevalent species. Bacteroides* were highest in relative abundance in Irel and lowest in Sud while *Lactobacillus* and *Methanobrevibacter* were enriched in Sud and absent in Irel.

Sud lacked enrichment in *Lachnospira* and the species *Eubacterium eligens, Dialister invisus* and *Alistipes finegoldii*. Sud and SudIrel exhibited microbiome memory for *Prevotella_9, Alloprevotella, Marvinbryantia* and *Holdemanella,* found to be relatively abundant only in Sud and SudIrel. *Megasphaera, Olsenella, Klebsiella* and *Catenibacterium* were absent from the Irel cohort. *Bifidobacterium* and *Dorea* were highest in SudIrel while *Weissella* were only high in Sud.

Conclusion: Distinctions and preservations in gut microbiome are evident in African and European populations while those populations who migrate show 'middle lane' microbiome composition integrated by memory and reshaping.

Appendix 8

Abstract Title:

The oral manifestations of the Sudanese smokeless tobacco Toombak and its effects on long term cortisol body production

Background:

Toombak is produced from the *Nicotiana Rustica* tobacco plant and is primarily utilised by males from a young age in Sudan. This study served to analyse 1] oral manifestations of Toombak outlined in the literature as well as from a retrospective cohort review of Sudanese patients [346] and 2] chronic Toombak effects on cortisol production in a subset of users [19] in comparison to controls or non-users [18].

Material and Methods:

Scalp hair samples were collected safely in Sudan and the Fagerstrom Test for Nicotine Dependence Smokeless Tobacco [FTND-ST] questionnaire was utilised on Toombak users to outline dependency through a six-question scale, where a score of five indicated marked dependence. Cortisol was further extracted from the scalp hair using a methanol related protocol and measured by the Elisa kit Enzo life sciences as per manufacturer's instructions explored with spectrophotometry at wavelength 405nm.

Results:

Limited papers exist in relation to Toombak oral manifestations. With a further inclusion of 346 patients [aged 17-65] from our centre [National Ribat University, Sudan], the most common oral manifestations from Toombak use included keratosis, localised periodontal disease, generalised periodontal disease, gingival ulceration, leukoplakia [homogenous], leukoplakia [speckled], and oral squamous cell carcinoma. Mean concentration of cortisol in hair samples for non Toombak users was 19.4 pg/ml while in users it was 9.7 pg/ml [P = 0.023]. FTND-ST scores ranged from 4-9 where 84% of Toombak users were severely dependant on the product.

Conclusion:

Toombak causes important oral changes as well as marked physical dependence that interestingly may have adverse long-term effects on adrenal gland and cortisol production. Reduced levels of long- term cortisol production was found amongst users compared to controls, indicating that chronic and high nicotine and nitrosamines exposure affects normal cortisol axis regulation, may damage adrenal glands and place Toombak users at risk of cortisol insufficiency.

Appendix 9 [Oral presentation]

Introduction and aim: Sudanese fermented foods are a great source of affordable daily nutrition for many families and age groups all over the country. They include a diverse categorisation of starting ingredients and traditional methodologies of production that have been preserved for centuries. They are influenced by the rich and diverse ethnic backgrounds of the Sudanese people, different seasons throughout the year including drought, economic stability of the varying regions, migration patterns and inner displacement. Examples include crop-based fermentation products, sorghum, and millet. Sorghum products include, pancakes [Kisra], porridges [Aseeda], juices [Abreh], and alcoholic beverages [Mareesa]. Millet products include Damirga porridge and fermentation water. Plant-based fermented foods include, Cassia Obtusifolia [Kawal], sesame [Sigda], hibiscus [Furundu], watermelon, cucumis sativa, dates [date wines], boaob, aradeib, pumpkin, onions and mushroom based foods. Meat-based fermentation products [beef, sheep, goat, gazelle] include those from the small intestine [Musran], fat sheath [Miriss] and fresh bones [whole, pasted, sun dried smaller pieces or bone balls] – [Dodery]. Fish-based fermentation products include sun dried [Kajeik], pasted fish [Mendeshi] and moist fish [Faseekh].

Methodology: To understand the methodologies of fermented foods and drinks of Sudan we collected 46 samples of fermented foods from various urban and rural locations including shop- bought and homemade preparations. 16s RNA and ITS sequencing of Sudanese fermented foods was achieved.

Results and discussion: Many *Lactobacillus* species were found in these foods that included, *Lactobacillus* species; *L. lactis, L. fermentum, L. pontis. L. ingluviei, L. brevis. L. hamster, L amylolyticus, L. kefiri* and *L. nagelii*. Others were *Pediococcus*

acidilactici, Bacillus subtilis and coagulans Acetobacter pasteurianus and *tropicalis*. In animal-based products, *Bacteroides* and *Lysinibacillus* were increased in relative abundance while in fish-based products *Cetobacterium* and *Sporosarcina* were increased. **Conclusion:** Sudanese fermented foods are a rich source for nutrients and have a unique microbiome, depending on source and methodologies. Many of these foods are also a rich source of natural probiotic genera [*Lactobacillus, Acetobacter, Bacteroides*]. While the natural integrity of methodologies must be preserved, there is currently a profound risk of contamination from exterior sources such as the presence of *Ignatzschineria* and *Oblitimonas* as well as other pathogenic genera within these foods. North Sudan must therefore look for new and non-contaminated preparatory methods for the stabilisation of these rich foods.

Appendix 10 [Oral presentation]

Abstract Title - EU investment in Science Research in Sudan and Microbe role in Africa

In 2002, the UNDP reported that Sudan is one of the least developed countries in the world, with the ministry of finance and economic planning estimating that 65% of the population live below the poverty line. Sudan's multidimensional poverty index is also very high with around 1 million people living in and with poor health, education and living standards.

Microbiome science has been introduced in Sudan with the dual Irish- Sudanese partnership through Esther Alliance. This Alliance works to achieve global health partnerships in order to tackle major diseases and improve healthcare in less developed countries by building expertise and sharing experience. This quickly turned into a dedicated and trusted relationship in the sharing of knowledge, personnel, and investment into coherently aligned goals. Both partners have advocated the importance of microbiome science as a fundamental necessity for a country such as Sudan looking to improve its living outcomes for its population.

Sudan is a landscape for change and absorption with bright resilient researchers, openness and transparency and a willingness for international learning. Ireland on the other hand is a landscape for outreach with a pioneering team in global microbiome research and a passion for merged science.

In Sudan, the team has currently worked to achieve effective microbiome sampling, transport, and storage of research material. It has helped Sudan be introduced to next generation sequencing technologies of short hypervariable regions of microbial 16S rRNA and has allowed for the generation of new data sets.

Examples of future projects include the Maternal Irish study molding for the understanding of currently lagging critical aspects of Sudanese maternal and neonatal health and the Malaria – Microbiome relation to drug resistance

Appendix 11

Buccal microbiome; a pilot study on the Sudanese population and the effects of smokeless tobacco use; Toombak, on micro-organisms of the buccal cavity.

Introduction:

Smokeless tobacco Toombak is widely used particularly amongst males in the Sudan, Africa. The microbiome changes to the oral mucosa in relation to Toombak use has not been previously explored. Toombak is currently the most evident cause of premalignant conditions and oral cancer development in Sudan, where the buccal mucosa is one of the most aggressive cancer locations.

Material and Methods.

Participants were chosen from the oral and maxillofacial department, National Ribat University Khartoum Sudan, attending for dental clinic. Those with active or severe periodontal disease, active or rampant caries and periapical infection; were excluded from the study. 71 participants were included of which 26 non-users and 45 Toombak users were achieved. Swab analysis was carried out by rolling for 30 seconds in the right buccal mucosa in the absence of food, drink or exterior substances and were immediately frozen for DNA extraction and Illumina Miseq sequencing pipeline as per Qiagen and Illumina protocol guidelines respectively.

Results:

274 features [Operational Taxonomic Units or OTUS] were compared in the buccal mucosa of users of the Sudanese smokeless tobacco; Toombak and controls [non-users]. While phyla were generally similar in abundance; both *Actinobacteria* and *Fusobacteria* were increased in buccal mucosa of Toombak users. Using DeSeq2, *Moraxellaceae* [q=0.037], *Leptotrichiaceae* [q=0.019] and *Staphylococcaceae* [q=0.037] were the most significantly increased families in the buccal mucosa amongst Toombak users while *Enterobacteriaceae* [q=0.037] were increased amongst non-users. Indeed, the genera; *Leptotrichia* [q=0.029] were the most abundant amongst the buccal mucosa of Toombak users as well as *Staphylococcus* [q=0.047] and *Cutibacterium* [q= 0.047]. In the non-user group, *Escherichia-Shigella* [q= 0.033] and *Prevotella_6* [q= 0.047] were found to be increased. Utilising pattern search: other genera in the buccal mucosa highly correlating with Toombak users included

Corynebacterium_1, *Lautropia*, *Abiotrophia, Selenomonas_3* and *Peptococcus* as well as *uncultured bacterium*. Incorporating BLASTn NCBI analysis the most increased identifiable species amongst buccal mucosa of non-users was *Shigella sonnei* [FDR =0.0075, 99.14% per ident], while *Leptotrichia wadei* [q = 0.01, 98.42% per ident], *Actinomyces massiliensis* [q =0.024, 99.35% per ident] and *Staphylococcus caprae* [q = 0.05, 100% per ident] were the significantly increased in the buccal mucosa of Toombak users. Core microbiome consisted of 126 OTUs shared between the buccal mucosa of users and non –users, however, 12 unique OTUs existed in the buccal mucosa of non – users while there were 11 OTUs unique to Toombak users. LEfSe highlighted, *Staphylococcus*, *Corynebacterium_1* and *Cutibacterium* as the most discriminant genera of the buccal microbiome amongst Toombak users while *Scardovia* and *Prevotella_6* were discriminant of the buccal mucosa of non – users. Further OTU – species - LEfSe interconnection highlighted 6 discriminant OTUS on the buccal mucosa of the Toombak group of which *Gemella morbillorum* [98.28% per ident], *Prevotella nigrescens* [99.35% per ident], and *Lachnoanaerobaculum gingivalis* [99.32% per ident] could be identified with BLASTn while species distinguishing buccal mucosa of non – users included *Rothia mucilaginosa* [98.88% per ident], *Veillonella atypica* [98.28% per ident], *Streptococcus sobrinus* [99.35% per ident.], *Prevotella salivae* [99.35% per ident], *Prevotella pallens* [99.35% per ident] and *Bifidobacterium dentium* [99.33% per ident].

Conclusion:

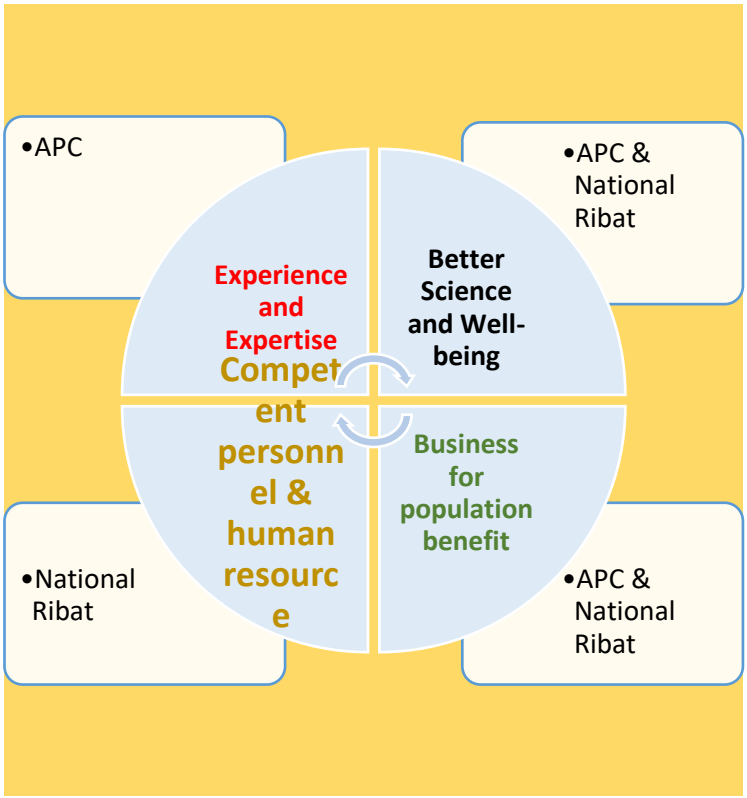
While there are similarities in the core microbiome of the buccal mucosa between controls and Toombak users, Toombak users carry several key features modifying their buccal microbiome into a more disease-associated composition Increased *Fusobacteria* have been associated with oral cancer development states as well as gastrointestinal [colorectal cancer], while increased *Staphylococcus* and *Cutibacterium*, may be associated with skin microflora, introduced during Toombak placement. *Corynebacterium_1* in another study by our centre was the most abundant genus in Toombak samples, the effects of which still need further analysis. Increases in *Lautropia* may be associated with locational immune deregulation. The increased abundance of *Abiotrophia* may harbour an increased risk of infective endocarditis development.

Appendix 12

Abstract Title: The National Ribat University Khartoum Sudan – APC Microbiome Institute Cork, Ireland partnership

The main objective of this partnership focuses on building relations and capacity building for moving forward in the establishment of Microbiome science in Sudan with a logical outcome model for the best fit between partners [Figure 2]. It also aims to allow the implementation of improved targets to health and the merging and co-adapting of undiscovered traits and trends in health and disease from the Sudanese population lining this with Irish Science. The microbiome partnership between Ireland and Sudan is on the first Esther partnerships to involve institutional and personnel capacity development directly in a specific research field, reflecting the vital urgency to target Sudanese research improvement. Research in Sudan aims to be functional and diverse. Future plans include bilateral visits exchanges, departmental outsourcing, building primary microbiome infrastructure in Sudan, strengthening of research ties in a multidisciplinary format and achieving further funding targets.

Figure 2:



Appendix 13

Toombak… the new king of smokeless tobacco

Objective: Toombak is a smokeless tobacco of Sudan that is made from the finely ground leaves of the *Nicotiana Rustica*, a tobacco species with a particularly high content of nicotine, tobacco specific nitrosamines and minor alkaloids. Tobacco is mixed with sodium bicarbonate [about 4:1], water is added, and the mixture becomes one of a high pH [11.0 - 11.8] where nicotine is completely protonated and the rate of absorption spikes in the blood rapidly. It is a powerfully addictive and harmful product. Recent reports through consultations, social media and word of mouth indicate the increasing use of Toombak within the USA and amongst American teenagers and young adults. The knowledge of Toombak amongst oral medicine clinicians must therefore be fully developed as well as further awareness protocols amongst society initiated.

Methods: 300 questionnaires were distributed to Toombak users within the Dental department of National Ribat University in the year 2017. 256 responded with mean age group of 36 years. Patients were asked in regards the initiation of Toombak, social implications including early age initiation [EAI], health awareness and difficulty of cessation. Furthermore, an assessment by an oral medicine clinician was carried out of the oral cavity.

Results: Of 256 responders, 41 patients had manifestations of disease in relation to the use of Toombak. This ranged from poor oral hygiene [37], halitosis [24] and periodontal disease [16] to oral epithelial dysplasia [7] and oral cancer [5]. 94% of patients were of EAI with 76% unable or unwilling to cessate the habit.

Conclusion: Toombak is a type of smokeless tobacco originating from Sudan but with new trends of worldwide spread. More young US citizens may be at risk of developing the habit of Toombak and awareness amongst US clinicians is vital. Toombak carries several health risks and is highly addictive.

Acknowledgements

We would like to acknowledge that this current project was supported by APC Microbiome Ireland, a research centre funded by Science Foundation Ireland [SFI], through the Irish Government's National Development Plan [grant no. 12/RC/2273_P2].

We would also like to acknowledge the Esther Alliance group for their partnership support and funding.