HHS Public Access

Author manuscript

*Lett Appl Microbiol*. Author manuscript; available in PMC 2023 March 01.

Published in final edited form as:

*Lett Appl Microbiol*. 2022 March ; 74(3): 444–451. doi:10.1111/lam.13623.

# Shotgun metagenome sequencing of a Sudanese toombak snuff tobacco: genetic attributes of a high tobacco-specific nitrosamine containing smokeless tobacco product

**R.E. Tyx**[1],

**A.J. Rivera**[1],

**S.B. Stanfill**[1],

**G.S. Zaatari**[2],

**C.H. Watson**[1]

[1]Division of Laboratory Sciences at the Centers for Disease Control and Prevention, Atlanta, GA, USA

[2]Department of Pathology and Laboratory Medicine, American University of Beirut, Beirut, Lebanon

## Abstract

The most alarming aspect of the Sudanese toombak smokeless tobacco is that it contains high levels of highly toxic tobacco-specific nitrosamines (TSNAs). Understanding the microbiology of toombak is of relevance because TSNAs are an indirect result of microbial-mediated nitrate reductions. We conducted shotgun metagenomic sequencing on a toombak product for which relevant features are presented here. The microbiota was composed of over 99% Bacteria. The most abundant taxa included Actinobacteria, specifically the genera *Enteractinococcus* and *Corynebacterium,* while Firmicutes were represented by the family Bacillaceae and the genus *Staphylococcus.* Selected gene targets were nitrate reduction and transport, antimicrobial resistance, and other genetic transference mechanisms. Canonical nitrate reduction and transport genes (i.e. *nar*) were found for *Enteractinococcus* and *Corynebacterium* while various species of *Staphylococcus* exhibited a notable number of antimicrobial resistance and genetic transference genes. The nitrate reduction activity of the microbiota in toombak is suspected to be a contributing factor to its high levels of TSNAs. Additionally, the presence of antimicrobial resistance and transference genes could contribute to deleterious effects on oral and gastrointestinal health of the

**Correspondence** Robert E. Tyx, Centers for Disease Control and Prevention, Division of Laboratory Sciences 4770 Buford Highway, NE MS S110-3; Atlanta, GA 30341, USA. rtyx@cdc.gov.

Robert E. Tyx and Angel J. Rivera contributed in equal measure to the totality of the content in this report.

Author Contributions

Robert Tyx, Angel Rivera and Stephen Stanfill conceived and designed the experiments. Robert Tyx and Angel Rivera performed and analysed the experiments. Ghazi Zaatari obtained the sample and contributed knowledge of toombak processing. Robert Tyx, Angel Rivera, Stephen Stanfill, Ghazi Zaatari and Clifford Watson wrote and revised the manuscript.

Conflict of Interest

None declared.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

end user. Overall, the high toxicity and increased incidences of cancer and oral disease of toombak users warrants further investigation into the microbiology of toombak.

**Keywords**

metagenomics; microbial communities; nitrosamines; toombak

## Introduction

One of the most harmful aspects of smokeless tobacco use is the carcinogenic compound content. Compared to other smokeless tobacco, the Sudanese oral snuff, known as toombak, contains the highest levels of tobacco-specific nitrosamines (TSNAs), which are considered one of the most carcinogenic components of tobacco and are causally associated with cancer (Idris et al. 1991). Accordingly, toombak which is found to be used in high incidence (~47%) by males 30 years of age and older residing in rural areas of the country, has been linked to a high incidence of oral cancer in Sudan and constitutes a serious public health risk (Idris et al. 1995, 1998; Ahmed 2013; Sami et al. 2021).

Toombak is made of fermented leaves from the tobacco *Nicotiana rustica,* a species known to hold high levels of nicotine and other minor alkaloids, direct precursors for TSNAs. However, the generation of TSNAs is not limited by the concentration of alkaloids but rather by the amount of nitrite generated by post-harvest microbial metabolism, increasing the potential for high TSNA concentrations (Idris et al. 1991; Burton et al. 1994; Wahlberg and Ringberger 1999). Hand-harvested and piled in small heaps, tobacco leaves are sun-cured for approximately 45 days. After curing, leaves are bundled, moistened and fermented for a few weeks at 30–45°C. The fermented leaves are then ground, placed in burlap sacks and aged for approximately 1 year (Arbyn et al. 2007). Prior to sale, toombak tobacco is mixed with sodium bicarbonate and placed in an airtight container for approximately 2 h (Idris et al. 1998; Arbyn et al. 2007). The resulting toombak is an alkaline (pH 8–11) moist mixture, with strong pungent aromas, and a wide range of nicotine content (8–102 mg $g^{-1}$, dry wt) (Arbyn et al. 2007). During the growing season, nitrate from soils or fertilizers are absorbed by tobacco plants while microbes present in the immediate and surrounding environment become associated with those plants (Davis and Nielsen 1999). Thus, the prevailing thought is that the combination of micro-organisms, nitrate and alkaloids in this tobacco, play a key role in the formation of TSNA (Chattopadhyay et al. 2021; Rivera and Tyx 2021; Vishwakarma and Verma 2021). The processing steps followed in making toombak allow for the introduction and proliferation of micro-organisms resulting in the persistence of numerous viable bacterial species, as identified by culturing media plate counts (Idris et al. 1998). Their presence can lead to active metabolism at various stages throughout the toombak making process. For instance, during curing, moisture loss results in ruptures of the tobacco cell membranes and the release of nitrate that becomes available for microbial metabolism (Spiegelhalder and Fischer 1991; Idris et al. 1995; Davis and Nielsen 1999). Consequently, respiratory nitrate reduction may initiate conversions of nitrate to nitrite, particularly under hypoxic conditions (Nishimura et al. 2007). The nitrite byproduct is well known to react readily with other biomolecules and could be considered cytotoxic if not

further metabolized. Additionally, nitrite is excreted from the cell by various transporters in some bacterial species (Moir and Wood 2001). Released nitrite can ultimately react with tobacco alkaloids forming TSNAs (Spiegelhalder and Fischer 1991; Nishimura et al. 2007; Rutqvist et al. 2011).

The process of toombak preparation, described by Sami et al. (2021), begins with air curing, followed by a 'natural or "compost" like fermentation', then is milled for consistency, and finally addition of alkaline carbonates contributes to a high level of free, or unionized nicotine (Stanfill et al. 2011). Tobacco fermentation has only been characterized in depth in cigar tobacco (Di Giacomo et al. 2007), but has similarities to the processing of smokeless tobacco, including toombak. This process is characterized by microbial successions that rapidly change the chemical and physical environment (i.e. rapid sugar catabolism, increased temperature and pH) increasing nitrite and TSNA concentrations (Di Giacomo et al. 2007). Early in fermentation, Firmicutes genera (i.e. *Aerococcus, Staphylococcus* and *Lactobacillus*) predominate in the community. As temperature and pH increase, the proliferation of *Bacillus* spp. is observed followed by Actinobacteria genera (*Corynebacterium* and *Yaniella*) becoming more prevalent as these two properties further increase. It is assumed that toombak fermentation is a similar process. This is supported by the finding of at least two 16S rRNA analyses that revealed the genus *Corynebacterium* as the most abundant bacterial taxa in the final toombak product (Tyx et al. 2016; Sami et al. 2021).

This study presents a brief overview of a whole metagenome shotgun sequencing analysis of both the organisms and the associated genes that may contribute to the harmful aspects of toombak tobacco. An understanding of the metabolic processes in toombak production could shed light on means of decreasing the levels of harmful by-products in toombak and other tobacco products.

## Results and Discussion

### Taxonomic distributions

Data were submitted to IMG/M-ER to determine phylogeny, abundances and genetic content. A presentation of the phylogenetic abundance at the highest level of taxonomy using the built-in IMG/M-ER generated phylogenetic abundance is presented in Table S1. The first notable aspect of this toombak metagenome is the absence of fungi, unlike domestic US moist snuff product, where fungi achieved some prominence (Rivera et al. 2020). Fungi have been identified and studied in certain smokeless products, and have been found to be variable in abundance likely based on conditions of fermentation (Rivera and Tyx 2021; Vishwakarma and Verma 2021). The presence of fungi has not yet been explored in depth in toombak, but fungi may be heavily involved in the early fermentation of toombak, as is the case in cigar tobacco (Di Giacomo et al. 2007).

The gene copy hits in this toombak metagenome (at 30%+ protein sequence identity) were almost entirely (>99%) attributable to Bacteria, with a nominal percentage of sequences (0·2%) associated with Eukaryote and Archaea, while a 0·4% of the sequences observed were viral (Table S1). Bacteria identified were mostly from the phyla Actinobacteria

(74·7%) (top genera: *Enteractinococcus, Corynebacterium, Yaniella*) and Firmicutes (19·3%) (top genera: *Aerococcus, Atopostipes, Bacillus, Carnobacterium, Enterococcus, Jeotgalicoccus, Oceanobacillus, Staphylococcus, Virgibacillus*) with smaller proportions of Proteobacteria (0·49%) (top genus: *Pseudomonas*).

A more in-depth characterization of the bacteria in the sample is presented, based on gene copy abundances of scaffolds with the RNA polymerase β-subunit (*rpoB*) housekeeping gene (Fig. 1). In the phylum Actinobacteria, the most abundant genera included *Enteractinococcus* at 65·8% relative abundance, *Corynebacterium* 22·5%, and the family Micrococcaceae (other than the genus *Enteractinococcus*) 2·05% (Table 1). In the phylum Firmicutes, the family Bacillaceae members were identified at 1·95% abundance, and the genera *Staphylococcus* 1·92%, *Atopostipes* 1·89% and *Virgibacillus* 1·05% were also found in relative abundance greater than 1% (Table 1). It should be noted that some taxa using the *rpoB* gene results could not be identified to the genus level, and are annotated at a higher level (i.e. Family or Order).

It should be noted that this particular sample was previously characterized by our laboratory using 16S rRNA gene amplicon sequencing (Tyx et al. 2016). That study showed that the family Corynebacteriaceae as the highest in abundance, with Aerococcaceae as the second and Yaniellaceae as the third most abundant. In the present study, we identified Micrococcaceae as the most abundant, followed by Corynebacteriaceae and Bacillaceae. Although results for both studies should be considered semi-quantitative, reported differences in the studies may be explained below.

Differences in apparent abundances in different studies may be resulting from one or more of a variety of explanations. The amplification and sequencing of a short part of the 16S rRNA gene result in some taxa being unable to be distinguished from each other at the genus or family level. This results in some groups' lower levels of taxonomy being clustered together at a higher level of taxonomy. There is amplification bias that is due to the nature of 16S amplicon sequencing, which uses PCR prior to going into the sequencing library preparation steps (Wu et al. 2010). Some sequences are more likely to amplify based on primers being used. Additionally, there may be bias in the shotgun metagenome that is quite different from 16S. Because the shotgun metagenome results presented here are based on the *rpoB* gene and the abundance assigned to the particular scaffolds that contained that gene, there could be additional bias introduced based on the conservation and ease of assembly of the genes around *rpoB* gene in the various species found in toombak. Database discrepancies and ongoing changes to taxonomic names provide for further inconsistencies. Another obvious difference between these studies is the taxa being called with Greengenes in the older 16S study, where *Yaniella* genus is attributed to Yaniellaceae Family, while in the current IMG/M-ER's taxonomy system, *Yaniella* falls under the family Micrococcaeae.

Several studies have recently reviewed the presence of microbes in smokeless tobacco products (Chattopadhyay et al. 2021; Rivera and Tyx 2021; Vishwakarma and Verma 2021). Of note is a recently published article that presented the results of several toombak samples that included samples from different regions (Sami et al. 2021). This study used 16S (V3–V4) rRNA gene amplicon sequencing, highlighting similar results to prior 16S studies

that included toombak (e.g. high relative abundances of *Corynebacterium, Atopostipes, Yaniella,* etc.). The species present in toombak have mostly been identified previously in similar tobacco products, but one notable difference is that the toombak microbiomes appear to be heavily laden with genera and clades of species that are known to reduce nitrate (*Corynebacterium* spp., *Enteractinococcus* and *Staphylococcus* spp.) (Collins 1987; Place et al. 2002; Bernard et al. 2010; Cao *et al.* 2012). Some of these species or genera have been identified as being associated with increased TSNAs in smokeless tobacco (Law et al. 2016). In contrast, many other smokeless products, such as in US moist snuff tend to have the dominant genera associated with fermentation such as *Marinilactibacillus* and *Tetragenococcus,* that are not known to reduce nitrate to nitrite (Rivera et al. 2020; Tyx et al. 2020; Rivera and Tyx 2021).

To validate genus and species identifications observed in IMG/M, we used read mapping of the trimmed reads to annotated genes. The recombinase gene, *recG*, a conserved bacterial housekeeping gene (Wen et al. 2005), was used to verify the presence of representative species of abundant genera. Several species in the genera *Corynebacterium* (*C. ammoniagenes, C. casei,* and *C. stationis*) and *Staphylococcus (S. xylosus, S. equorum,* and *S. succinus*), as well as a few others previously identified in smokeless tobacco products (e.g. *Atopostipes, Brachybacterium*) (Rivera and Tyx 2021; Vishwakarma and Verma 2021), were investigated and affirmed to be present (Table 2).

## Nitrogen utilization genes

Bacterial pathways that catalyse nitrate reduction to nitrite may be key in the formation of TSNAs in smokeless tobacco (Wahlberg and Ringberger 1999; Fisher et al. 2012; Tyx et al. 2016; Rivera and Tyx 2021). The canonical genes associated with this process are nitrate reductases, especially the dissimilatory nitrate reductase, encoded by the *nar* genes (*narGHJI*) (Gonzalez et al. 2006). A clade of three closely related *Corynebacterium* species (*C. stationis, C. ammoniagenes* and *C. casei*) that were found to be prominent in both fermented cigar tobacco (Di Giacomo et al. 2007) and this African toombak sample, are known dissimilatory nitrate reducers (Bernard et al. 2010; Sami et al. 2021). These species contain an incomplete denitrification pathway, where nitrite is not further reduced, potentially resulting in its incorporation into TSNAs (Law et al. 2016). We used read-mapping of sequences to a reference gene, *recG,* to confirm the presence of these species in the toombak sample, and then used read-mapping to the *narG* genes to confirm the presence of nitrate reductase (*narG*) and the nitrite-exporting transporter (*narK*) (Table 2). These genes from the most abundant genus, *Enteractinococcus,* were confirmed to be present (Table 2).

The genus Corynebacterium, presumed to predominate in late-stage fermentation, can tolerate elevated temperatures, pH, and salt content. In facilities that make toombak, the pungent aroma present is described as a suffocatingly strong ammonia smell (G.S. Zaatari, private communication). This observation could be in part due to the presence of ammonia-generating bacteria such as *C. ammoniagenes,* which is so named because it has urease activity, resulting in the metabolism of urea to ammonia (Collins 1987).

The third most prominent genus in toombak was found to be *Staphylococcus*. We confirmed the presence of *narG* from *Staphylococcus* species (likely *S. xylosus* and/or *S. equorum*). The *narK* gene for the nitrate/nitrite transporter in the genomes of these two *Staphylococcus* species is also referred to as a *narK* homolog and referred to as *narT* in the KEGG Orthology terms (*narT* = K10850, *narK* = K02575). In COG orthology terms, these transporters are both classified similarly into COG2223. These species also harbour a nitrite reductase gene (*nirB*) that allows for nitrite to be further metabolized to ammonia and subsequently utilized in assimilatory pathways (Neubauer et al. 1999). The presence of *nirB* in these species was also confirmed using read mapping (Table 2).

An ongoing generation of nitrite by species found in toombak products, especially those that contain incomplete denitrification pathways (containing nitrate reductase but not nitrite reductase), may result in the high levels of TSNAs seen in these products (up to 1 mg $g^{-1}$ of product of total TSNAs) (Stanfill et al. 2011). Levels of TSNAs in toombak are much higher when compared with other smokeless tobacco products. The level of NNK in one toombak product, for example, was found to be more than 50 times higher than maximum concentrations found in any US moist snuff products measured in that study (Stanfill et al. 2011). Native flora that reduce nitrate may also potentially contribute to formation of nitrite in the oral cavity, due to the excess of nitrate provided during smokeless use (Welch et al. 2016).

## Other genetic features

The data set was analysed for antimicrobial resistance and mobile genetic elements the CARD and ICEBERG databases, respectively (Table S2) (Bi et al. 2012; McArthur et al. 2013). From the ICEBERG results, only a single mobile element was identified in abundance, an integrating conjugative element ICE6013 (identified in *Staphylococcus aureus*), with 18·7-fold coverage and 76·6% of the sequence covered (Table S2). More antimicrobial resistance genes were found using the COG hits in IMG. This data is summarized in Table S3.

## Limitations of the study

Toombak snuff is a cottage industry product and as such it does not share the same manufacturing practices as other smokeless tobaccos. Although the ageing or fermentation steps are similar to other moist snuff products, there are dissimilarities to other smokeless tobacco products. For instance, products are not generally created in large lots but rather by individuals in smaller batches, thus making access and sampling challenging. Hence, the study holds some limitations that the authors wish to highlight.

The study presents data from a single toombak snuff product. However, the data presented here provide foundational knowledge of the microbiota of toombak tobacco products and a publicly accessible, rich data set for further investigation of microbial presence and associated metabolic pathways in toombak. This allows for in-depth comparisons to other smokeless tobacco products, in terms of microbial genetic capacities that may reflect their chemical constituent aspects.

Short sequence shotgun metagenomics may be associated with potential pitfalls, such as accurate classification of sequences to particular genera or species. Classification of a sequence is highly dependent on having a thorough database. This may be particularly problematic for international tobacco and other environmental samples, as these types of samples tend to be much less studied compared to clinical or human-associated samples.

Additionally, although we present results for the identification of specific genes, the presence of these genes does not necessarily indicate expression or function in a sample. Hence, we suggest gene expression profiling warrants further research to investigate metabolic capacities of the species present in these types of products.

The question of what taxa are specifically associated with TSNA formation in smokeless tobacco in general remains a knowledge gap in the literature. Law et al. (2016) presented a study that potentially addressed some portion of this gap, although the products analysed were all created in a laboratory rather than commercial products. Discerning the microbial processes taking place during smokeless tobacco production could facilitate a better understanding of how nitrite levels and subsequent TSNA production occurs. Identifying and targeting the presence and activity of the micro-organisms that contribute to the formation of those compounds in tobacco products could be beneficial in reducing harmful exposures. The potential to reduce or displace species with the capability of forming and exporting nitrite may be a worthwhile avenue to explore, in decreasing the levels of nitrosamines in toombak and other smokeless tobacco products.

## Materials and Methods

Metagenomic analysis was performed on a single toombak product-type previously analysed in an initial 16S rRNA study in our laboratory (Tyx et al. 2016). Ghazi Zaatari, M.D. (American University of Beirut; Beirut, Lebanon) obtained the Sudanese toombak sample from a store in Khartoum, Sudan. The sample was shipped for analysis to the Tobacco Products Laboratory at the Centers for Disease Control and Prevention (Atlanta, GA). Immediately upon receipt, the sample was catalogued and stored at −20°C for later use.

### DNA extraction

Sample DNA extraction, library preparation and shotgun sequencing of the metagenome were completed as described by Tyx and colleagues (Rivera et al. 2020; Tyx et al. 2020). DNA was sheared to approximately 400 base pair (bp) fragments and libraries prepared using a TruSeq® Nano DNA LS Sample Preparation kit (Illumina, Inc., San Diego, CA), and nucleic acid quantified using a fluorometer (Qubit® 2.0, Life Technologies, Carlsbad, CA). Insert size and quality was determined by DNA High Sensitivity assay chip on a 2100 Bioanalyzer (Agilent Technologies; Waldbornn, Germany).

### Sequencing and processing

Sequencing was performed using an Illumina MiSeq with 2 × 300 V3 sequencing kit. A total of 34 247 581 read pairs were generated before filtering. Sequence data processing is also previously described by Tyx et al. (2020). Read processing was performed using the bbduk.sh script (https://sourceforge.net/projects/bbmap/) to remove sequence adapters

and phiX phage control sequences. Quality filtering was adjusted for a minimum insert size of 60 bp and otherwise default parameters were used with SICKLE ver. 1.33 (Joshi and Fass 2011). A total of 33 849 942 read pairs remained after quality filtering, with an average GC content of 52%. SPAdes META ver. 3.10 was used for final assemblies which consisted of 96 735 293 bases assembled into 67 562 contigs on 67 218 scaffolds with a N50 of 5597 bp and a max. contig size of 367 844 bp (Nurk et al. 2017). The GC content of the assembly was 44·7%. Sequencing reads were mapped to the assembly, resulting in an average fold coverage of 165·69×. Mapping of the assembly and reference genes to the trimmed sequencing reads was performed using the BBMAP.sh script (ver. 38.44), using default parameters.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

All funding for the project was provided by the U.S. Centers for Disease Control and Prevention, Division of Laboratory Sciences. There was no specific grant for funding.

### Disclaimer

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention. The use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention, the Public Health Service or the U.S. Department of Health and Human Services.

## Data Availability Statement

Assembled sequence data was uploaded to the U.S. Department of Energy (DOE), Joint Genome Institute (JGI) Integrated Microbial Genomes/Metagenomes Expert Review system (IMG/M-ER), with IMG Taxon ID number 3300034651 (Markowitz et al. 2014). The assembled data contained 46 896 sequences annotated into 122 487 protein-coding genes and 1851 RNA genes. Raw sequence reads were submitted to the NCBI Sequence Read Archive and are available currently as accession number SRP158247.

## References

Ahmed HG (2013) Survey on knowledge and attitudes related to the relation between tobacco, alcohol abuse and cancer in the northern state of Sudan. Asian Pac J Cancer Prev 14, 2483–2486. [PubMed: 23725160]

Arbyn M, Bosch X, Cuzick J, Denny L, Galloway D, Giuliano AR, Herrero R, Lambert P et al. (2007) IARC monographs programme on the evaluation of carcinogenic risks to humans. In Human Papillomaviruses. pp. 9–31. Lyon, France: International Agency for Research on Cancer.

Bernard KA, Wiebe D, Burdz T, Reimer A, Ng B, Singh C, Schindle S and Pacheco AL (2010) Assignment of *Brevibacterium stationis* (ZoBell and Upham 1944) Breed 1953 to the genus Corynebacterium, as *Corynebacterium stationis* comb. nov., and emended description of the genus Corynebacterium to include isolates that can alkalinize citrate. Int J Syst Evol Microbiol 60, 874–879. [PubMed: 19661509]

Bi D, Xu Z, Harrison EM, Tai C, Wei Y, He X, Jia S, Deng Z et al. (2012) ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. Nucleic Acids Res 40, D621–626. [PubMed: 22009673]

Burton HR, Dye NK and Bush LP (1994) Relationship between tobacco-specific nitrosamines and nitrite from different air-cured tobacco varieties. J Agricult Food Chem 42, 2007–2011.

Cao Y-R, Jiang Y, Jin R-X, Han L, He W-X, Li Y-L, Huang X-S, Xue Q-H *Enteractinococcus coprophilus* gen. nov. sp. nov. of the family Micrococcaceae isolated from Panthera tigris amoyensis faeces and transfer of Yaniella fodinae Dhanjal *et al.* 2011 to the genus Enteractinococcus as *Enteractinococcus fodinae* comb. nov. Int J Syst Evol Microbiol 62, 2710–2716.

Chattopadhyay S, Malayil L, Mongodin EF and Sapkota AR (2021) A roadmap from unknowns to knowns: advancing our understanding of the microbiomes of commercially available tobacco products. Appl Microbiol Biotechnol 105, 2633–2645. [PubMed: 33704513]

Collins MD (1987) Transfer of *Brevibacterium ammoniagenes* (Cooke and Keith) to the Genus *Corynebacterium* as *Corynebacterium ammoniagenes* comb. nov. Int J Syst Bacteriol 37, 442–443.

Davis DL and Nielsen MT (1999) Tobacco: Production, Chemistry and Technology. Hoboken, NJ, USA: Wiley.

Di Giacomo M, Paolino M, Silvestro D, Vigliotta G, Imperi F, Visca P, Alifano P and Parente D (2007) Microbial community structure and dynamics of dark firecured tobacco fermentation. Appl Environ Microb 73, 825–837.

Fisher MT, Bennett CB, Hayes A, Kargalioglu Y, Knox BL, Xu D, Muhammad-Kah R and Gaworski CL (2012) Sources of and technical approaches for the abatement of tobacco specific nitrosamine formation in moist smokeless tobacco products. Food Chem Toxicol 50, 942–948. [PubMed: 22142690]

Gonzalez PJ, Correia C, Moura I, Brondino CD and Moura JJ (2006) Bacterial nitrate reductases: molecular and biological aspects of nitrate reduction. J Inorg Biochem 100, 1015–1023. [PubMed: 16412515]

Idris A, Ahmed H and Malik M (1995) Toombak dipping and cancer of the oral cavity in the Sudan: a case-control study. Int J Cancer 63, 477–480. [PubMed: 7591252]

Idris AM, Ibrahim SO, Vasstrand EN, Johannessen AC, Lillehaug J, Magnusson B, Wallström M, Hirsch J-M et al. (1998) The Swedish snus and the Sudanese toombak: are they different? Oral Oncol 34, 558–566. [PubMed: 9930371]

Idris AM, Nair J, Ohshima H, Friesen M, Brouet I, Faustman EM and Bartsch H (1991) Unusually high levels of carcinogenic tobacco-specific nitrosamines in Sudan snuff (toombak). Carcinogenesis 12, 1115–1118. [PubMed: 2044192]

Joshi NA and Fass JN (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. https://github.com/najoshi/sickle

Law AD, Fisher C, Jack A and Moe LA (2016) Tobacco, microbes, and carcinogens: correlation between tobacco cure conditions, tobacco-specific nitrosamine content, and cured leaf microbial community. Microb Ecol 72, 120–129. [PubMed: 27023797]

Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E and Pillay M (2014) IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res 42, D568–D573. [PubMed: 24136997]

McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ et al. (2013) The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57, 3348–3357. [PubMed: 23650175]

Moir JW and Wood NJ (2001) Nitrate and nitrite transport in bacteria. Cell Mol Life Sci 58, 215–224. [PubMed: 11289303]

Neubauer H, Pantel I and Gotz F (1999) Molecular characterization of the nitrite-reducing system of *Staphylococcus carnosus*. J Bacteriol 181, 1481–1488. [PubMed: 10049379]

Nishimura T, Vertès AA, Shinoda Y, Inui M and Yukawa H (2007) Anaerobic growth of *Corynebacterium glutamicum* using nitrate as a terminal electron acceptor. Appl Microbiol Biotechnol 75, 889–897. [PubMed: 17347820]

Nurk S, Meleshko D, Korobeynikov A and Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. Genome Res 27, 824–834. [PubMed: 28298430]

Place RB, Hiestand D, Burri S and Teuber M (2002) *Staphylococcus succinus* subsp. casei subsp. nov., a dominant Isolate from a surface ripened cheese. Syst Appl Microbiol 25, 353–359. [PubMed: 12421073]

Rivera AJ, Tyx RE, Keong LM, Stanfill SB and Watson CH (2020) Microbial communities and gene contributions in smokeless tobacco products. Appl Microbiol Biotechnol 104, 10613–10629. [PubMed: 33180172]

Rivera AJ and Tyx RE (2021) Microbiology of the American smokeless tobacco. Appl Microbiol Biotechnol 105, 4843–4853. [PubMed: 34110473]

Rutqvist LE, Curvall M, Hassler T, Ringberger T and Wahlberg I (2011) Swedish snus and the GothiaTek® standard. Harm Reduction J 8, 1–9.

Sami A, Elimairi I, Patangia D, Watkins C, Ryan CA, Ross RP and Stanton C (2021) The ultra-structural, metabolomic and metagenomic characterisation of the sudanese smokeless tobacco 'Toombak'. Toxicol Rep 8, 1498–1512. [PubMed: 34401360]

Spiegelhalder B and Fischer S (1991) Formation of tobacco-specific nitrosamines. Crit Rev Toxicol 21 , 241.

Stanfill SB, Connolly GN, Zhang L, Jia LT, Henningfield JE, Richter P, Lawler TS, Ayo-Yusuf OA et al. (2011) Global surveillance of oral tobacco products: total nicotine, unionised nicotine and tobacco-specific N-nitrosamines. Tob Control 20, e2.

Tyx RE, Rivera AJ, Keong LM and Stanfill SB (2020) An exploration of smokeless tobacco product nucleic acids: a combined metagenome and metatranscriptome analysis. Appl Microbiol Biotechnol 104, 751–763. [PubMed: 31820070]

Tyx RE, Stanfill SB, Keong LM, Rivera AJ, Satten GA and Watson CH (2016) Characterization of bacterial communities in selected smokeless tobacco products using 16S rDNA analysis. PLoS One 11, e0146939. [PubMed: 26784944]

Vishwakarma A and Verma D (2021) Microorganisms: crucial players of smokeless tobacco for several health attributes. Appl Microbiol Biotechnol 105, 6123–6132. [PubMed: 34331556]

Wahlberg I and Ringberger T (1999) Smokeless tobacco. In Tobacco: production, chemistry, and technology ed. Davis DEL and Nielsen MR Oxford: Blackwell Science Ltd.

Welch JLM, Rossetti BJ, Rieken CW, Dewhirst FE and Borisy GG (2016) Biogeography of a human oral microbiome at the micron scale. Proc Natl Acad Sci 113, E791–E800. [PubMed: 26811460]

Wen Q, Mahdi AA, Briggs GS, Sharples GJ and Lloyd RG (2005) Conservation of RecG activity from pathogens to hyperthermophiles. DNA Repair (Amst) 4, 23–31. [PubMed: 15533834]

Wu JY, Jiang XT, Jiang YX, Lu SY, Zou F and Zhou HW (2010) Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. BMC Microbiol 10, 255. [PubMed: 20937143]

**Significance and Impact of the Study:**

A great deal of morbidity and mortality in Sudan is attributed to toombak due to carcinogenicity associated with the abundance of tobacco-specific nitrosamines (TSNAs). TSNAs are formed through a reaction of alkaloids, such as nicotine, and nitrite. Nitrite is a product of microbial metabolic processes. We used shotgun metagenome sequencing to identify microbial populations and the genetic content of a representative toombak sample. A better understanding of the microbial communities in toombak may be beneficial in development of mitigation or regulatory strategies targeting a reduction or elimination of microbes resulting in products with a potential for reduced harm.

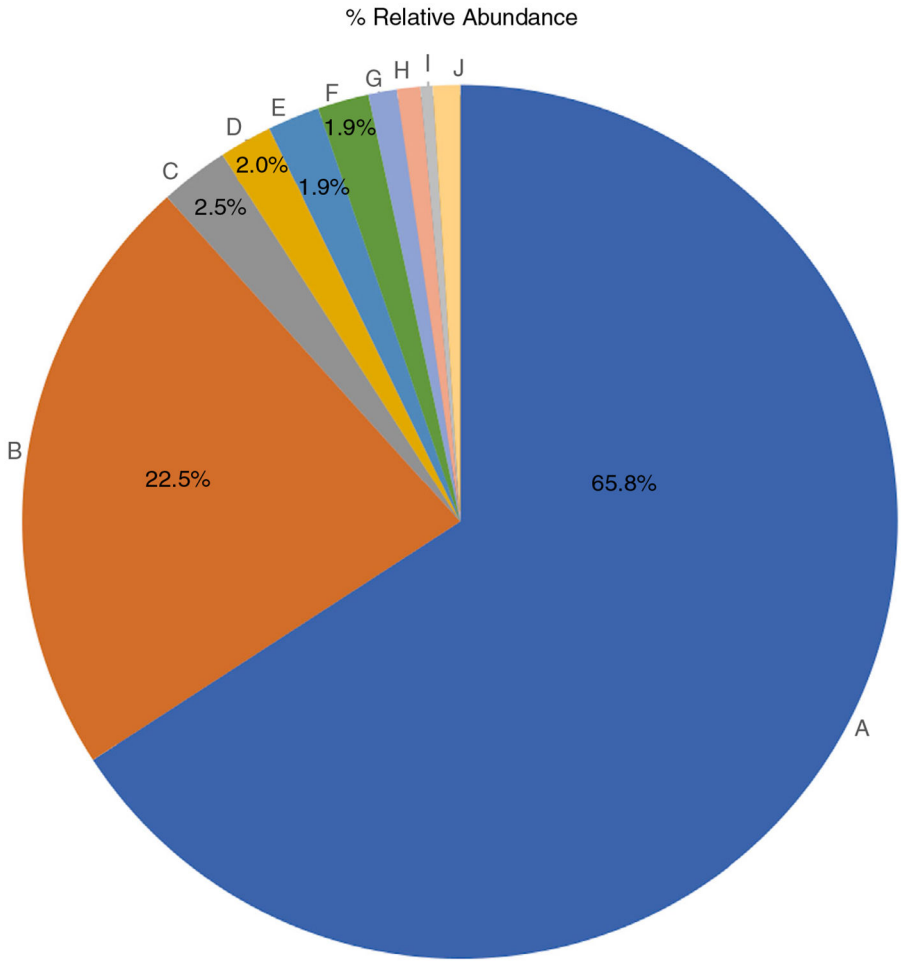


**Figure 1.**

Prevalent bacterial taxa. (■) *Enteractinococcus;* (■) *Corynebacterium;* (■) Other Micrococcaceae (*Yaniella*); (■) Other Bacillaceae (*Oceanobacillus* Bacillus); (■) *Staphylococcus;* (■) *Atopostipes;* (■) *Virgibacillus;* (■) *Oceanobacillus;* (■) Lactobacillales (Unable to classify further); (■) unclassified.

**Table 1**

Prevalent bacterial genera based on scaffold hits of the *rpoB* (RNA polymerase β-subunit) gene

| Phylum | Family/Order | Genus | % Relative abundance | Hits |
|---|---|---|---|---|
| Actinobacteria | Micrococcaceae | *Enteractinococcus* | 65·8 | 15239 |
| Actinobacteria | Corynebacteriaceae | *Corynebacterium* | 22·5 | 5210 |
| Actinobacteria | Micrococcaceae | * | 2·05 | 475 |
| Firmicutes | Bacillaceae | * | 1·95 | 452 |
| Firmicutes | Staphylococcaceae | *Staphylococcus* | 1·92 | 445 |
| Firmicutes | Carnobacteriaceae | *Atopostipes* | 1·89 | 438 |
| Firmicutes | Bacillaceae | *Virgibacillus* | 1·05 | 243 |
| Firmicutes | Bacillaceae | *Oceanobacillus* | 0·860 | 199 |
| Actinobacteria | Micrococcaceae | *Yaniella* | 0·484 | 112 |
| Firmicutes | Order Lactobacillales | * | 0·454 | 105 |
| Others, including unclassified | | | 1·00 | 232 |

Taxonomic Group is given with the following levels of taxonomy (unless stated otherwise): Phylum; Family; Genus. Percent relative abundance is based on the number of hits of that particular taxonomic group divided by the total number of hits, multiplied by 100.

* Unclassified at indicated level.

## Table 2

Read mapping genes of interest to reference species

| Bacterial Species | *recG* | *narG* | *narK* | *nirB* |
|---|---|---|---|---|
| Actinobacteria | | | | |
| *Brachybacterium*[*] | 14·2 | 7·76 | 28·5 | 38·8 |
| *Corynebacterium* | 875 | 1071 | 798 | – |
| *Enteractinococcus* | 361 | 1427 | 40·8 | – |
| Firmicutes | | | | |
| *Aerococcus*[*] | 25·8 | – | – | – |
| *Atopostipes* | 84·7 | – | – | – |
| *Staphylococcus* | 198 | 199 | 75·2 | 151 |
| *Virgibacillus* | 3·95 | 7·76 | 0 | – |

[*] *Brachybacterium* genus represented by *B. faecium* 6-10 DSM4810; *Corynebacterium* by *C. ammoniagenes; Enteractinococcus* by *E. helveticum* UASW1574.

[†] *Aerococcus* genus represented by *A. viridans; Atopostipes* by *A. suicloacalis* DSM15692; *Staphylococcus* by *S. xylosus; Virgibacillus* by *V. sp.* SK37.