



Geographic and position-based variations in phyllospheric bacterial communities present on flue-cured tobacco

Jianbin Ye¹ · Yilang Ding² · Xiaona Qi² · Jia Xu¹ · Xuepeng Yang² · Zhan Zhang³

Received: 24 August 2021 / Revised: 27 October 2021 / Accepted: 28 October 2021 / Published online: 18 November 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Although tobacco leaves (TLs) contain abundant bacteria, how the geography and leaf position of TLs affect these bacteria is unclear. Here, TLs at different positions from Henan (HN, strong flavor style) and Yunnan (YN, fresh flavor style) provinces were collected, and the bacteria were characterized by Illumina sequencing at harvest and 1 year of storage. Bacterial communities were very different between TLs originating from different geographical areas and positions, and beta diversity analysis showed that leaf position was the most important factor for phyllospheric bacterial communities, followed by geographical area and storage time. At the genus level, *Subdoligranulum*, *Thermus*, and *Acinetobacter* were obviously more abundant in HN than in YN, while *Blautia* and *Ruminococcus* were significantly more abundant in YN. These differences in bacterial communities decreased after 1 year of storage, indicating that the microbiota tends to become similar during tobacco processing. Storage time also affected the phyllospheric bacteria of TLs, as the bacterial communities shifted significantly on both HN and YN TLs after 1 year of storage. Significant differences in the predicted genes were also observed between the different geographic locations and leaf positions. Potential human pathogens, including *Acinetobacter*, *Methylobacterium*, and *Escherichia-Shigella*, were greatly different between TLs originating from different areas and positions. These data suggested that geographic variations and positions were associated with phyllospheric bacterial communities on TLs, which may be related to not only the flavor style and quality of TLs but also the potential health risks to humans.

Key points

- Tobacco leaf position and tobacco growth location affected bacterial communities.
- Microbial communities of TLs shifted significantly after one year of storage.
- Potential human pathogens differed at different leaf positions and growth locations.

Keywords Geographic · Phyllospheric · Bacterial communities · Illumina sequencing

Introduction

Flue-cured tobacco, a high value economic crop, is planted worldwide and commonly used as raw material for a series of tobacco products (Huang et al. 2010). Billions of cigarettes are sold worldwide every year comprised of more than 1,000,000 tons of flue-cured tobacco leaves (FTLs). The FTLs harvested from around the world differ according to their type, storage time, growth location, and leaf position, and these differences are related to the different qualities of FTLs (Sun et al. 2011; Zhang et al. 2018). However, the relationships between these factors and phyllospheric bacterial communities are not well documented.

Recently, traditional culture methods and molecular techniques have been widely used to illustrate the bacterial communities of tobacco leaves. Previous studies have screened

✉ Xuepeng Yang
yangxp0000@163.com

✉ Zhan Zhang
zhangzhan2059729@126.com

¹ Key Laboratory of Translational Tumor Medicine in Fujian Province, Putian University, Putian City 351100, Fujian Province, China

² School of Food and Biological Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, Henan Province, China

³ Technology Center, China Tobacco Henan Industrial Co., Ltd., Zhengzhou 450001, China

the bacterial communities of tobacco leaves by using 16S ribosomal RNA gene (rDNA) clone libraries, polymerase chain reaction-denaturing gradient gel electrophoresis (PCR-DGGE), or high-throughput sequencing (Huang et al. 2010; Ye et al. 2017; Zhao et al. 2007). Most of the studies have focused on the aging process, which includes bacterial fermentation, for improving the quality of tobacco leaves. The results demonstrate that bacterial communities inhabiting tobacco leaves could be affected by the aging process as well as threshing and redrying, which may further affect the constituents of tobacco leaves. In addition, few studies focusing on cigar and smokeless tobacco products have been reported, and results have suggested that the storage time and additives on the tobacco products largely affected the bacterial communities (Chattopadhyay et al. 2019; Chopyk et al. 2017; Smyth et al. 2017).

Additionally, the cigarettes made from FTLs can cause many significant health problems, such as cancer, oral lesions, and nicotine addiction. Many studies have evaluated the relationship between the harmful constituents of TLs and health problems. Few studies, however, have focused on the potential roles of the bacterial components in these smoking-associated illnesses (Sapkota et al. 2010). Several studies have shown that adverse health outcomes are linked to some potentially harmful constituents of tobacco, including bacteria, fungi, and microbial-derived toxins (Fisher et al. 2012; Rooney et al. 2005). These bacteria or other microbes can be resistant to the extreme tobacco microenvironment (low moisture, pH, and high temperature), which allows them to survive the whole process of tobacco product manufacturing and storage. For instance, *Bacillus* species were commonly identified in cured tobacco leaves as well as other tobacco products (Han et al. 2016; Smyth et al. 2017). These bacteria present in tobacco leaves are not only related to the production of microbial-derived toxins and other secondary metabolites (such as tobacco-specific nitrosamines, TSNA) but also include potential human and respiratory pathogens (Chattopadhyay et al. 2019; Sapkota et al. 2010). Thus, to better understand the adverse health outcomes of tobacco and smoke, more information on the bacterial communities inhabiting tobacco leaves needs to be obtained.

However, to the best of our knowledge, there are no data regarding the geographic and position-associated variations in bacterial communities that inhabit tobacco leaves. In China, there are several areas from which flue-cured tobacco leaves are produced, including Yunnan and Henan Provinces as the two main production regions. The FTLs originating from Yunnan have a fresh flavor style, while Henan FTLs have a strong flavor style. The tobacco microenvironment of leaves from these two different areas may cause variation in the bacterial populations. Moreover, based on their different positions on the plant (upper, middle, and bottom), FTLs also have different qualities and microenvironments.

Several studies have shown that FTLs originating from different areas have different nicotine and TSNA levels and other harmful chemical components (Nurnasari and Subiyakto 2015; Sun et al. 2011). Microbes play key roles in the formation of these constituents, such as TSNA (Law et al. 2016; Wei et al. 2014) and flavor (Wang et al. 2018).

Thus, the aim of this study was to investigate the bacterial communities inhabiting FTLs originating from two different areas by using a next-generation sequencing method. In addition, position-associated variations in the bacterial microbiota of the FTLs were also examined. The results of the present study may help to reveal the microbial-related health risks associated with tobacco products as well as the relationship between the qualities of TLs and phyllospheric bacterial communities.

Materials and methods

Sample collection

Two large groups of tobacco leaves (category K326) were collected from Henan Province (assigned as HN1) and Yunnan Province (assigned as YN1). Each large group was composed of three subgroups with different kinds of tobacco leaves, including the bottom leaves (position X, named HN1-X or YN1-X), middle leaves (position C, named HN1-C or YN1-C), and upper leaves (named position B, HN1-B or YN1-B). Each subgroup was collected from three different cities, specifically, Xuchang (HN1-X1, B1, C1), Sanmenxia (HN1-X2, B2, C2), and Pingdingshan (HN1-X3, B3, C3) in Henan Province and Lijiang (YN1-X1, B1, C1), Chuxiong (YN1-X2, B2, C2), and Qujing (HN1-X3, B3, C3) in Yunnan Province. Then two different subsets of leaves were taken from each sample for independent DNA extraction and sequencing and were considered biological repetitions. Thus, 36 tobacco leaf (TL) samples were collected first. Then, these 36 TLs were stored under the same natural conditions without adding any fumigant for 1 year. During the storage process, the humidity and temperature of the storage environment was controlled at 60–65% and 25 °C, respectively. Then, the samples were collected (named HN2 or YN2) to test the changes in bacterial communities. Sample information is listed in Supplemental Table S1.

DNA extraction and PCR amplification

FTL samples (50 g, with the same water content for all samples) were cut into ~5 cm *5 cm pieces under sterile conditions and washed with 1 L PBS buffer for three times. To collect as many bacterial populations as possible from the surface of FTLs, the washes were shaken at 200 rpm for 2 h. Additionally, all the alive and dead bacteria were collected

according to previous reports (Chattopadhyay et al. 2019; Zhang et al. 2020) as the proportion of dead bacteria are relatively small and may not largely affect the final community and functional analyses. Then, the suspensions were concentrated with a 0.45 µm filter membrane. Total DNA was extracted from the membrane using Fast DNA SPIN kits (MP Biomedicals, Santa Ana, CA, USA) according to the instructions. The extracted DNA was measured using a NanoDrop-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The V3-V4 hypervariable region of the bacterial 16S rRNA gene was amplified by PCR using universal primers (319F 5'-ACTCCTACGGGAGGCAGCAG-3' and 806R 5'-GGACTACHVGGGTWTCTAAT-3'), where the barcode was a 12 bp sequence unique to each sample. The PCRs were carried out as follows: 98 °C for 2 min, followed by 25 cycles at 98 °C for 15 s, 55 °C for 30 s, and 72 °C for 45 s and a final extension at 72 °C for 10 min. PCRs were performed in a total reaction volume of 25 µL, which consisted of 0.25 µL Q5 High-Fidelity DNA polymerase (5 unit/µL), 0.25 µL extracted DNA template, 5 µL Q5 reaction buffer (5×), 2 µL 2.5 mM dNTPs, 1 µL each primer (10 µM), and 13.75 µL ddH₂O. PCR results were confirmed by gel electrophoresis before further procedures.

Illumina NovaSeq sequencing and sequence quality control

Successful PCR amplicons were extracted from agarose gels, purified using Agencourt AMPure Beads (Beckman Coulter, Indianapolis, IN) and quantified using a PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA). Confirmed PCR amplicons were pooled in equal amounts, after which 16S rRNA sequencing was performed using the Illumina NovaSeq platform (paired-end sequenced, 2 × 250 bp) according to standard protocols. Sequencing was carried out by a commercial company (Shanghai Personal Biotechnology Co., Ltd., Shanghai, China) using the NovaSeq 6000 SP Reagent Kit v1.5 (300 cycles) (Illumina, USA).

The sequence reads were processed through Quantitative Insight Into Microbial Ecology (QIIME, version 1.9.1) (Caporaso et al. 2010). Raw sequencing reads with exact matches to the barcodes were assigned to respective samples and identified as valid sequences. Then, the sequences were screened for low-quality bases and short read lengths and filtered according to the following exclusion criteria: sequences less than 150 bp, sequences with an average quality score of < 20 over a 25 bp sliding window, sequences containing ambiguous bases, and mononucleotide repeats of > 8 bp. Paired-end read pairs were assembled using Fast Length Adjustment of Short Reads (FLASH v1.2.11) (Magoč and Salzberg 2011). Sequences that overlapped for more than 15 bp were merged, reads that could not be merged were discarded, and the mismatch rate of overlapping sequences was

less than 0.1. Before further analysis, the chimeric sequences were removed using UCHIME (Edgar et al. 2011).

Bioinformatics analysis

Quality trimmed sequences were clustered into operational taxonomic units (OTUs) at 97% sequence similarity by UCLUST (Edgar 2010). The representative sequence was selected from each OTU, and taxonomic classification was conducted using the Ribosomal Database Project (RDP) Classifier (<http://rdp.cme.msu.edu/>). The OTU table was generated to display the abundance of each OTU in each sample as well as the taxonomy of these OTUs. All sequences taxonomically assigned to likely chloroplasts and the phylum *Cyanobacteria* were removed. OTUs with a relative abundance less than 1% at the genus level were assigned to “others.”

Bacterial alpha diversity was determined using the observed richness metric and Shannon index calculated by the Phyloseq R package (McMurdie and Holmes 2013). Significant differences between samples were tested using ANOVA with Tukey’s honestly significant difference (HSD) post hoc test. The beta diversity was calculated using Bray–Curtis dissimilarity and compared using analysis of similarities (ANOSIM) (McMurdie and Paulson 2015) between samples. Considering the uneven sequencing depth of each sample, diversity was compared by selecting equal numbers of sequence reads for all samples. *p*-values of ≤ 0.05 were considered statistically significant. A heatmap was created and visualized with R version 3.2.2 and *vegan* heatmap (Ploner 2012).

Prediction of functional metagenomic profiles

Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) was used against the GreenGenes database and then used for predicting the KEGG (orthologs) functional profiles of microbial communities based on the 16S rRNA sequences (Langille et al. 2013). Based on the KEGG database, three levels of metabolic pathway information and EC numbers were screened. Functional profiles were visualized by STAMP (Parks et al. 2014).

Availability of supporting data

Sequencing data of all TLs included in this study have been deposited in the NCBI BioProject database under BioProject accession number PRJNA646576 (<https://www.ncbi.nlm.nih.gov/bioproject/>).

Results

Sequencing data analysis

DNA extraction was completed on the 36 FTL samples (each sample was sequenced twice), and all of them were sequenced successfully. In total, 3,790,546,508 sequences were obtained with an average of 127,038 ($\pm 52,643$ SD) (Supplemental Table S1) sequences for each sample. To ensure that samples included in the final dataset had appropriate sequence coverage, a Good's coverage cutoff of 0.9 was chosen; samples falling below this cutoff were removed from the analysis. After filtering, these sequences were clustered into 108,307 OTUs (97% identity) across all 36 samples (72 sequence samples), with 52,260 from Henan samples (HN) and 56,047 from Yunnan samples (YN).

Alpha and beta diversity analyses

Sequence data were analyzed by geography (HN, YN), storage times (1, 2), and position (B, C, X) to illustrate the alpha and beta diversity. Alpha diversity metrics were calculated and are shown in Fig. 1. The alpha diversity was much different ($p < 0.01$) between HN1 (observed: 1745.9 ± 219.7 , Chao1: 2015.6 ± 170.9) and YN1 (observed: 1550 ± 572.8 , Chao1: 1821.7 ± 573.9) but was similar between HN2 (1808.4 ± 145.5 ; Chao1: 2016.1 ± 147.0) and YN2 (observed: 1800.4 ± 358.4 , Chao1: 2038.6 ± 421.4). On average, both the richness (Observed) and diversity (Chao1) were similar between HN1 and HN2 but were obviously different ($p < 0.01$) between YN1 and YN2. This phenomenon could be attributed to differences in the initial bacterial communities between the TLs, which implied that the bacterial

communities present on TLs were very important and further affected the final composition of bacteria. Interestingly, the alpha diversity (observed and Chao1) varied between different positions (B, C, X) over geography (HN, YN) and storage times (1, 2) (Fig. 1B–D). The observed alpha diversity of the lower leaves (position X) was the consistently highest, followed by that of the middle leaves (position C) and upper leaves (position B).

Beta diversity analysis (differences in composition between samples) was displayed by using the Bray–Curtis dissimilarity measure plotted using PCoA (Fig. 2). The geography and leaves' position were the variables that explained most of the variance in bacterial microbiota between samples (Fig. 2). The first principal component (PC1) explained 32.2% of the total variability between bacterial communities. The results showed that the samples differed greatly between these four groups (HN1, YN1, HN2, and YN2). The PERMANOVA analysis was further carried out to evaluate clustering of three factors, and the results showed that the largest significant clustering was for leaf position within the groups HN1 and YN1 (comparison of B vs. C, B vs. X, and C vs. X; see the Supplemental Table S2). Then, significant clustering by geography (HN1 vs YN1, $R^2 = 0.2280$, $p = 0.001$) and storage time (HN1 vs. HN2, $R^2 = 0.2584$, $p = 0.001$; YN1 vs YN2, $R^2 = 0.2136$, $p = 0.001$) was also observed. However, the differences in phyllospheric bacterial communities based on geography were reduced after 1 year of storage (HN2 vs. YN2, $R^2 = 0.1028$, $p = 0.001$). These results indicated that the leaf position was the most important factor affecting the phyllospheric bacterial communities of TLs, followed by the sample location and then storage time.

Fig. 1 Alpha diversity value of TLs. **A** and **B**: Chao1 values of the four groups; **c** and **d**, observed values. (HN1, Henan TLs collected within 5 days; HN2, Henan TLs stored for 1 year; YN1, Yunnan TLs collected within 5 days; YN2, Yunnan TLs stored for 1 year; HN1 (or YN1, HN2, YN2)-B represents the upper leaves of HN1; HN1 (or YN1, HN2, YN2)-C represents the middle leaves; HN1 (or YN1, HN2, YN2)-X represents the bottom leaves.)

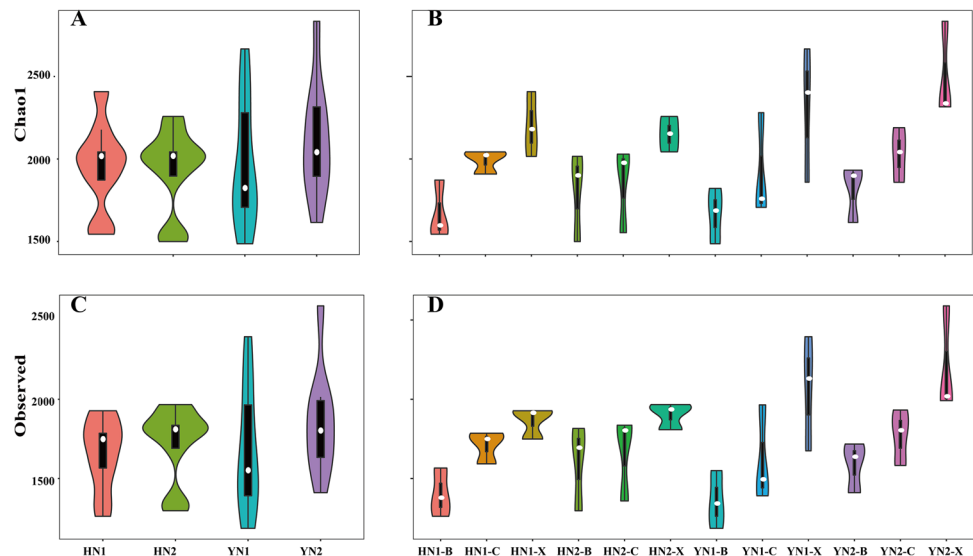


Fig. 2 PCoA plots of Bray–Curtis computed distances between TLs. PCoA results were based on the OTU level. (HN1, Henan TLs collected within 5 days; HN2, Henan TLs stored for 1 year; YN1, Yunnan TLs collected within five days; YN2, Yunnan TLs stored for 1 year; HN1 (or YN1, HN2, YN2)-B represents the upper leaves of HN1; HN1 (or YN1, HN2, YN2)-C represents the middle leaves; HN1 (or YN1, HN2, YN2)-X represents the bottom leaves)

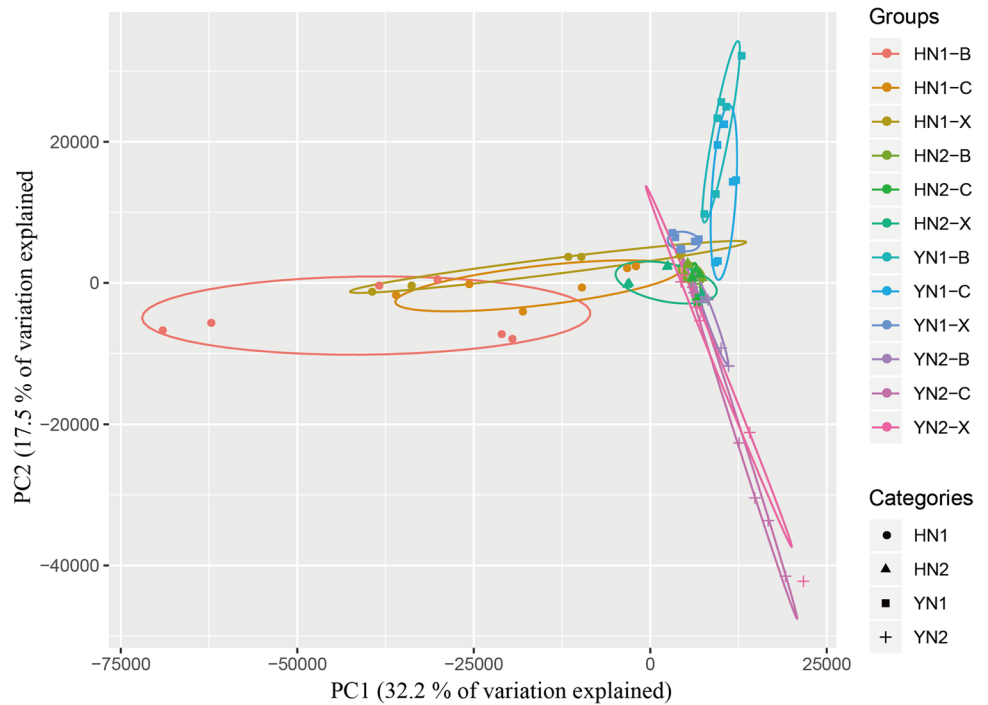
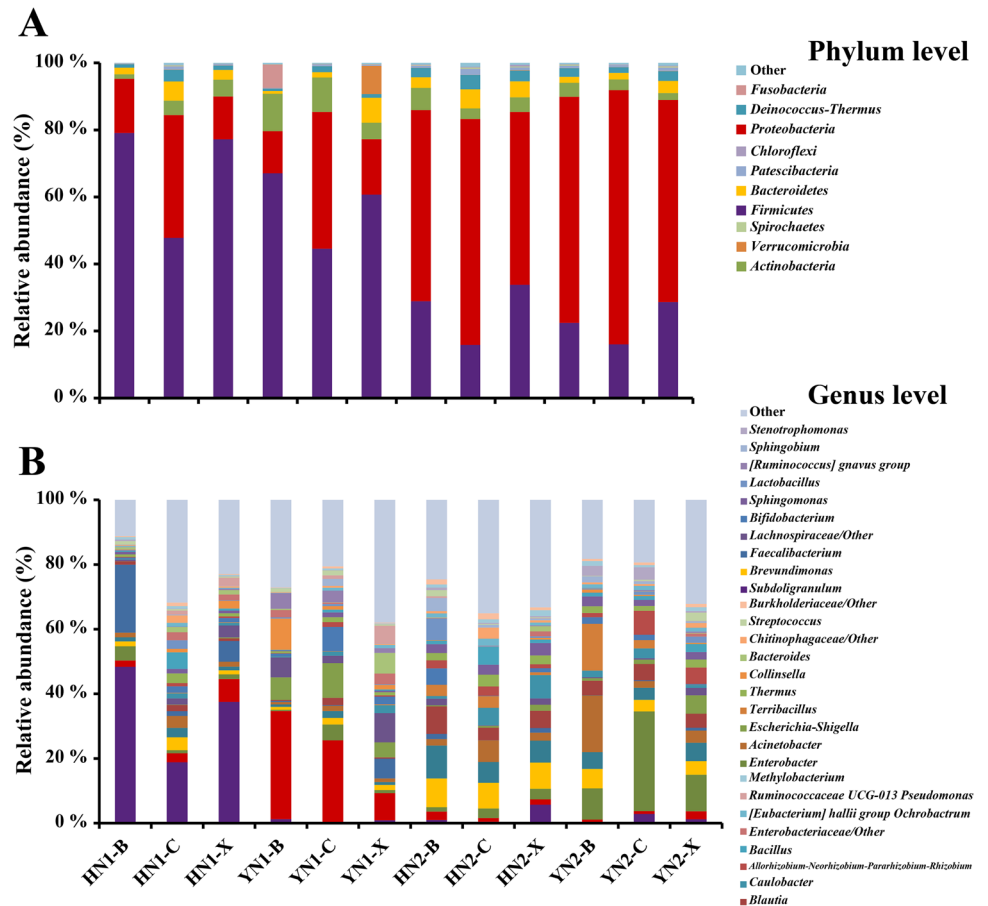


Fig. 3 Bacterial communities of TLs at the phylum level and genus level. (HN1, Henan TLs collected within 5 days; HN2, Henan TLs stored for 1 year; YN1, Yunnan TLs collected within 5 days; YN2, Yunnan TLs stored for 1 year; HN1 (or YN1, HN2, YN2)-B represents the upper leaves of HN1; HN1 (or YN1, HN2, YN2)-C represents the middle leaves; HN1 (or YN1, HN2, YN2)-X represents the bottom leaves)



Taxonomic analysis

The bacterial community composition of TLs is shown in Fig. 3. At the phylum level, *Firmicutes* and *Proteobacteria* were the two most dominant bacteria for all TLs. However, large shifts in bacterial community composition were apparent after 1 year of storage. *Firmicutes* was reduced from ~65.6% (HN1) to ~25.5% (HN2), while *Proteobacteria* was increased from 15.6% (HN1) to 56.8% (HN2). Similar results were observed for the TLs from YN Province; *Firmicutes* was reduced from ~61.3% (YN1) to ~24.1% (YN2), while *Proteobacteria* was increased from 14.8% (YN1) to 68.7% (YN2).

At the genus level, the most dominant genera were greatly different across all samples. On average, *Subdoligranulum* (~34.9%) was the predominant genus in the HN1 samples, while *Blautia* (~21.3%) was the predominant genus in the YN1 samples. Although the dominant genera were similar within both the HN1 and YN1 groups, the proportions of bacteria were much different across the different positions (B, C, and X). After 1 year of storage, *Brevundimonas* (~7.46%) and *Caulobacter* (~7.01%) were the two dominant genera in the HN2 samples, while *Enterobacter* (~15.2%) and *Acinetobacter* (~7.42%) were the two dominant genera in the YN2 samples.

Geographic and position-associated variation in bacterial communities

The heatmap in Fig. 4 shows the differences in the top 20 bacteria at the genus level. TLs of HN1 and YN1 were the original samples collected from Henan and Yunnan. From the heatmap, the genera of *Subdoligranulum*, *Thermus*, and *Acinetobacter* were obviously more abundant in HN1 than in YN1. In contrast, the genera *Blautia* and *Ruminococcus* were significantly more abundant in YN1 than those in HN1. Nevertheless, after 1 year of storage, these differences in bacterial communities were reduced between HN2 and YN2, which agreed with the PERMANOVA analysis results ((HN1 vs. YN1, $R^2 = 0.2280$, $p = 0.001$; HN2 vs. YN2, $R^2 = 0.1028$, $p = 0.001$). In addition, the bacterial community compositions of TLs from different positions (B, C, X) were much different in all area-based groups (HN1, YN1, HN2, and YN2). The hierarchical cluster results (Supplemental Figs. S1–S4) also suggested that the bacterial compositions were closely related to the positions of the TLs. The bacterial communities of TLs in the same positions were usually grouped together and were significantly different from those of TLs in other positions.

The geographic and position-associated variations in bacterial communities were then further explained by the statistically significantly different ($p < 0.05$) relative abundances between HN1, HN2, YN1, and YN2 (Fig. 5). There

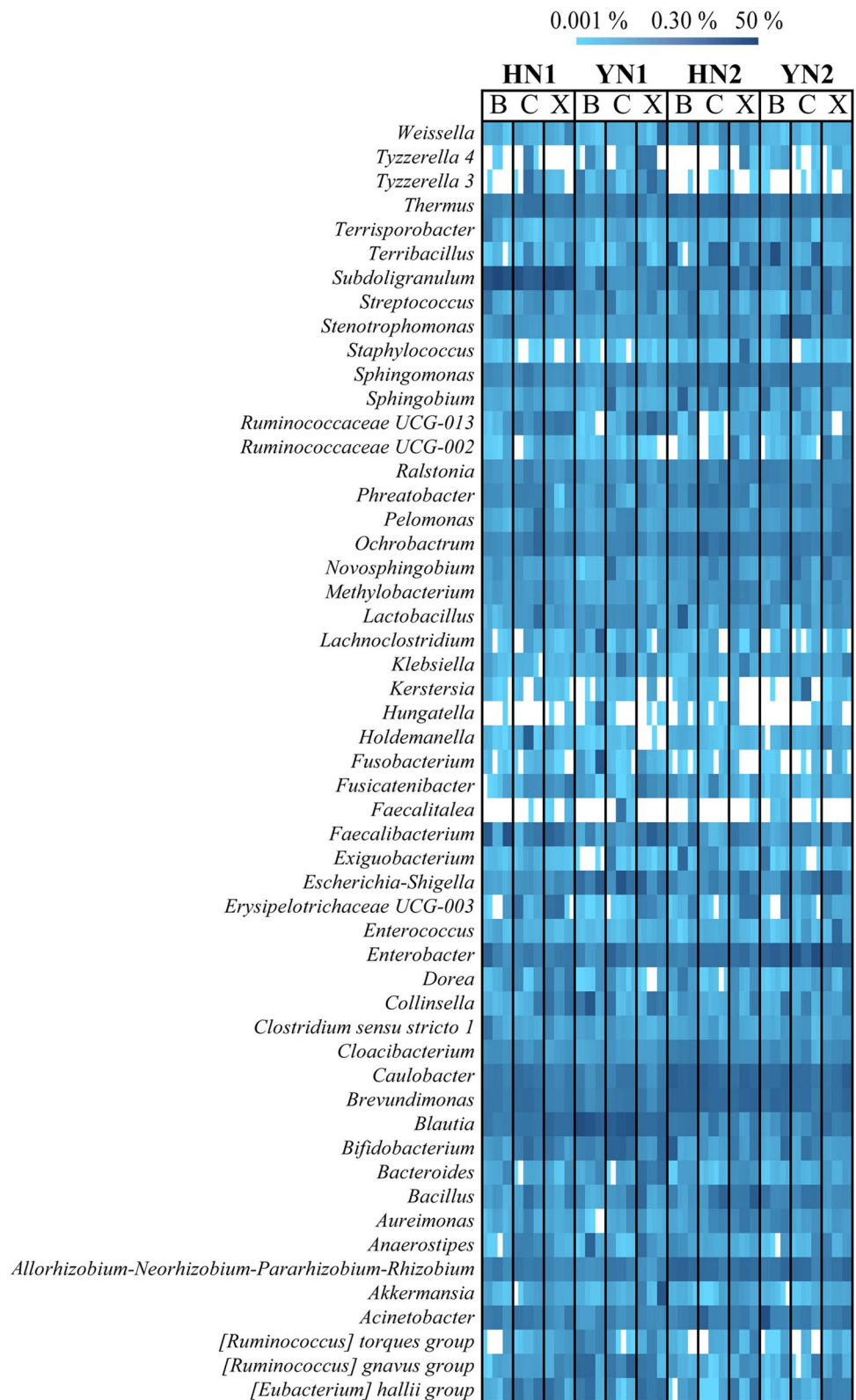
were 33 bacterial genera that showed significantly different abundances between HN1 and YN1 (Fig. 5A). Most of the different bacteria belonged to the phylum *Firmicutes*. Fifteen significantly differentially abundant bacteria were identified in the upper leaves (Position B), with 12 (*Tyzzarella*, *Sellimonas*, *Hungatella*, *Blautia*, etc.) in YN1 and 3 in HN1 (*Subdoligranulum*, *Paraprevotella*, *Brevibacterium*). Sixteen significantly differentially abundant bacteria were identified in the middle leaves (Position C), and 11 significantly differentially abundant bacteria were identified in the bottom leaves (Position X). These results implied that position and geography were both highly associated with the dominant bacteria present on the TLs.

The differences in dominant bacteria were also observed in the TLs after 1 year of storage. After comparing the storage times (HN1 vs. HN2, YN1 vs. YN2), the most significantly differentially abundant bacteria were identified in HN1 (Fig. 5B) and YN1 (Fig. 5C). Among 27 statistically significantly different genera of bacteria, 24 were found to be more abundant in HN1 at either one position or two positions (such as *Subdoligranulum* and *Butyrivibrio*) of TLs, and only 2 (*Staphylococcus* and *Blastomonas*) were found to be more abundant in HN2 from the bottom leaves (Position X). Similarly, 45 significantly different genera of bacteria were found between YN1 and YN2, of which 36 (including 16 of Position B, 12 of Position X, 3 of Position C and 5 of two positions) were more abundant in YN1 and only 10 were more abundant in YN2, with *Acidaminococcus* and *Barnesiella* being more abundant both in bottom leaves of YN1 and middle leaves of YN2. Twenty-three significantly different genera of bacteria were found between HN2 and YN2 (Fig. 5D), but most of the different genera of bacteria were either more abundant in HN2 or in YN2 leaves at different positions.

Functional prediction of bacterial communities

PICRUSt was employed to compare the predicted functional potential of bacterial microbiota from different TLs. In total, 328 KEGG level 3 modules were obtained from all TLs. Of these, 41 were significantly different between different TLs of different geographic locations ($p < 0.05$), and 72 (total) were significantly different between TLs from different positions (27 (C_vs_B), 11 (X_vs_B), 46 (X_vs_C), $p < 0.05$) (Supplemental Table S3). For KEGG level 1, metabolism was the most abundant pathway, including some common metabolic pathways in level 3 identified in tobacco leaves, such as starch and sucrose metabolism (0.4051 ~ 1.1667%), naphthalene degradation (0.1295 ~ 0.4382%), nicotinamide and nicotinamide metabolism (0.3445 ~ 0.4673%), nitrogen metabolism (0.6557 ~ 0.9422%), and nitrotoluene degradation (0.0225 ~ 0.1319%). Among these pathways, starch and sucrose metabolism and nitrotoluene degradation were

Fig. 4 Heatmap of bacterial communities in TLs originating from different provinces and subjected to different storage times. (HN1, Henan TLs collected within 5 days; HN2, Henan TLs stored for 1 year; YN1, Yunnan TLs collected within 5 days; YN2, Yunnan TLs stored for 1 year)



significantly different between the TLs originating from different areas (HN1 vs. YN1, Supplemental Fig. S5). Considering the storage time, the pathways of “naphthalene

degradation,” “starch and sucrose metabolism,” and “nicotinamide and nicotinamide metabolism” were significant in Henan TLs after 1 year storage (HN1 vs. HN2). Similarly,

Fig. 5 Geographic variation in bacterial communities and the effect of time storage on the microbiota. (HN1, Henan TLs collected within 5 days; HN2, Henan TLs stored for 1 year; YN1, Yunnan TLs collected within 5 days; YN2, Yunnan TLs stored for 1 year). **A**, HN1 vs. YN1; **B**, HN1 vs. HN2; **C**, YN1 vs. YN2; **D**, HN2 vs. YN2



four of the above pathways (except the nitrogen metabolism) shifted obviously in Yunnan TLs after storage (YN1 vs. YN2).

Human diseases ranged between 0.680 and 1.3106%, and both human diseases at level 1 and infectious diseases at level 2 were similar between TLs. However, when looking into the pathogenic bacteria of the infectious disease pathway, statistical significance ($p < 0.05$) was observed between the HN1 and YN1 TLs (Supplemental Fig. S6A). These pathogenic bacteria, including *Acinetobacter*, *Methylobacterium*, and *Escherichia-Shigella* were identified in all samples. Nevertheless, these differences were no longer statistically significant after 1 year of storage (HN2 vs. YN2) (Supplemental Fig. S6B). After 1 year of storage,

Acinetobacter and *Escherichia-Shigella* were significantly different between YN1 and YN2 (Supplemental Fig. S6D) but were not significantly different between HN1 and HN2 (Supplemental Fig. S6C).

Discussion

Recently, studies have shown that tobacco leaves are colonized by many bacteria and fungi, which could change over the process of tobacco product manufacturing and storage and could contribute to changes in the components of TLs (Di Giacomo et al. 2007; Ye et al. 2017). Smokers can be easily infected by these bacteria or fungi (Bagaitkar et al.

2009), and the leaf chemistry may also be closely related to the microbiota communities (Law et al. 2016). Thus, identifying the microbiota inhabiting tobacco leaves is of great significance to avoid risks to human health. This study identified the bacterial communities inhabiting tobacco leaves originating from different areas as well as different positions on the plant to help understand the potential microbial-related health risks associated with tobacco products made by using different tobacco leaves.

The results of this study revealed that the bacterial diversity of tobacco leaves was related to their geographic origins. Several previous studies have demonstrated that geographical origin is a critical factor for tobacco leaves, as the components of TLs were largely different (Sun et al. 2018; Zhang et al. 2013). Based on flavor styles, Chinese flue-cured tobacco leaves (FTLs) can be divided into three different categories, which are also related to geography (Zhang et al. 2020). FTLs from Yunnan tend to have fresh flavor styles, while FTLs from Henan have a strong flavor style. Moreover, the microbes also contributed significantly to desirable tobacco characteristics (Law et al. 2016; Zhang et al. 2020). However, how the production areas and positions of tobacco leaves affect the bacteria is rarely reported. The alpha- and beta-diversity were significantly different between the tobacco leaves from different areas (HN1 vs. YN1). In other plants, bacteria inhabiting the surface of plants are thought to be related to the environment as well as to geography (Kim et al. 2018; Padaga et al. 2000), which may contribute to the different qualities of these plants. The different leaf chemistry of plants from Yunnan and Henan has been measured and found to be related to climate factors, including rainfall, sunshine, and temperature (Zhao et al. 2013). When comparing the KEGG functional prediction results, “starch and sucrose metabolism” and “nitrotoluene degradation” were significantly different between HN1 and YN1 (Supplemental Fig. S5), further indicating that the different qualities and flavor styles of tobacco could be partly attributed to the microbiota inhabiting the TLs. Although these TLs belong to the same category (K326), the flavor styles are very different between HN and YN TLs in this study, and our results suggest that phyllospheric bacterial communities are significantly correlated with the flavor style of TLs.

Furthermore, the different bacterial communities of TLs from Henan and Yunnan were displayed by the bacterial composition. Both the Henan and Yunnan TLs were dominated by *Firmicutes* and *Proteobacteria* at the phylum level. However, the bacterial composition was different at the genus level, at which HN1 TLs were dominated by the genera *Subdoligranulum*, *Thermus*, and *Acinetobacter*, while YN1 TLs were dominated by the genera *Blautia* and *Ruminococcus*. By using traditional techniques, few specific bacterial species, such as *Bacillus* spp., *Actinomyces* spp., and

Pseudomonas spp., have been identified in tobacco products (Di Giacomo et al. 2007; Huang et al. 2010). The different dominant bacteria at the genus level identified in the present study indicate that the bacterial diversity may be underestimated by using culture-dependent techniques or other sequencing protocols. Several former studies showed that tobacco products (including TLs) harbor a rich and diverse bacterial microbiota (Chopyk et al. 2017; Di Giacomo et al. 2007; Han et al. 2016; Smyth et al. 2017; Tyx et al. 2016) using high-throughput sequencing technologies. These studies investigated the effects of brands, storage conditions, and times on tobacco products, but no reports were focused on the geographic variation of bacterial communities of TLs. The TLs used in this study were collected after harvest and curing and then stored under natural conditions in local places for 1 year. Therefore, the large difference in bacterial composition between them could be attributed to the environment where the TLs were harvested and stored.

Position-associated variation in bacterial communities was also observed. Different positions (bottom, middle and upper) of TLs are key factors related to their quality (Zhang et al. 2018). The growth stages and harvest times were found to affect the leaf chemistry, including organic acid and sugar contents (Xiang et al. 2010; Zhang et al. 2018). To the best of our knowledge, no reports have focused on the bacterial communities associated with growth positions of TLs. Here, the phyllospheric bacterial communities and diversity were both significantly different between tobacco leaves at different positions (B, C, and X) within HN1 and YN1. Moreover, the heatmap results also showed that bacterial compositions were closely related to the positions of TLs (Supplemental Figs. S1–S4), and the bacterial populations were grouped together. The dominant bacteria remained different after 1 year of storage (HN2 vs. YN2) between different TL positions. Although many tobacco characteristics were not investigated in this study, our results imply that leaf positions are a critical factor for the phyllospheric bacteria present on TLs. Moreover, based on the different qualities of leaves at various positions, our results indicated that the effects of leaf position on the microbiota could be attributed to the chemical compositions and qualities of TLs. However, the relationships between these qualities and phyllospheric bacteria should be further investigated by profiling the characteristics and bacterial communities at the same time.

The variations in phyllospheric bacterial communities and functions were also observed among FTLs under storage (HN1 vs HN2, YN1 vs YN2). The bacterial diversity and dominant bacteria were both significantly different after 1 year of storage (Fig. 3), which suggested that the storage process is a factor in the selection of phyllospheric bacteria present on TLs. This result agrees with those of previous studies that focused on the changes in the microbiota inhabiting tobacco leaves during tobacco processing, such

as fermentation and curing (Ye et al. 2017; Zhao et al. 2007). The functional prediction showed that several common metabolic pathways (such as “starch and sucrose metabolism”) in TLs shifted after storage (Supplemental Fig. S5), implying that the phyllospheric bacteria present on TLs are correlated with the metabolism of TLs during the storage time. For potential human pathogens, significant differences in *Acinetobacter* and *Escherichia-Shigella* were observed between YN1 and YN2, while no significant difference was observed between HN1 and HN2. This means that the risk to human health caused by potential pathogens harbored on TLs is related to sample locations.

Although the WHO has made many efforts to control the consumption of tobacco, more than one-third of the population is affected by cigarette smoking directly or indirectly (secondhand smoke) (Astell-Burt et al. 2018; Zhang et al. 2011). The adverse health effects of cigarettes (such as cancer, oral lesions, and nicotine addiction) are probably attributed to the components of FTLs, tobacco smoke, and the microorganisms present on TLs. In China, TLs are harvested and cured by farmers, and so-called flue-cured tobacco is then be purchased by industrial companies. Thousands of components have been identified in FTLs and smoke, and many are considered carcinogenic toxins (Stepanov et al. 2008; Talhout et al. 2011). During the manufacturing procedure, these bacteria inhabiting TLs may affect the quality of tobacco products and cause threats to human health (Law et al. 2016; Wang et al. 2018). Evidence that cigarette tobacco and other tobacco products potentially harbor microbes has been provided in previous studies (Sapkota et al. 2010; Smyth et al. 2017). In this study, many OTUs were identified as *Acinetobacter*, *Escherichia-Shigella*, *Pseudomonas*, *Methylobacterium*, and *Staphylococcus*, and all of these genera included potential human bacterial pathogens (Chattopadhyay et al. 2019). The *Acinetobacter* spp. are usually detected in the hospital environment (Kappstein et al. 2000), and *Escherichia-Shigella* are human pathogens causing intestinal disease (Bin et al. 2018; Khalil et al. 2018). *Methylobacterium* has been identified as a human pathogen causing peritonitis (Zhao et al. 2011). Regarding these potential human pathogens, significant differences ($p < 0.05$) in *Acinetobacter*, *Escherichia-Shigella*, and *Methylobacterium* were observed between HN1 and YN1. The geographic variations in potential human pathogens reveal that TLs harvested from different areas may cause different health risks for farmers. No significant difference in potential human pathogens was observed in Henan TLs after 1 year of storage (HN1 vs. HN2), while significant differences in *Acinetobacter* and *Escherichia-Shigella* were observed in Yunnan TLs (YN1 vs. YN2). These results implied that the initial human pathogens present on the TLs could be a risk to human health, and it is important to remove human pathogens before TLs are subjected to further processing.

Nevertheless, after 1 year of storage, no significant difference in potential human pathogens was observed between the TLs that originated from different provinces (HN2 vs. YN2). These results are consistent with those of previous studies showing that the bacteria inhabiting the surface of TLs could be altered after fermentation (Di Giacomo et al. 2007; Zhao et al. 2007) and suggest that the fermentation of TLs for 1–3 years is necessary for tobacco production, which will not only improve the qualities of TLs but also reduce the risk to human health.

In conclusion, the findings in this study suggest that the bacterial microbiota harbored by TLs varies between different geographic locations as well as different positions on the plant. In addition, the difference in bacterial diversity decreased after 1 year of storage. The human pathogens were also differed across positions and geography, but the differences will be reduced after storage. The public health implications of these results remain uncertain, but further studies are needed to characterize the changes in human pathogens during cigarette production and consumption, especially to illustrate the role of human pathogens in affecting tobacco users’ oral microbiota and health. In addition, the relationships between the characteristics of TLs and phyllospheric bacteria should also be further investigated.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00253-021-11671-7>.

Acknowledgements We are grateful to the employees of China Tobacco Henan Industrial Co., Ltd. for collecting the tobacco leaf samples. The project was supported in part by the Doctor Support Grants of Putian University.

Author contribution Study concepts: JY and ZZ. Study design: JY, YD, and QN. Experimental studies: YD, XQ, and JX. Data acquisition: YD, XQ, and JX. Data analysis: YD and XQ. Manuscript preparation: JY and XY. Manuscript revision: JY, XY, and ZZ.

Funding This study was funded by Doctor-Supported Grants of the Putian University.

Data Availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability The sequence data of this study have been deposited in the NCBI BioProject database under BioProject accession number PRJNA646576 (<https://www.ncbi.nlm.nih.gov/bioproject/>).

Declarations

Ethical approval This study was conducted without the use of human or animal participants.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

- Astell-Burt T, Zhang M, Feng X, Wang L, Li Y, Page A, Zhou M, Wang L (2018) Geographical inequality in tobacco control in China: multilevel evidence from 98 058 participants. *Nicotine Tob Res* 20:755–765. <https://doi.org/10.1093/ntr/ntx100>
- Bagaitkar J, Demuth DR, Scott DA (2009) Tobacco use increases susceptibility to bacterial infection. *Tob Induc Dis* 4:12. <https://doi.org/10.1186/1617-9625-4-12>
- Bin P, Tang Z, Liu S, Chen S, Xia Y, Liu J, Wu H, Zhu G (2018) Intestinal microbiota mediates enterotoxigenic *Escherichia coli*-induced diarrhea in piglets. *BMC Vet Res* 14:385. <https://doi.org/10.1186/s12917-018-1704-9>
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7:335–336. <https://doi.org/10.1038/nmeth.f.303>
- Chattopadhyay S, Smyth EM, Kulkarni P, Babik KR, Reid M, Hittle LE, Clark PI, Mongodin EF, Sapkota AR (2019) Little cigars and cigarillos harbor diverse bacterial communities that differ between the tobacco and the wrapper. *PLoS ONE* 14:e0211705. <https://doi.org/10.1371/journal.pone.0211705>
- Chopyk J, Chattopadhyay S, Kulkarni P, Clay E, Babik KR, Reid MC, Smyth EM, Hittle LE, Paulson JN, Cruz-Cano R, Pop M, Buehler SS, Clark PI, Sapkota AR, Mongodin EF (2017) Mentholation affects the cigarette microbiota by selecting for bacteria resistant to harsh environmental conditions and selecting against potential bacterial pathogens. *Microbiome* 5:22. <https://doi.org/10.1186/s40168-017-0235-0>
- 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* 73:825–837. <https://doi.org/10.1128/AEM.02378-06>
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27:2194–2200. <https://doi.org/10.1093/bioinformatics/btr381>
- Fisher MT, Bennett CB, Hayes A, Kargalioglu Y, Knox BL, Xu D, Muhammad-Kah R, 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. <https://doi.org/10.1016/j.fct.2011.11.035>
- Han J, Sanad YM, Deck J, Sutherland JB, Li Z, Walters MJ, Duran N, Holman MR, Foley SL (2016) Bacterial populations associated with smokeless tobacco products. *Appl Environ Microbiol* 82:6273–6283. <https://doi.org/10.1128/AEM.01612-16>
- Huang J, Yang J, Duan Y, Gu W, Gong X, Zhe W, Su C, Zhang KQ (2010) Bacterial diversities on unaged and aging flue-cured tobacco leaves estimated by 16S rRNA sequence analysis. *Appl Microbiol Biotechnol* 88:553–562. <https://doi.org/10.1007/s00253-010-2763-4>
- Kapstein I, Grundmann H, Hauer T, Niemeyer C (2000) Aerators as a reservoir of acinetobacter junii: an outbreak of bacteraemia in paediatric oncology patients. *J Hosp Infect* 44:27–30. <https://doi.org/10.1053/jhin.1999.0648>
- Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, Brewer TG, Engmann CM, Houtp ER, Kang G, Kotloff KL, Levine MM, Luby SP, MacLennan CA, Pan WK, Pavlinac PB, Platts-Mills JA, Qadri F, Riddle MS, Ryan ET, Shoultz DA, Steele AD, Watson JL, Sanders JW, Mokdad AH, Murray CJL, Hay SI, Reiner RC Jr (2018) Morbidity and mortality due to shigella and enterotoxigenic *Escherichia coli* diarrhoea: the Global Burden of Disease Study 1990–2016. *Lancet Infect Dis* 18:1229–1240. [https://doi.org/10.1016/s1473-3099\(18\)30475-4](https://doi.org/10.1016/s1473-3099(18)30475-4)
- Kim MS, Bae JW, Park EJ (2018) Geographic and host-associated variations in bacterial communities on the floret surfaces of field-grown broccoli. *Appl Environ Microbiol* 84:e02837-e2817. <https://doi.org/10.1128/AEM.02837-17>
- Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepille 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* 31:814–821. <https://doi.org/10.1038/nbt.2676>
- Law AD, Fisher C, Jack A, 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. <https://doi.org/10.1007/s00248-016-0754-4>
- Magoč T, Salzberg SL (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27:2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
- McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>
- McMurdie PJ, Paulson JN (2015) Biomformat: an interface package for the BIOM file format. R/Bioconductor package version 1.0.0. Bioconductor. <http://bioconductor.org/packages/release/bioc/html/biomformat.html>. Accessed 20 May 2021
- Nurnasari E, Subiyakto (2015) Chemical compound of essential oils from three different area of tobacco leaves (*Nicotiana tabacum* L.) in Indonesia. *J Life Sci Biomed* 5:163–166
- Padaga M, Heard GM, Paton JE, Fleet GH (2000) Microbial species associated with different sections of broccoli harvested from three regions in Australia. *Int J Food Microbiol* 60:15–24. [https://doi.org/10.1016/s0168-1605\(00\)00329-9](https://doi.org/10.1016/s0168-1605(00)00329-9)
- Parks DH, Tyson GW, Hugenholtz P, Beiko RG (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30:3123–3124. <https://doi.org/10.1093/bioinformatics/btu494>
- Ploner A (2012) Heatplus: heatmaps with row and/or column covariates and colored clusters. R Package Version 2 (0). Bioconductor. <http://bioconductor.org/packages/release/bioc/html/Heatplus.html>. Accessed 20 May 2021
- Rooney AP, Swezey JL, Wicklow DT, McAtee MJ (2005) Bacterial species diversity in cigarettes linked to an investigation of severe pneumonitis in U.S. military personnel deployed in operation Iraqi freedom. *Curr Microbiol* 51:46–52. <https://doi.org/10.1007/s00284-005-4491-z>
- Sapkota AR, Berger S, Vogel TM (2010) Human pathogens abundant in the bacterial metagenome of cigarettes. *Environ Health Perspect* 118:351–356. <https://doi.org/10.1289/ehp.0901201>
- Smyth EM, Kulkarni P, Clay E, Stanfill S, Tyx R, Maddox C, Mongodin EF, Sapkota AR (2017) Smokeless tobacco products harbor diverse bacterial microbiota that differ across products and brands. *Appl Microbiol Biotechnol* 101:5391–5403. <https://doi.org/10.1007/s00253-017-8282-9>
- Stepanov I, Jensen J, Hatsukami D, Hecht SS (2008) New and traditional smokeless tobacco: comparison of toxicant and carcinogen levels. *Nicotine Tob Res* 10:1773–1782. <https://doi.org/10.1080/14622200802443544>

- Sun JG, Jie-Wang HE, Feng-Guang WU, Shu-Xing TU, Yan TJ, Hui SI, Hao X (2011) Comparative analysis on chemical components and sensory quality of aging flue-cured tobacco from four main tobacco areas of China. *Agric Sci China* 10:1222–1231. [https://doi.org/10.1016/S1671-2927\(11\)60113-2](https://doi.org/10.1016/S1671-2927(11)60113-2)
- Sun B, Zheng AH, Zhang F, Wei KS, Chen Q, Luo Y, Zhang Y, Wang XR, Lin FC, Yang J, Tang HR (2018) Metabolic profiles of Cuibi-1 and Zhongyan-100 flue-cured tobacco leaves in different growing regions by gas chromatography/mass spectrometry. *R Soc Open Sci* 5:180261. <https://doi.org/10.1098/rsos.180261>
- Talhout R, Schulz T, Florek E, van Benthem J, Wester P, Opperhuizen A (2011) Hazardous compounds in tobacco smoke. *Int J Environ Res Public Health* 8:613–628. <https://doi.org/10.3390/ijerph8020613>
- Tyx RE, Stanfill SB, Keong LM, Rivera AJ, Satten GA, Watson CH (2016) Characterization of bacterial communities in selected smokeless tobacco products using 16S rDNA analysis. *PLoS ONE* 11:e0146939. <https://doi.org/10.1371/journal.pone.0146939>
- Wang F, Zhao H, Xiang H, Wu L, Men X, Qi C, Chen G, Zhang H, Wang Y, Xian M (2018) Species diversity and functional prediction of surface bacterial communities on aging flue-cured tobacco. *Curr Microbiol* 75:1306–1315. <https://doi.org/10.1007/s00284-018-1525-x>
- Wei X, Deng X, Cai D, Ji Z, Wang C, Yu J, Li J, Chen S (2014) Decreased tobacco-specific nitrosamines by microbial treatment with *Bacillus amyloliquefaciens* DA9 during the air-curing process of burley tobacco. *J Agric Food Chem* 62:12701–12706. <https://doi.org/10.1021/jf504084z>
- Xiang G, Yang H, Yang L, Zhang X, Cao Q, Miao M (2010) Multivariate statistical analysis of tobacco of different origin, grade and variety according to polyphenols and organic acids. *Microchem J* 95:198–206. <https://doi.org/10.1016/j.microc.2009.12.001>
- Ye J, Yan J, Zhang Z, Yang Z, Liu X, Zhou H, Wang G, Hao H, Ma K, Ma Y, Mao D, Yang X (2017) The effects of threshing and redrying on bacterial communities that inhabit the surface of tobacco leaves. *Appl Microbiol Biotechnol* 101:4279–4287. <https://doi.org/10.1007/s00253-017-8143-6>
- Zhang J, Ou JX, Bai CX (2011) Tobacco smoking in China: prevalence, disease burden, challenges and future strategies. *Respirology* 16:1165–1172. <https://doi.org/10.1111/j.1440-1843.2011.02062.x>
- Zhang L, Wang X, Guo J, Xia Q, Zhao G, Zhou H, Xie F (2013) Metabolic profiling of Chinese tobacco leaf of different geographical origins by GC-MS. *J Agric Food Chem* 61:2597–2605. <https://doi.org/10.1021/jf400428t>
- Zhang L, Zhang X, Ji H, Wang W, Liu J, Wang F, Xie F, Yu Y, Qin Y, Wang X (2018) Metabolic profiling of tobacco leaves at different growth stages or different stalk positions by gas chromatography–mass spectrometry. *Ind Crops Prod* 116:46–55. <https://doi.org/10.1016/j.indcrop.2018.02.041>
- Zhang Q, Geng Z, Li D, Ding Z (2020) Characterization and discrimination of microbial community and co-occurrence patterns in fresh and strong flavor style flue-cured tobacco leaves. *Microbiologyopen* 9:e965. <https://doi.org/10.1002/mbo3.965>
- Zhao M, Wang B, Li F, Qiu L, Li F, Wang S, Cui J (2007) Analysis of bacterial communities on aging flue-cured tobacco leaves by 16S rDNA PCR-DGGE technology. *Appl Microbiol Biotechnol* 73:1435–1440. <https://doi.org/10.1007/s00253-006-0625-x>
- Zhao JJ, Chopra T, Bernabela L, Revankar SG (2011) Peritonitis due to *Methylobacterium mesophilicum* complicating continuous ambulatory peritoneal dialysis. *Infect Dis Clin Pract* 19:83–84. <https://doi.org/10.1097/IPC.0b013e3181f5eac5>
- Zhao Y, Zhao C, Lu X, Zhou H, Li Y, Zhou J, Chang Y, Zhang J, Jin L, Lin F, Xu G (2013) Investigation of the relationship between the metabolic profile of tobacco leaves in different planting regions and climate factors using a pseudotargeted method based on gas chromatography/mass spectrometry. *J Proteome Res* 12:5072–5083. <https://doi.org/10.1021/pr400799a>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.