

ORIGINAL ARTICLE

Open Access



# Assessment of 16S rRNA gene primers for studying bacterial community structure and function of aging flue-cured tobaccos

Fan Wang<sup>1,2</sup>, Xiao Men<sup>1,2</sup>, Ge Zhang<sup>1,2</sup>, Kaichao Liang<sup>3</sup>, Yuhua Xin<sup>3</sup>, Juan Wang<sup>3</sup>, Aijun Li<sup>3</sup>, Haibo Zhang<sup>1,2\*</sup>, Haobao Liu<sup>3,4\*</sup> and Lijun Wu<sup>5\*</sup>

## Abstract

Selection of optimal primer pairs in 16S rRNA gene sequencing is a pivotal issue in microorganism diversity analysis. However, limited effort has been put into investigation of specific primer sets for analysis of the bacterial diversity of aging flue-cured tobaccos (AFTs), as well as prediction of the function of the bacterial community. In this study, the performance of four primer pairs in determining bacterial community structure based on 16S rRNA gene sequences in AFTs was assessed, and the functions of genes were predicted using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt). Results revealed that the primer set 799F–1193R covering the amplification region V5V6V7 gave a more accurate picture of the bacterial community structure of AFTs, with lower co-amplification levels of chloroplast and mitochondrial genes, and more genera covered than when using the other primers. In addition, functional gene prediction suggested that the microbiome of AFTs was involved in kinds of interested pathways. A high abundance of functional genes involved in nitrogen metabolism was detected in AFTs, reflecting a high level of bacteria involved in degrading harmful nitrogen compounds and generating nitrogenous nutrients for others. Additionally, the functional genes involved in biosynthesis of valuable metabolites and degradation of toxic compounds provided information that the AFTs possess a huge library of microorganisms and genes that could be applied to further studies. All of these findings provide a significance reference for researchers working on the bacterial diversity assessment of tobacco-related samples.

**Keywords:** Microbial community diversity, 16S rRNA gene sequencing, Optimal primer sets, Aging flue-cured tobaccos, PICRUSt, Functional profiling prediction

## Introduction

Microorganisms are dominant drivers of biogeochemical processes (Shinichi et al. 2015), yet a large number of microorganisms ( $\geq 99\%$ ) from the environment remain uncultivated in the laboratory, and these regulate ecosystem processes and even affect our lives (Kaeberlein et al.

2002; Su et al. 2012). Therefore, characterizing microbial community diversity can have theoretical and practical significance in understanding the relationships between microorganism and their hosts, treatment of environmental pollutants, utilization of microbial resources, and human medical health. 16S rRNA gene sequences are considered the predominant and most reliable approach for studying microbial structures in humans, the guts of animals, natural habitats, and the rhizospheres of plants (Beckers et al. 2016). 16S rRNA gene variable regions and the primer sets used for amplification of 16S rRNA gene sequences can result in significantly different bacterial community profiles (Jason et al. 2013; Prosdocimi et al. 2013). Determining optimal primer pairs for 16S rRNA

\*Correspondence: zhanghb@qibebt.ac.cn; liuhaobao@caas.cn; 2818429250@qq.com

<sup>1</sup> CAS Key Laboratory of Biobased Materials, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, China

<sup>4</sup> Tobacco Research Institute of Chinese Academy of Agriculture Sciences, Qingdao 266101, Shandong, China

<sup>5</sup> Yunnan Academy of Tobacco Sciences, Kunming 650106, China  
Full list of author information is available at the end of the article

gene sequencing, therefore, plays an important role in microorganism diversity analysis of special hosts.

Several researchers have suggested that the primers V3F and V4R, covering the V3V4 amplification region, should be the preferred primer set for future studies to achieve accurate bacterial community diversities, and a large number of studies have focused on this region (Castelino et al. 2017; Cai et al. 2013; Parulekar et al. 2017; Takahashi et al. 2014). Analogously, for plant samples, not including subterranean plant parts but only above-ground parts such as shoot-tip, leaves, etc., the V3V4 regions has been used predominantly for identifying 16S rRNA gene sequences (Thomas and Sekhar 2017; Tyx et al. 2016; Wang et al. 2018). Nevertheless, due to the homology between bacterial 16S rRNA genes and plant chloroplast and mitochondrial DNA, plant DNA may be coextracted during bacterial DNA extraction from various plant samples, which can cause contaminating sequences (Dyall et al. 2004; Raven 1970). Usually, a high abundance of *Cyanobacteria* at the phylum level, unidentified at the genus level, present in raw sequencing results indicates the occurrence of contaminating sequences from an organellar origin such mitochondrial and/or chloroplast DNA (Beckers et al. 2016). To exclude the interference of contaminating sequences, a common method is to filter out the chloroplast DNA, mitochondrial DNA, and other unknown sequences during data processing. However, this might not reveal the real microbiome structures of samples. To date, although several studies have focused on characterization of bacterial communities in different tobaccos using 16S rRNA gene sequences with the V3V4 and V5V6V7 primers (Huang et al. 2010; Tyx et al. 2016), few report have showed optimized primers for bacterial microbiome investigation of tobaccos, while there are as many as 10,000 chloroplast DNA copies in tobacco leaf cells (Shaver et al. 2006). Microorganisms from aging flue-cured tobaccos (AFTs) would be useful in producing valuable products, treating wastewater, and creating biofuels and a wide range of chemicals and enzymes. There is, therefore, an urgent need to investigate the perfect primers for tobacco bacterial microbiome research.

PICRUSt is an optimal tool to predict the functions of a microbial community's metagenome from its 16S profile. It predicts genes presented in organisms and uses existing annotations of gene content employing the KEGG (Kyoto Encyclopedia of Genes and Genomes) Orthology and Clusters of Orthologs Groups (Langille et al. 2013). PICRUSt has been used to determine the potential functions of the microbial community in a number of samples (Koo et al. 2017a, b; Wu et al. 2016), which might provide information for development and utilization of microorganism resources. Some studies showed that bacteria

in tobacco could generate toxins and pro-inflammatory biomolecules and generate nitrite (Tyx et al. 2016). Thus, characterizing bacterial diversity in tobacco samples and predicting the potential functions of the bacterial community might be a good way to reduce the levels of certain harmful compounds in tobacco and improve the quality of tobacco products, and other products, such as coffee berry, tea, and sauce, etc.

In this study, we used four 16S rRNA gene primer sets to investigate the bacteria in tobacco samples to select the optimal primer sets for characterizing bacterial diversity in tobacco samples. Moreover, the bacterial community function was further predicted using PICRUSt based on 16S rRNA gene sequencing for exploring microorganisms that produced natural products and degraded toxic compounds. The aim of this study was to provide a valuable reference to peer researchers working on bacterial diversity determination of tobacco-related samples, and other plant samples, using 16S rRNA gene sequencing technology, and provide useful information about microorganisms in AFTs for improving the quality and reducing the levels of certain harmful compounds in tobaccos and other fermentation products in further studies.

## Materials and methods

### Experimental design

High-throughput sequencing of 16S rRNA genes was employed to investigate the optimized primer set which could be used to improve the determination of the microbial diversity of AFTs. To evaluate bacterial 16S rRNA gene primer pairs in regard to bacterial diversity coverage and/or accurate taxonomic assignment, four primers pairs (V1F–V3R, V3F–V4R, V4F–V5R, and V5F–V7R) targeting different 16S rRNA gene regions were employed. After individual PCR (Polymerase Chain Reaction) amplification, a total of four fragments were obtained, as shown in Table 1. To minimize non-specific amplification and investigate the potential biases of the Illumina high-throughput sequencing technology, each 16S primer carried a unique eight-base sequence for each sample. Finally, the generated sequencing reads were analyzed by subsequent bioinformatics. In addition, PICRUSt (Langille et al. 2013) was used to predict the functional composition of the metagenome we obtained from the Illumina MiSeq platform analysis of the samples.

### Sampling and DNA isolation

Samples were collected from Honghe cigarette factory (Qujing, Yunnan) and Chuxiong cigarette factory (Yuanjiang, Yunnan), respectively. These samples were stored and transported using an ice bath before reaching the laboratory and then stored at  $-20^{\circ}\text{C}$ . No specific permits were required for this study.

**Table 1 Primers used in current study**

Primer pairs	Primer sequence (5'-3')	M	References
8F	TGGAGAGTTTGATCCTGG CTCAG	V1V2V3	Sun et al. (2014)
533R	TACCGCGGCTGCTGGCAC		
336F	GTACTCCTACGGGAG GCAGCA	V3V4	Munyaka et al. (2015)
806R	GTGGACTACHVGGGTWT CTAAT		
515F	GTGCCAGCMGCCGCG GTAA	V4V5	Tuan et al. (2014)
909R	CCCCGYCAATTCMTT RAGT		
799F	AACMGGATTAGATACC CKG	V5V6V7	Beckers et al. (2016)
1193R	ACGTCATCCCCACCTTCC		

Primers are indicated as forward (F) or reverse (R)

M, Hypervariable region of the 16S rRNA operon targeted by primer pairs

An UltraClean<sup>®</sup> Soil DNA Isolation Kit (Mobio Inc., Carlsbad, CA, USA) was used to extract total DNA from the collected samples individually according to the manufacturer's manual. The DNA extracted from three technical replicates of each sample was pooled into one DNA sample to minimize any potential DNA extraction bias. The concentration of each DNA extract was determined with 1% agarose gel electrophoresis and a Thermo NanoDrop 1000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA).

#### PCR amplification and Illumina MiSeq sequencing

As been shown in Table 1, the V1V2V3 region of the bacterial 16S rRNA gene was amplified with the universal primers 8F–533R. The PCR program was as follows: 95 °C for 5 min, 26 cycles at 95 °C for 45 s, 55 °C for 50 s, and 72 °C for 45 s, with a final extension of 72 °C for 10 min. The V3V4 region was amplified with the primers 336F–806R following the PCR program 95 °C for 5 min, 27 cycles at 95 °C for 45 s, 50 °C for 50 s, and 72 °C for 45 s, with a final extension of 72 °C for 10 min. The primers 515F–909R covered the V4V5 amplification regions and employed the following PCR program: 95 °C for 5 min, 28 cycles at 95 °C for 45 s, 55 °C for 50 s, and 72 °C for 45 s, with a final extension of 72 °C for 10 min. The V5V6V7 region was amplified with the primers 799F–1193R. The PCR program was 95 °C for 5 min, 27 cycles at 95 °C for 45 s, 55 °C for 50 s, and 72 °C for 45 s, with a final extension of 72 °C for 10 min. All primers contained an 8-nucleotide barcode sequence unique to each sample. All reactions were performed in triplicate in 50 µL volumes containing 5 µL of 10× Pyrobest Buffer, 4 µL of 2.5 mM dNTPs, 2 µL of each primer (10 µM),

0.3 µL of Pyrobest DNA Polymerase (2.5 U/µL, TaKaRa, Japan), and 30 ng of template DNA.

Amplicons were extracted from 2% agarose gels and purified using an AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, USA) according to the manufacturer's instructions and quantified using QuantiFluor<sup>™</sup>-ST (Promega Corporation, Madison, WI, USA). Purified amplicons were pooled in equimolar amounts and paired-end sequenced (2 × 300) on an Illumina MiSeq platform according to standard protocols.

#### Bioinformatic analysis

The extraction of high-quality sequences was firstly performed with the QIIME package (Gregory et al. 2010). Raw sequences were selected based on sequence length, quality, primer, and tag, and the low-quality sequences were removed. The Illumina Miseq sequencing data have been deposited in NCBI Sequence Read Archive database with the SRA accession number SRP139912. The unique sequence set was classified into operational taxonomic units (OTUs) under the threshold of 97% identity using UCLUST (Edgar 2010). Chimeric sequences were identified and removed using Usearch (version 8.0.1623). The taxonomy of each 16S rRNA gene sequence was analyzed with UCLUST against the Silva119 16S rRNA gene database using a confidence threshold of 90%.

#### Prediction of the functional composition of the metagenome using PICRUSt

The PICRUSt software package (Langille et al. 2013) was used to infer the potential genetic capability and specific contributions of bacterial taxa to the metagenomes of the tobacco samples. PICRUSt requires a phylogenetic tree of marker genes that includes complete reference genomes and the sequences from the samples under study. For this research, we used the KEGG database (Kanehisa 2000) for annotations, and the 16S rRNA gene sequence from each of these genomes was obtained from the United States Department of Energy Joint Genomic Institute's Integrated Microbial Genomes (IMG) database (Markowitz et al. 2012).

## Results

#### Illumina MiSeq sequencing

Two samples were collected from Honghe cigarette factory (Qujing, Yunnan) and Chuxiong cigarette factory (Yuanjiang, Yunnan), and sent for high-throughput sequencing of the 16S rRNA gene with four selected bacterial 16S rRNA gene primer pairs (Table 1). Generated reads and OTUs from different primers are shown in Table 2. Rarefaction analysis was performed to evaluate whether the size of the clone library represented the diversity in the original samples (Zhao et al. 2007).

**Table 2** Numbers of sequences analyzed for the two samples

Regions	Samples	Reads	OTU	Ace	Chao	Shannon	Simpson	Coverage	Observed species
V1V2V3	A1	17,349	133	133.51	134.99	1.83	0.637187	0.9981	127
	B1	9763	88	89.39	89.52	2.58	0.448405	0.9994	88
V3V4	A1	18,879	175	178.99	179.67	3.21	0.362518	0.9995	173
	B1	21,213	170	171.17	170.37	3.83	0.203652	0.9992	168
V4V5	A1	35,591	211	216.33	214.51	2.64	0.466091	0.9995	209
	B1	39,179	259	259.77	260.03	3.46	0.319225	0.9997	258
V5V6V7	A1	23,913	230	235.58	234.11	5.59	0.038895	0.9989	226
	B1	21,074	233	236.45	238.65	5.41	0.046749	0.9993	232

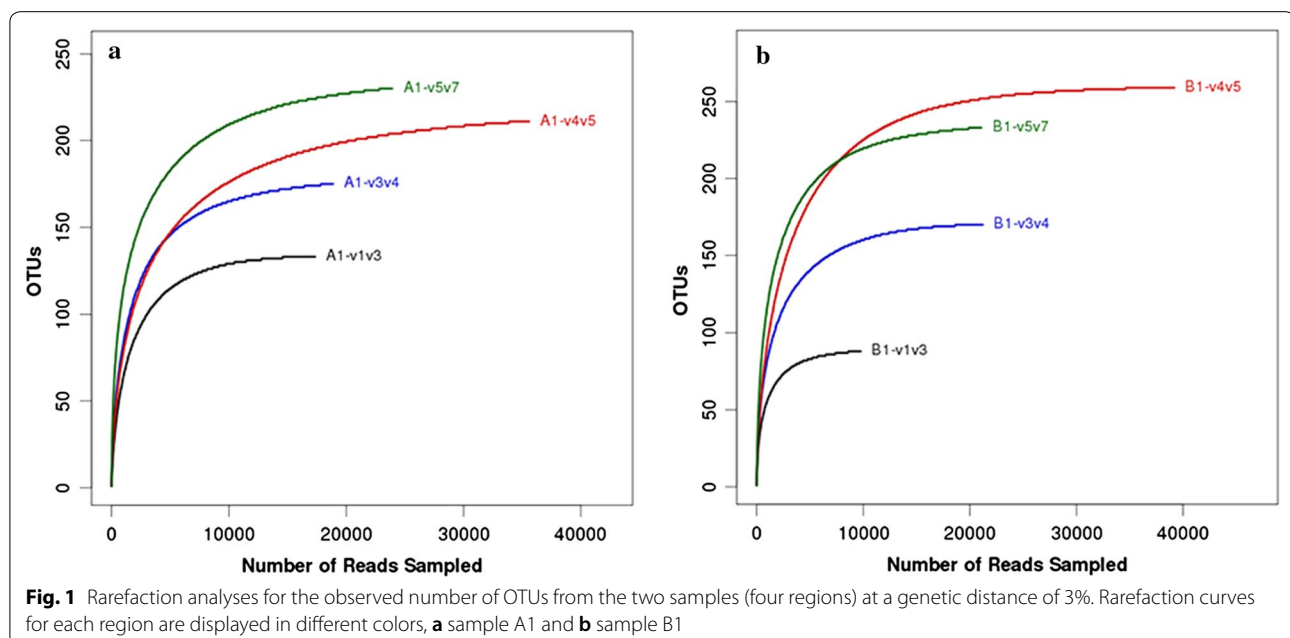
Operational Taxonomic Units (OTUs), estimated OTU richness, sample coverage, and diversity indices of Shannon and Simpson were calculated from 16S rRNA gene sequences of two samples

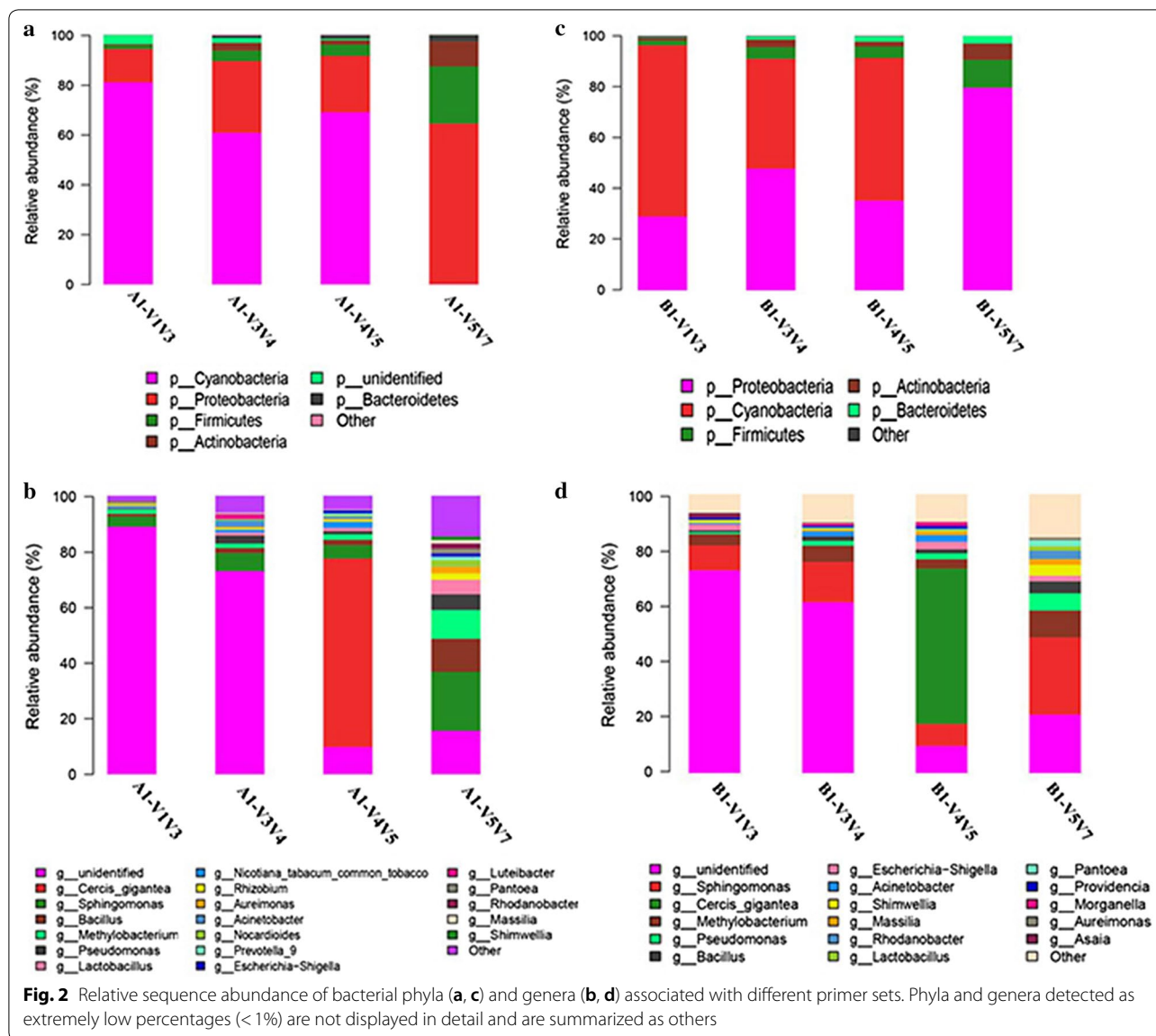
In our study, individual rarefaction curves did reflect that the primers give an acceptable measure of species diversity (Fig. 1). The comparison of the primer pairs (four regions) revealed the highest OTU richness and observed species for the V4V5 and V5V6V7 regions. The two samples yielded similar numbers of OTUs with 211 OTUs (A1-V4V5), 259 OTUs (B1-V4V5), 230 OTUs (A1-V5V6V7), and 233 OTUs (B1-V5V6V7), and these were similar to the number of observed species, 209 (A1-V4V5), 258 (B1-V4V5), 226 (A1-V5V6V7), and 232 (B1-V5V6V7), respectively (Table 2).

#### Comparison of diversity and abundance of bacterial taxa

Samples exhibited a similar range of taxonomic diversity at the phylum and genus levels. Overall, 18 and 16

genera from five phyla were detected with an abundance of at least 0.1% in the two samples, respectively (Fig. 2). At the phylum level (Fig. 2a, b), detection of bacterial taxa was dependent on the choice of 16S regions: e.g., *Cyanobacteria* (V1V2V3 > V3V4 > V4V5 > V5V6V7), *Proteobacteria* and *Actinobacteria* (V5V6V7 > V3V4 > V5V6 > V1V2V3), and *Firmicutes* (V5V6V7 > V3V4 or V5V6 > V1V2V3). At genus level also, the results showed that the abundance of detected bacteria was closely related to the covered regions of the selected primer pairs. For example, *Sphingomonas*, *Bacillus*, *Methylobacterium*, *Lactobacillus*, *Nocardioides*, and *Pseudomonas* showed high abundance with primers 799F–1193R, and extremely low abundance with primers 8F–533R. Similar cases were also observed for

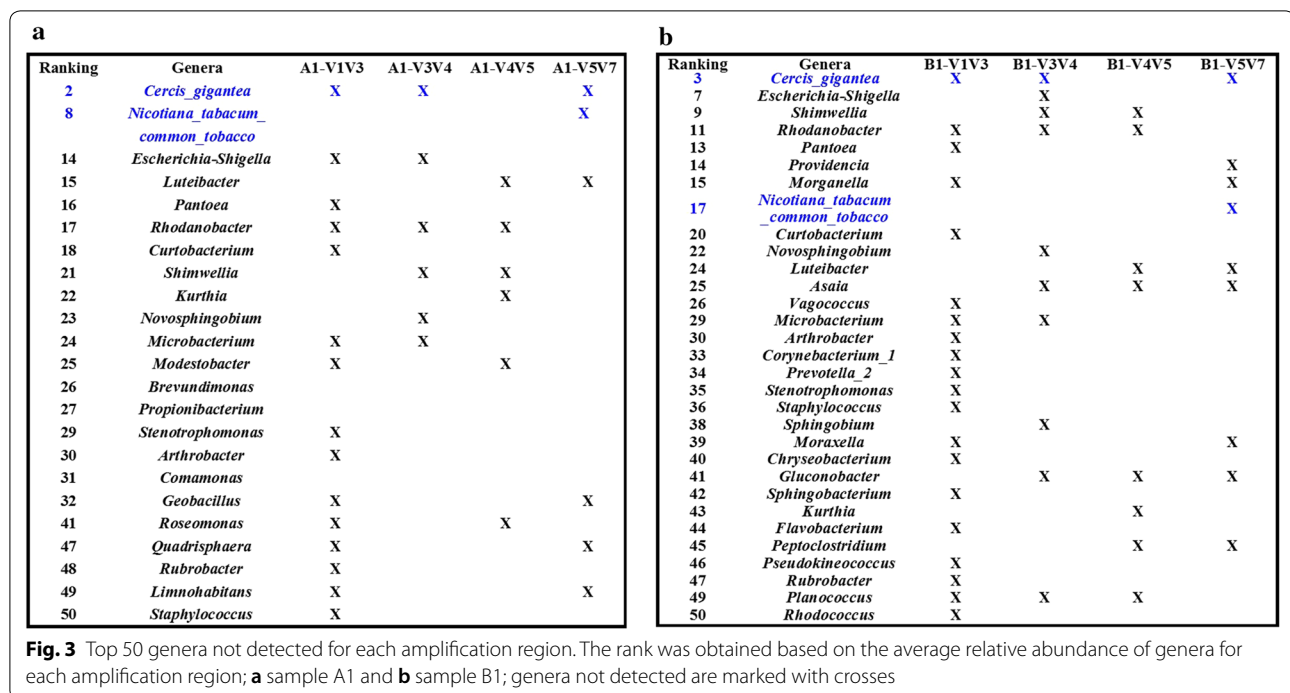




*Rhizobium*, *Aureimonas*, *Prevotella\_9*, and others. *Pantoea* and *Rhodanobacter* displayed high abundance with primers 799F–1193R, but were not detection with the other primers. Unidentified taxa showed high abundance with 8F–533R, but low abundance with primers 515F–909R and 799F–1193R. Of note, *Cercis gigantea* and *Nicotiana tabacum* (common tobacco) were two plant species detected. *Cercis gigantea* was detected only with primers 515F–909R with high abundance, and *N. tabacum* (common tobacco) was not detected with primers 799F–1193R. Figure 3 further illustrates that primers 799F–1193R amplified more genera than the other primers from the top 50 genera.

**Primers for the V5–V7 region are highly recommended**

The primers 799F–1193R covering region V5V6V7 performed better than the others, which was evidenced as follows. First, primers 799F–1193R had low co-amplification levels of chloroplast and mitochondrial genes, which was concluded from the absence of detection of *Cyanobacteria* at the phylum level and *C. gigantea* and *N. tabacum* (common tobacco) at the genus level (Fig. 2). Second, primers 799F–1193R covered more genera than the other primers (Figs. 2c, d, 3). Third, primers 799F–1193R gave the highest OTU richness (Table 2). To achieve a relatively accurate assessment of bacterial community structure and diversity, here we highly recommend 799F–1193R as the preferred



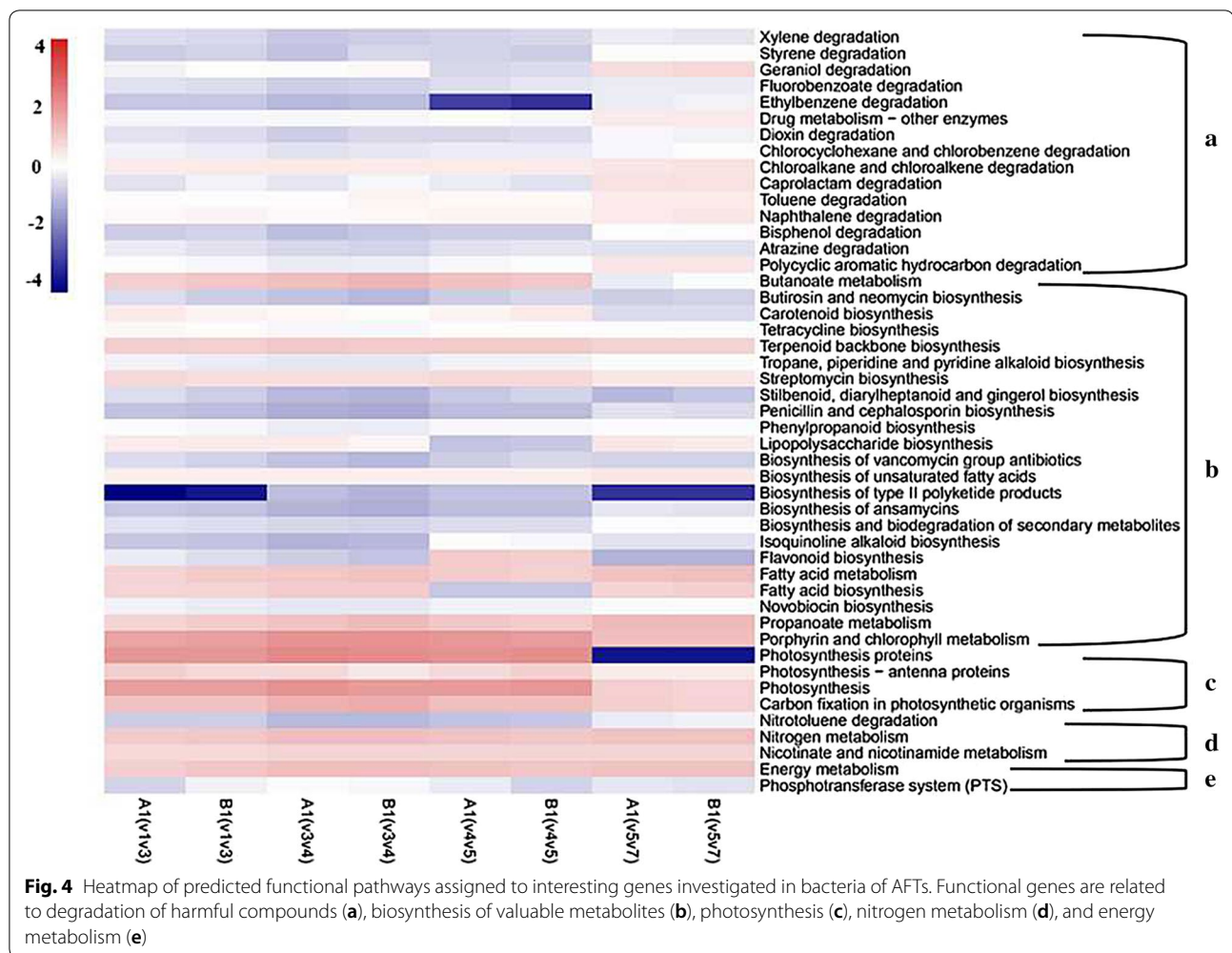
primers for future studies of tobacco-related samples, and even other plant samples, such as ensilage, etc.

### Functional gene prediction

AFTs microbiome functions were predicted with PICRUSt and annotated using the KEGG database, which showed that the microbiome of AFTs is involved in diverse pathways at level 3 KEGG Orthology (Additional file 1: Tables S1–S4). Functional genes related to degradation of harmful compounds, biosynthesis of valuable metabolites, photosynthesis, nitrogen metabolism, and energy metabolism were interestingly investigated. The heatmap obtained from the analysis of interesting functional genes is presented in Fig. 4. The abundance of photosynthesis-related genes detected, especially photosynthesis proteins, was much lower using the V5V6V7 region, and other genes showed slightly significant differences in abundance when using different 16S regions. This analysis demonstrated that primers 799F–1193R covering the V5V6V7 region as the preferred sequencing region should be used in future tobacco-related samples studies. Moreover, several remarkable genes were involved in biosynthesis of flavors and fragrances, and degradation of harmful compounds, especially nicotine, which could be employed to enhance the quality of tobacco leaves during the aging process. Besides, the results concluded that bacteria on AFTs are important microorganisms and natural resources of genes for future exploration.

### Discussion

There is no uniform argument for which primer set we should employ for particular samples. Studies have suggested that the V3, V4, or V3V4 regions are highly recommended for work employing 16S rRNA gene sequences as these provided adequate and accurate information for taxonomic classification of bacterial communities (Castelino et al. 2017; Cai et al. 2013). The V3V4 regions are also highly recommended by genome sequencing companies. However, it is confusing to find that similar type of samples sometimes use different primer sets (Huang et al. 2010; Tyx et al. 2016), while different types of samples use similar primers (Huang et al. 2010; Thomas and Sekhar 2017). For dealing with the problem of selection of preferred primer sets in the characterization of bacterial diversity of AFTs by 16S rRNA gene sequence, we employed four primer sets in this study. Conclusively, the choice of primer set could have a dramatic impact on the resulting accuracy of bacterial community structure determination. The results showed that the primers 799F–1193R had low co-amplification levels of chloroplast and mitochondrial genes, and also covered more genera than other primers. Moreover, the conclusion was also confirmed by the results of functional gene prediction. It is obvious that the abundance of functional genes related to photosynthesis was far lower using the V5V6V7 region than the other 16S regions. Therefore, we highly recommend that the preferred primers 799F–1193R are used to study the accurate diversity of



tobacco-related samples, and even other plant samples in future studies, and the plant samples are not only the above ground parts of plants, but also the subterranean parts, as suggested by several other reports (Beckers et al. 2016; Huang et al. 2010; Sun et al. 2008).

A large number of reports have highlighted the importance of microbial communities in degradation of harmful compounds, natural product biosynthesis, and other metabolism processes (Abia et al. 2018; Chen et al. 2016; Thelusmond et al. 2016; Wang et al. 2016; Yan et al. 2015). PICRUSt provided a convenient way to predict the functional composition of a metagenome using a database of reference genomes. In addition, it gave useful insights into the large number of uncultured microorganisms (Langille et al. 2013). In the present study, we employed PICRUSt to predict the functional gene contents and abundances among microbial communities on AFTs, and revealed several attractive genes relevant to (1) photosynthesis, (2) nitrogen metabolism, (3) biosynthesis of valuable metabolites, (4) degradation of toxic

compounds, and (5) energy metabolism. Genes relevant to photosynthesis are possessed by plants, algae, and cyanobacteria, and were investigated in this paper, leading to the recommendation of primers 799F–1193R as the preferred sequencing region in future tobacco-related samples. Moreover, in other reports, genes relevant to photosynthesis were candidate genes involved in cold responses, energy metabolism, and carbohydrate metabolism (Bryant and Frigaard 2006; Legrand et al. 2013). Genes relevant to nitrogen metabolism are pivotal in the tobacco-related samples. It is well known that nicotine and tobacco-specific nitrosamines, the major nitrogen compounds in tobacco plants, play a critical role in smoking addiction and are harmful to humans (Liu et al. 2015; Wei et al. 2014). PICRUSt offers a promising way to discover harmful nitrogen compound-degrading microorganisms for improving the quality of tobaccos, treating tobacco waste, and producing valuable intermediates of these compounds. Genes relevant to biosynthesis of valuable metabolites, especially aroma compounds, would

have a bright prospect when exploring microorganisms that produce natural products. Similarly, genes relevant to degradation of toxic compounds, which include polycyclic aromatic hydrocarbons and other pollutants, have the potential to deal with contaminants of AFTs. Genes relevant to energy metabolism, which is a process of generating energy for the whole the organism, have a key role in the above actions of bacteria on AFTs. This might justify the conclusion that AFTs are a treasury of microbes that could be widely used in all walks of life.

In this study, the performance of four primer pairs for determination of bacterial community structure on AFTs using 16S rRNA gene sequences were evaluated experimentally, and the corresponding functional genes were predicted using PICRUSt. Results suggested that the amplification region V5V6V7 with primers 799F–1193R generated a more accurate picture of the bacterial community structure on AFTs. First, primers 799F–1193R had low co-amplification levels of chloroplast and mitochondrial genes, which was concluded from the result of low abundance of *Cyanobacteria* at the phylum level and no *C. gigantea* or *N. tabacum* (common tobacco) at the genus level. Then, primers 799F–1193R covered more genera than the other primers. Finally, the PICRUSt results displayed that the abundance of photosynthesis-related genes was much lower in the amplification region V5V6V7. Therefore, we highly recommend that primers 799F–1193R should be used in future studies of plant samples, especially tobacco-related samples. In addition, functional gene prediction showed that the microbiome on AFTs was involved in a number of pathways. The study paid most attention to functional genes related to photosynthesis, nitrogen metabolism, biosynthesis of valuable metabolites, degradation of toxic compounds, and energy metabolism. A high abundance of functional genes involved in nitrogen metabolism was observed, reflecting a high level of bacteria on AFTs involved in degrading harmful nitrogen compounds and generating nitrogenous nutrients for others. Additionally, the functional genes involved in biosynthesis of valuable metabolites and degradation of toxic compounds provided information that the AFTs possess a huge library of microorganisms and genes that could be applied to further studies.

## Additional file

**Additional file 1.** Additional tables.

## Abbreviations

AFTs: aging flue-cured tobaccos; PICRUSt: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States; KEGG: Kyoto Encyclopedia

of Genes and Genomes; PCR: Polymerase Chain Reaction; IMG: Integrated Microbial Genomes.

## Authors' contributions

Planning and designing of the study: HBZ, FW and HBL; experimentation: WF, KCL and GZ; data analysis: FW YHX and AJL; manuscript drafting: FW and XM; supply material: JW and LJW. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> CAS Key Laboratory of Biobased Materials, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, China. <sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China. <sup>3</sup> Hainan Cigar Research Institute Hainan Provincial Branch of China National Tobacco Corporation, Haikou 571100, Hainan, China. <sup>4</sup> Tobacco Research Institute of Chinese Academy of Agriculture Sciences, Qingdao 266101, Shandong, China. <sup>5</sup> Yunnan Academy of Tobacco Sciences, Kunming 650106, China.

## Acknowledgements

This study was supported by Hainan's Key Project of Research and Development Plan (No. ZDYF2017155), Yunnan Academy of Tobacco Sciences Project (NO. 2017CP01), Youth Innovation Promotion Association CAS No. 2017252.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

The data supporting the conclusions of this article are all available.

## Consent for publication

All authors gave their consent for publication.

## Ethics approval and consent to participate

Not applicable.

## Funding

Hainan's Key Project of Research and Development Plan (No. ZDYF2017155); Yunnan Academy of Tobacco Sciences Project (No. 2017CP01); Youth Innovation Promotion Association CAS No. 2017252.

## Publisher's Note

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

Received: 26 March 2018 Accepted: 3 November 2018

Published online: 10 November 2018

## References

- Abia ALK, Alisoltani A, Keshri J, Ubomba-Jaswa E (2018) Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. *Sci Total Environ* 616–617:326–334. <https://doi.org/10.1016/j.scitotenv.2017.10.322>
- Beckers B, Beeck M, Thijs S, Truyens S, Weyens N, Boerjan W, Vangronsveld J (2016) Performance of 16s rDNA primer pairs in the study of rhizosphere and endosphere bacterial microbiomes in metabarcoding studies. *Front Microbiol* 7:1–15. <https://doi.org/10.3389/fmicb.2016.00650>
- Bryant DA, Frigaard NU (2006) Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol* 14(11):488–496. <https://doi.org/10.1016/j.tim.2006.09.001>
- Cai L, Ye L, Tong Y, Lok S, Zhang T (2013) Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLoS ONE* 8(1):e53649
- Castelino M, Eyre S, Moat J, Fox G, Martin P, Ho P, Upton M, Barton A (2017) Optimisation of methods for bacterial skin microbiome investigation: primer selection and comparison of the 454 versus MiSeq platform. *BMC Microbiol* 17(1):23. <https://doi.org/10.1186/s12866-017-0927-4>



- Chen L, Brookes PC, Xu J, Zhang J, Zhang C, Zhou X, Luo Y (2016) Structural and functional differentiation of the root-associated bacterial microbiomes of perennial ryegrass. *Soil Biol Biochem* 98:1–10. <https://doi.org/10.1016/j.soilbio.2016.04.004>
- Dyall SD, Brown MT, Johnson PJ (2004) Ancient invasions: from endosymbionts to organelles. *Science* 304(5668):253–257. <https://doi.org/10.1126/science.1094884>
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Gregory C, Justin K, Jesse S, Kyle B, Frederic DB, Elizabeth KC, Noah F, Antonio GP, Julia KG, Jeffrey IG, Gavin AH, Scott TK, Dan K, Jeremy EK, Ruth EL, Catherine AL, Daniel M, Brian DM, Meg P, Jens R, Joel RS, Peter JT, William AW, Jeremy W, Tanya Y, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335–336. <https://doi.org/10.1038/nmeth0510-335>
- Huang JW, Yang JK, Duan YQ, Gu W, Gong XW, 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(2):553–562. <https://doi.org/10.1007/s00253-010-2763-4>
- Jason AP, Aymé S, Omry K, Zhao J, Susannah GT, Jeffery LD, Edward SB, Ley RE (2013) Diversity and heritability of the maize rhizosphere microbiome under field conditions. *PNAS* 110(16):6548–6553
- Kaeberlein T, Lewis K, Epstein SS (2002) Isolating “Uncultivable” microorganisms in pure culture in a simulated natural environment. *Science* 296(5570):1127–1129
- Kanehisa G (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30
- Koo H, Hakim JA, Morrow CD, Eipers PG, Davila A, Andersen DT, Bej AK (2017a) Comparison of two bioinformatics tools used to characterize the microbial diversity and predictive functional attributes of microbial mats from lake obersee, antarctica. *J Microbiol Methods* 140:15–22. <https://doi.org/10.1016/j.mimet.2017.06.017>
- Koo H, Mojib N, Hakim JA, Hawes I, Tanabe Y, Andersen DT, Bej AK (2017b) Microbial communities and their predicted metabolic functions in growth laminae of a unique large conical mat from lake untersee, east antarctica. *Front Microbiol* 8:1347. <https://doi.org/10.3389/fmicb.2017.01347>
- 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(9):814–821. <https://doi.org/10.1038/nbt.2676>
- Legrand S, Marque G, Blassiau C, Bluteau A, Canoy AS, Fontaine V, Jaminon O, Bahrman N, Mautord J, Morin J, Petit A, Baranger A, Riviere N, Wilmer J, Delbreil B, Lejeune-Henaut I (2013) Combining gene expression and genetic analyses to identify candidate genes involved in cold responses in pea. *J Plant Physiol* 170(13):1148–1157. <https://doi.org/10.1016/j.jplph.2013.03.014>
- Liu JL, Ma GH, Chen T, Hou Y, Yang SH, Zhang KQ, Yang J (2015) Nicotine-degrading microorganisms and their potential applications. *Appl Microbiol Biotechnol* 99(9):3775–3785. <https://doi.org/10.1007/s00253-015-6525-1>
- Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC (2012) IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* 40:115–122. <https://doi.org/10.1093/nar/gkr1044>
- Munyaka PM, Eissa N, Bernstein CN, Khafipour E, Ghia JE (2015) Antepartum antibiotic treatment increases offspring susceptibility to experimental colitis: a role of the gut microbiota. *PLoS ONE* 10(11):e0142536. <https://doi.org/10.1371/journal.pone.0142536>
- Parulekar NN, Kolekar P, Jenkins A, Kleiven S, Utkilen H, Johansen A, Sawant S, Kulkarni-Kale U, Kale M, Saebø M (2017) Characterization of bacterial community associated with phytoplankton bloom in a eutrophic lake in South Norway using 16S rRNA gene amplicon sequence analysis. *PLoS ONE* 12(3):e0173408. <https://doi.org/10.1371/journal.pone.0173408>
- Prosdocimi EM, Novati S, Bruno R, Bandi C, Mulatto P, Giannico R, Casiraghi M, Ferri E (2013) Errors in ribosomal sequence datasets generated using PCR-coupled ‘panbacterial’ pyrosequencing, and the establishment of an improved approach. *Mol Cell Probes* 27(1):65–67. <https://doi.org/10.1016/j.mcp.2012.07.003>
- Raven PH (1970) A multiple origin for plastids and mitochondria. *Science* 169(3946):641–646
- Shaver JM, Oldenburg DJ, Bendich AJ (2006) Changes in chloroplast DNA during development in tobacco, *Medicago truncatula*, pea, and maize. *Planta* 224(1):72–82. <https://doi.org/10.1007/s00425-005-0195-7>
- Shinichi S, Luis PC, Samuel C, Jens RK, Karine L, Guillem S, Bardya D, Georg Z, Daniel RM, Adriana A, Francisco MC, Paul IC, Corinne C, Francesco DO, Stefan E, Isabel F, Josep MG, Lionel G, Falk H, Florian K, Cyrille L, Gipsi LM, Julie P, Bonnie TP, Marta RL, Hugo S, Sara VS, Céline D, Marc P, Sarah S, Stefanie KL, Chris B, Colomban V, Gabriel G, Nigel G, Pascal H, Daniele I, Olivier J, Fabrice N, Hiroyuki O, Stephane P, Sabrina S, Lars S, Matthew BS, Jean W, Patrick W, Eric K, Jeroen R, Silvia GA, Tara Oceans coordinators (2015) Structure and function of the global ocean microbiome. *Science* 348:6237
- Su C, Lei L, Duan Y, Zhang KQ, Yang J (2012) Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. *Appl Microbiol Biotechnol* 93(3):993–1003. <https://doi.org/10.1007/s00253-011-3800-7>
- Sun L, Qiu F, Zhang X, Dai X, Dong X, Song W (2008) Endophytic bacterial diversity in rice (*Oryza sativa* L.) roots estimated by 16S rDNA sequence analysis. *Microb Ecol* 55(3):415–424. <https://doi.org/10.1007/s00248-007-9287-1>
- Sun Z, Li G, Wang C, Jing Y, Zhu Y, Zhang S, Liu Y (2014) Community dynamics of prokaryotic and eukaryotic microbes in an estuary reservoir. *Sci Rep* 4:6966. <https://doi.org/10.1038/srep06966>
- Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M (2014) Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PLoS ONE* 9(8):e105592. <https://doi.org/10.1371/journal.pone.0105592>
- Thelusmond JR, Strathmann TJ, Cupples AM (2016) The identification of carbamazepine biodegrading phylotypes and phylotypes sensitive to carbamazepine exposure in two soil microbial communities. *Sci Total Environ* 571:1241–1252. <https://doi.org/10.1016/j.scitotenv.2016.07.154>
- Thomas P, Sekhar AC (2017) Cultivation versus molecular analysis of banana (*Musa* sp.) shoot-tip tissue reveals enormous diversity of normally uncultivable endophytic bacteria. *Microb Ecol* 73(4):885–899. <https://doi.org/10.1007/s00248-016-0877-7>
- Tuan NN, Chang YC, Yu CP, Huang SL (2014) Multiple approaches to characterize the microbial community in a thermophilic anaerobic digester running on swine manure: a case study. *Microbiol Res* 169:717–724. <https://doi.org/10.1016/j.micres.2014.02.003>
- 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(1):e0146939. <https://doi.org/10.1371/journal.pone.0146939>
- Wang K, Ye XS, Zhang HJ, Chen HP, Zhang DM, Liu L (2016) Regional variations in the diversity and predicted metabolic potential of benthic prokaryotes in coastal northern Zhejiang, East China Sea. *Sci Rep* 6:38709
- Wang F, Zhao HW, Xiang HY, Wu LJ, Men X, Qi C, Chen GQ, Zhang HB, Wang Y, Xian M (2018) Species diversity and functional prediction of surface bacterial communities on aging flue-cured tobaccos. *Curr Microbiol* 75:1306–1315. <https://doi.org/10.1007/s00284-018-1525-x>
- Wei XT, Deng XW, Cai DB, Ji ZX, Wang CJ, Yu J, Li JP, Chen SW (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(52):12701–12706. <https://doi.org/10.1021/jf504084z>
- Wu Z, Hao Z, Sun Y, Guo L, Huang L, Zeng Y, Wang Y, Yang L, Chen B (2016) Comparison on the structure and function of the rhizosphere microbial community between healthy and root-rot *Panax notoginseng*. *Appl Soil Ecol* 107:99–107. <https://doi.org/10.1016/j.apsoil.2016.05.017>
- Yan Q, Bi Y, Deng Y, He Z, Wu L, Van Nostrand JD, Shi Z, Li J, Wang X, Hu Z, Yu Y, Zhou J (2015) Impacts of the Three Gorges Dam on microbial structure and potential function. *Sci Rep* 5:8605. <https://doi.org/10.1038/srep08605>
- Zhao MQ, Wang BX, Li FX, Qiu LY, Li FF, Wang SM, Cui JK (2007) Analysis of bacterial communities on aging flue-cured tobacco leaves by 16S rDNA PCR-DGGE technology. *Appl Microbiol Biotechnol* 73(6):1435–1440. <https://doi.org/10.1007/s00253-006-0625-x>