Quantification of Microbial Species in Solid State Fermentation Samples Using Signature Genomic Sequences

Zhaohui Xu*, Pooja Yadav*, Zhizhou Zhang†‡, Sankardas Roy§, Huimin Zhang*†
*Department of Biological Sciences, Bowling Green State University, Ohio, USA
{zxu, pyadav, huiminz}@bgsu.edu
†School of Chemical Engineering & Technology, Harbin Institute of Technology, Harbin, China
zhangzzbiox@hitwh.edu.cn
‡Shandong Gredmedic Co. Ltd., Weihai, China
§Department of Computer Science, Bowling Green State University, Ohio, USA
sanroy@bgsu.edu

***Abstract*—Solid state fermentation processes are mediated by the collective metabolism of specialized microbial communities. Monitoring the relative abundance of dominating species is a critical task in quality control, which is traditionally done by wet lab techniques, such as quantitative PCR (qPCR). In this study, we developed a computational method to quantify microbial species in metagenomes based on their signature genomic sequences, *i.e.*, unique k-mers. Bacterial species found in fermentation starters of a Chinese liquor producer were used as examples to demonstrate the development and application of the method. A database was constructed, comprising 562 complete genome sequences of 93 bacterial species that had been found in relevant fermentation samples. K-mers in length of 12 were extracted from each species and compared against each other to identify the ones that were unique to each species. The quantity of a species was determined by the average frequencies of unique k-mers encountered in the metagenome. Six dominating bacterial species were chosen as reporter species to test the quantification method. Four metagenome datasets were simulated, which contained various portions of sequence reads generated from the genomes of the reporter species. The amount of reads sampled from each reporter species followed a pre-determined ratio, *i.e.*, a known relationship in relative abundance. For each simulated dataset, the cell number of each reporter species was computed based on the unique k-mers found in the metagenome. In all datasets, the computed quantities of the reporter species reflected the expected relative abundance by displaying a linear relationship with the pre-determined ratio. This demonstrates that quantification based on a set of unique k-mers is a reliable way to detect relative abundance among species. Besides industrial fermentation, this method may also be applied to areas such as wastewater treatment, microbiota analysis, *etc*.**

***Index Terms*—Relative abundance, K-mers, Signature sequences, Fermentation, Microbial community**

I. Introduction

Many traditional beverages are produced by solid state fermentation, in which naturally occurring microbial communities are used to inoculate the feedstock, and fermentation takes place in ambient environment without vigorous stirring. In these processes, the collective metabolic capacities of specialized microbial communities convert the feedstock into beverages of distinctive flavors. Due to the variation of environmental conditions, such as seasonal temperatures and humidity, groundwater pH and hardiness, as well as nutritional compositions of local feedstock, microbial communities found in different geographic fermentation stations vary, and consequently, flavor of the products. For example, Chinese liquor, *i.e.*, Baijiu, is a type of alcoholic beverage being produced by thousands plants across the country using traditional solid fermentation methods [1], [2]. Each plant is operated under an established protocol that is optimized for local environmental conditions, and each plant supplies liquor of distinctive flavors that are recognized by specific consumer populations. Great efforts have been taken in recent years to isolate and characterize the microbial species in these fermentation samples and have offered valuable insights about the composition and metabolism of these microbes [3]–[10].

On the other hand, because of its exposure to the ambient environment, solid fermentation is prone to contamination and drifts in species compositions. This inevitably compromises the quality of the beverages and results in significant economic losses. Cycles of solid fermentation normally take months or even years to complete, and early detection of abnormal microbial compositions could prompt managerial actions to mitigate losses. In our previous study, we developed a set of real-time PCR primers for quantification of microbial species in several starter and pit mud samples of a liquor company in China (Gujing Group, Anhui, China) [11]. The fermentation facilities have been in continuous operation for nearly two thousand years and are a major contributor to local economy. The primers are instrumental in product quality control, but the design and screening of the primers was time consuming and labor intensive. A different fermentation process will require monitoring a different set of microbial species [12], and thus the entire primer design and screening work will have to be repeated. Therefore, it is desirable to have a method that can be easily automated and adapted to another fermentation process.

This study aims to develop a computational method to quantify microbes at the species level in a metagenome based on unique k-mers of each target species, *i.e.*, a DNA string of length k that is shared by all strains of the target species but not any other species. The intended application of our method is to monitor the fermentation process by routinely sequencing the metagenomes of the fermentation samples at designated time points and calculating the relative abundance of a set of reporter species. Because the metabolic activities of each microbial species are affected by every other species coexisting in the same environment, an apparent deviation from the expected ratio among the reporter species could be an early sign of abnormal metabolism and demands immediate attention. In fermentation samples, microbial communities are often dominated by a dozen or so known species, even though their relative abundance changes dramatically throughout the production. These dominating species are ideal candidates for reporter species, and the quantitative relationship among them is a reliable indicator of the overall health of fermentation processes.

Our method focuses on estimation of the relative quantities of reporter species and requires prior knowledge of the taxonomic membership of the community. The taxon knowledge can be acquired by existing technologies, such as 16S rRNA profiling, using TA-cloning [11] or next-generation sequencing technologies (NGS) [13]. In principle, these methods may also be used to quantify microbial compositions in complex communities, but they suffer bias caused by many factors, including primer design and PCR amplification, which could severely distort the quantities of bacteria actually present in a sample [14]–[16]. Kraken [17] is highly effective in assigning taxonomic labels to metagenomics DNA sequences and is an excellent choice for the initial classification purpose to determine what microbes are there in a fermentation sample. However, for later monitoring purpose, using Kraken would be too computationally expensive to us because it tries to classify and quantify all species contained in a sample, instead of a small set of reporter species. This is a drawback shared by MetaPhlA [18]. In addition, MetaPhlAn considers coding sequences only and leaves out useful information contained in the intergenic regions. GASiC [19] emphasizes on metagenomics abundance estimation, but it relies upon aligning reads against reference genomes and thus is even more computationally demanding. MetaID [20] applies a scoring function to assign weights to each k-mer and focuses on identification and quantification at the strain level. In solid fermentation, it is unrealistic to quantify microbes down to the strain level because of the high complexity of the microbial community, whose taxon composition has yet to be resolved at the strain level in most cases. Therefore, in this study, we decided to use unique k-mers at the species level for quantitative analysis. We employed the fermentation starters from the same liquor company as before [11] to demonstrate the development and application of our method. The difference of the current paper from our prior work [11] is discussed in Section III-A.

II. Materials and Methods

A. Database construction

The steps involved in this study are illustrated in Fig. 1, starting from the database construction stage. Complete genome sequences of the species that have been identified in the Gujing samples [11] were downloaded from NCBI GenBank. They constituted the entire genome database of this study. Throughout the study, sequence collection and analyses were done with Python (https://www.python.org/), and statistical analyses were performed with R (https://www.r-project.org/).

B. Identification of unique k-mers at the species level

This was the pre-processing stage (Fig. 1). K-mers from each genome were tallied. K-mers shared by all strains of a species were extracted to represent the species, and the counts of these common k-mers were means of their occurrences across these strains. Species k-mers were defined as k-mers found in the complete genome of a species if the species had a single complete sequence, or the common k-mers of that species if the species had multiple complete sequences. K-mers found in the target species but not in any other species in our database are referred to as unique k-mers, which were obtained by comparing the k-mers of each species against those of others.

C. Simulation of metagenomes

Metagenomes used in this study were simulated by generating random reads from the genomes of reporter species up to a desired data volume, followed by sampling the rest of the genomes in the database to meet a total data volume. All sampled reads were in length of 150 bases, a typical size of current Illumina sequencing reads. A cellular ratio, instead of absolute cell numbers, was used to allocate data volume among reporter species so that results are comparable across samples and studies [21]. Consequently, species with bigger genomes received higher DNA percentage quotas than species with smaller genomes. Because current sequencing technologies can generate at least 5-10 Gb raw data for metagenome sequencing (Novogene), the total data volume for our simulated metagenomes was set at 6 Gb.

D. Quantification of reporter species

Metagenome datasets were constructed to mimic the bacterial composition of fermentation starters prepared at 54-60°C, which are also called medium-high temperature starters (MH_Daqu). Simulated metagenome reads were screened to identify unique k-mers belonging to any of the reporter species. The count of each unique k-mer was adjusted by a weight, which was the reciprocal of the average occurrence of the k-mer per genome; this is called the modified count of the k-mer. The abundance of a species, *i.e.*, the number of cells belonging to that species, was calculated as the mean of the modified counts of its unique k-mers encountered in a metagenome, as shown in Eq. (1), where $\hat{s}$ is the estimate of cell number (or genome copy number), m is the

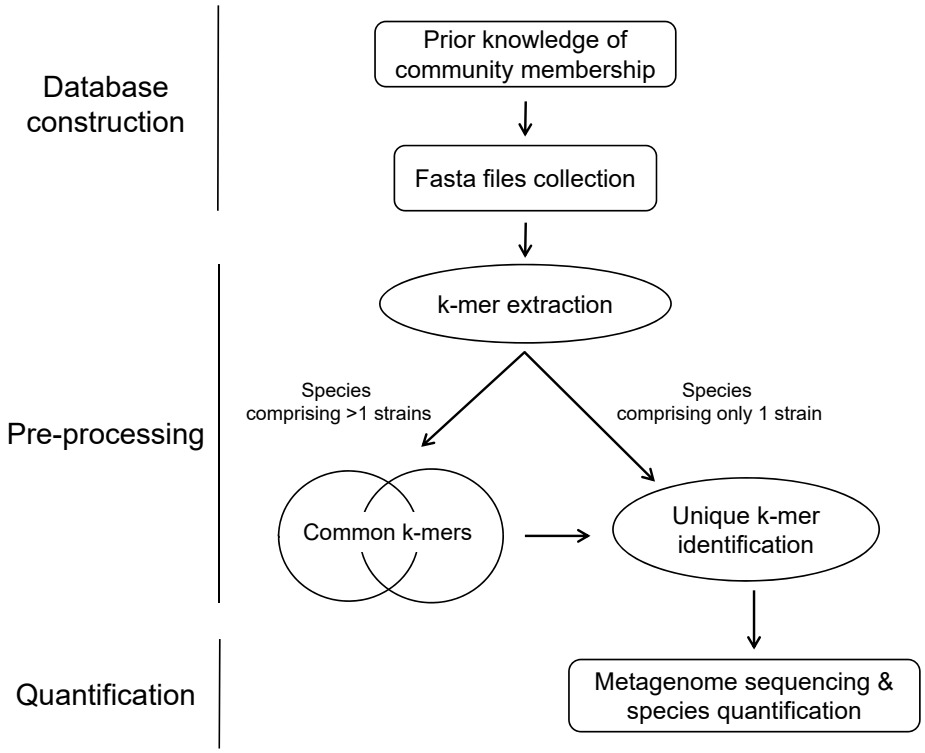

Fig. 1. Flow chart of the method. Our analytical method is composed of three stages: database construction, pre-processing, and quantification. In the first stage, based on prior knowledge on the composition of a microbial community, complete genome sequences of the relevant species are collected. In the second stage, k-mers are gathered from each genome, common k-mers are extracted from strains belonging to the same species, and then unique k-mers are identified for each species. In the last stage, metagenomes are sequenced, and the sequence reads are screened for unique k-mers linking with target species. The abundance of target species is computed based on the identified unique k-mers.

number of unique k-mers of the species that are found in the metagenome, c_i is the count of the ith k-mer, and w_i is the weight of the ith k-mer. For example, according to the results of the pre-processing stage, species X has five unique k-mers, a, b, c, d, and e, and their respective weights are 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, and $\frac{1}{5}$. In a simulated metagenome, the five k-mers were found to appear 0, 1, 6, 7, and 12 times. Then the estimated cell number of species X is $(1 * \frac{1}{2} + 6 * \frac{1}{3} + 7 * \frac{1}{4} + 12 * \frac{1}{5}) \div 4 \approx 2$. In calculating $\hat{s}$, to reduce the effects caused by outliers, up to 20% of the observations were trimmed from each end before the mean was computed (trim ≤ 0.2). The estimated cell numbers of all reporter species were then compared to the predetermined ratio that had been used to simulate the metagenome.

$$\hat{s} = \frac{\sum_1^m c_i w_i}{m} \tag{1}$$

III. Results

A. Construction of the genome database

Our previous studies identified a collection of bacteria in four fermentation samples from two plants of the Gujing Group [11]. These microbes represented the majority of microbes one expects to find in the facilities, whether the samples were from starters or pit mud or another fermentation stage. Only a subset of these microbes are expected to be detectable in a particular sample though, depending on where and when the sample will be taken. Our genome database included all of the available complete sequences of the identified species (last searched in April, 2017). Altogether, 562 genome sequences were collected, which represented 93 species. Among them, 54 species had genome sequence from a single strain, while 39 species had genomes from multiple strains, ranging from 2 to 167.

B. Determination of k-mer length

To accurately estimate the cell number of a species using Eq. (1), m must be big enough. Here we were aiming at the range of hundreds to thousands. In statistical terms, having a sample size in this range should provide a reliable estimation of the mean of a population. The average genome size in our database was about 3 million base pairs (Mbp). Since DNA molecules are double stranded, each genome contains up to 6 million k-mers. This is because the number of k-mers in a DNA string of n bases is $n - k + 1$ and, when k is far less than n ($k << n$), $n - k + 1$ approaches n. Also, DNA is composed of 4 nucleotides, and $4^{11} = 4194304$. This means any genome in our database could contain a complete set of all

possible 11-mers. Therefore, it is unrealistic to expect finding enough unique 11-mers for our quantification needs. We began testing at $k = 12$ and found that the number of unique k-mers we typically got was in hundreds or thousands (see below). When $k = 13$, the number of unique k-mers were in tens or hundreds of thousands. Although statistically speaking, a bigger sample size is preferred, processing tens or hundreds of thousands of k-mers would pose a daunting demand on computational resources. To balance the needs for statistical analysis and computational cost control, we decided to set the value of k to 12.

C. Identification of unique 12-mers at the species level

Because we intended to quantify microbes at the species level and many species had more than one complete genome sequences, we needed to identify a set of common k-mers that were in every strain to represent the species. The number of common 12-mers extracted for the 39 species ranged from 702027, in case of *Enterobacter hormaechei*, to 5247363, in case of *Bacillus anthracis*.

Unique 12-mers were identified for each species by comparing the 12-mers at the species level. For species having a single genome sequence, the k-mers from that single genome were used; for species having multiple genome sequences, their common k-mers were used. The number of unique 12-mers identified from the 93 species ranged from 0 for *Clostridium botulinum* and *E. hormaechei*, to 5108 for *Arthrobacter* sp. IHBB 11108 (as illustrated in Fig. 2).

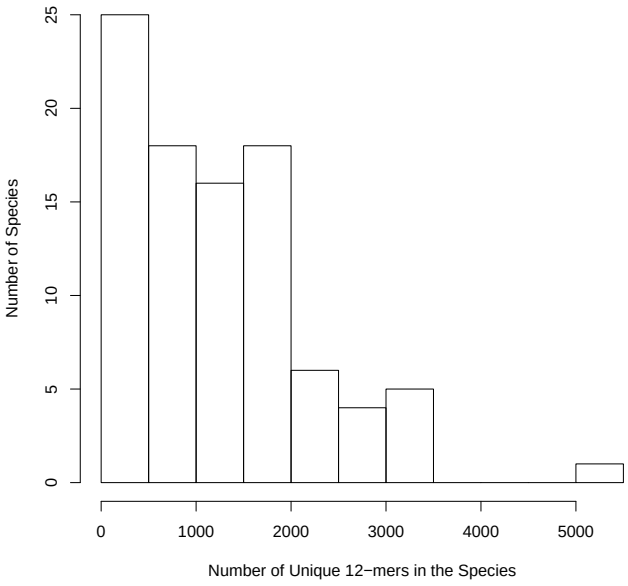

Fig. 2. Histogram of the distribution of unique 12-mers among the species. Unique 12-mers were identified at the species level for the 93 bacterial species previously known to exist in Gujing fermentation samples [11]. The horizontal axis represents the ranges of 12-mer counts found in the species, and the vertical axis counts the number of species belonging to each range. For example, the 3rd bin represents unique 12-mer counts from 1000 to 1500, and 16 species belong to this range.

D. Selection of reporter species

Among the dominating microbes found in the fermentation starters MH_Daqu, 13 bacterial species have been identified, namely, *B. licheniformis, B. subtilis, Virgibacillus halotolerans, Staphylococcus kloosii, Lactobacillus brevis, L. fermentum, L. plantarum, L. pontis, L. rossiae, E. hormaechei, Pantoea agglomerans, P. ananatis*, and *P. vagans*. The most populated one is *V. halotolerans*, which counts to about half of all identified species. Ten of the thirteen species had complete genome sequences available and were candidates for reporter species. Further analysis revealed that six of them had unique 12-mers close to or above a thousand (Table I). Therefore, these six species were selected as the reporter species for MH_Daqu samples. If one is interested in other samples than MH_Daqu, *e.g.*, pit mud samples, species dominating those samples should be considered as reporters.

E. Quantification of reporter species

Among the reporter species, their relative abundance was set at a ratio of 1:2:4:8:16:32 to test a wide range of values. The parts were randomly assigned to the species, rendering *P. vagans* to be 1 and *L. fermentum* to be 32 (Table I). (Since each species represents an independent observation, a different assignment of parts among the species should not affect the final results.) The ratio was then adjusted by the genome sizes of the species to compute DNA percentages in the simulated datasets (Table I). A total of four metagenome datasets were simulated with the entire (100%) or partial (25%, 50%, or 75%) dataset sampled from the reporter genomes (Table I).

With 20% of the observations were trimmed from both ends of the modified counts of unique k-mers (trim $= 0.2$), the calculated cell numbers of each species were plotted against the expected ratio. For all four datasets, the two variables fitted nicely to linear models, with r^2 above 0.99, even in the case of Comp_025 where only 25% of the data were from reporter genomes (as illustrated in Fig. 3). As a negative control, another mock metagenome containing 0% of reporter genomes but 100% background genomes was constructed and analyzed by following the identical procedure. When attempted to fit the data points to a linear model, we got $y = 17.5 - 0.31x$, $r^2 = 0.46$. These results suggest that Eq. (1) is a reliable estimate of cell numbers and a useful tool in depicting relative abundance among species.

F. Discussion

The concepts of common k-mers and k-mer weight in this study are different from what are defined in MetaID [20], which intends to identify and quantify prokaryotes at the strain level. In MetaID, common k-mers refer to k-mers found in more than 2 strains, regardless whether the strains belong to the same species, whereas in this study, common k-mers refer to k-mers that are shared by all strains of a species. In MetaID, weights are assigned to k-mers based on their presence across

TABLE I
COMPOSITIONAL DATA USED FOR METAGENOME SIMULATIONS

Species	No. of unique 12-mers	Ratio	Comp_025 (%)	Comp_050 (%)	Comp_070 (%)	Comp_100 (%)
B. licheniformis	1279	8	5.06	10.11	15.17	20.23
L. brevis	2166	16	6.15	12.30	18.44	24.59
L. fermentum	1956	32	9.30	18.60	27.89	37.18
L. plantarum	2225	2	0.96	1.93	2.89	3.86
P. ananatis	1217	4	2.83	5.66	8.49	11.32
P. vagans	986	1	0.70	1.41	2.11	2.82
All reporters			25	50	75	100
Background			75	50	25	0

genomes; the more ubiquitous the k-mer is, the lower the weight it receives. In this study, weights are assigned based on the presence of k-mers within a genome; the higher the count of a k-mer within the genome, the lower the weight for that k-mer.

The robustness of our method relies upon the large number of unique k-mers we use in estimating cell numbers for reporter species. It is worth noting that the unique k-mers identified by our method are relatively unique, even in the given database. That is because nearly half of our species have multiple genome sequences, and it is their common k-mers that are used to identify unique k-mers. The more genomes a species has, the more diverse the genomes are, the less the common k-mers, and hence the more unique k-mers we will find for that species. For example, if k-mer x only exists in *E. coli* strain A but not in strain B, when extracting common k-mers for *E. coli*, x will not be collected. When *E. coli* is compared to *B. subtilis* who does have x as a common k-mer, we will mistakenly think x is unique to *B. subtilis*. When the occurrence of x is used to calculate the cell numbers of *B. subtilis*, we will overestimate. This type of errors are inevitable and will increase the noise of our analysis. However, we do not depend upon just one k-mer to estimate the cell number of a species. Instead, we use hundreds to thousands k-mers, thus the noise created by a small portion of k-mers should not distort the overall estimation too much. Moreover, there are always microbes that do not have complete genome sequences available, even in well-studied industrial fermentation samples. This false uniqueness is an expected part of the analysis. In this sense, the unique k-mers here can be understood as signature k-mers. However, strictly speaking, uniqueness is always up to a certain level. So, for simplicity, we choose to use the term unique k-mers. On the other hand, because of the limitation on sequencing depth, some low frequent occurring species will be underestimated. To reduce the errors of various types, one can trim off the extreme numbers from both ends of a dataset to get a better estimation of the population mean. If the noise is too big to be handled by trimming the data, one can increase the value of k to enhance specificity.

The successful application of the method is dependent upon prior knowledge of the memberships of the microbial community, which is often the case for industrial processes, at least more and more so in recent years. The availability of genome sequences of the microbes is another requirement. Since industrial processes indubitably carry huge economical interest and the cost of DNA sequencing is constantly going downward, as we have witnessed in the past decade, it is guaranteed that more and more microbes will be sequenced in an accelerated pace, making precision quantification an achievable goal in the near future.

IV. CONCLUSIONS

Characterizing and monitoring the relative abundance of reporter species is an effective way to forecast the outcomes of fermentation. In this study, we developed a simple but robust method to quantify the relative abundance among a group of reporter species, based on their unique k-mers. The application of the method was demonstrated with fermentation starter samples from a Chinese liquor facility. The method should be applicable in similar industrial settings as well, such as kimchi fermentation [12] and wastewater treatment [22].

REFERENCES

[1] X.-W. Zheng and B.-Z. Han, "Baijiu, chinese liquor: History, classification and manufacture," *Journal of Ethnic Foods*, vol. 3, no. 1, pp. 19–25, 2016.

[2] G. Jin, Y. Zhu, and Y. Xu, "Mystery behind chinese liquor fermentation," pp. 18–28, 2017.

[3] M. Gou, H. Wang, H. Yuan, W. Zhang, Y. Tang, and K. Kida, "Characterization of the microbial community in three types of fermentation starters used for chinese liquor production," *Journal of the Institute of Brewing*, vol. 121, no. 4, pp. 620–627, 2015.

[4] C.-l. Wang, D.-j. Shi, and G.-l. Gong, "Microorganisms in daqu: a starter culture of chinese maotai-flavor liquor," *World Journal of Microbiology and Biotechnology*, vol. 24, no. 10, pp. 2183–2190, 2008.

[5] L. Xiu, G. Kunliang, and Z. Hongxun, "Determination of microbial diversity in daqu, a fermentation starter culture of maotai liquor, using nested pcr-denaturing gradient gel electrophoresis," *World Journal of Microbiology and Biotechnology*, vol. 28, no. 6, pp. 2375–2381, 2012.

[6] X.-W. Zheng, Z. Yan, B.-Z. Han, M. H. Zwietering, R. A. Samson, T. Boekhout, and M. J. Robert Nout, "Complex microbiota of a chinese fen liquor fermentation starter (fen-daqu), revealed by culture-dependent and culture-independent methods," pp. 293–300, 2012.

[7] R. Zhang, Q. Wu, and Y. Xu, "Aroma characteristics of moutai-flavour liquor produced with bacillus licheniformis by solid-state fermentation," *Letters in applied microbiology*, vol. 57, no. 1, pp. 11–18, 2013.

[8] L. Zhang, C. Wu, X. Ding, J. Zheng, and R. Zhou, "Characterisation of microbial communities in chinese liquor fermentation starters daqu using nested pcr-dgge," *World Journal of Microbiology & Biotechnology*, vol. 30, no. 12, pp. 3055–3063, Dec 2014.

[9] H. Wu, S. Zhang, Y. Ma, J. Zhou, H. Luo, and J. Yang, "Comparison of microbial communities in the fermentation starter used to brew xiaoqu liquor," *Journal of the Institute of Brewing*, vol. 123, no. 1, pp. 113–120, 2017.

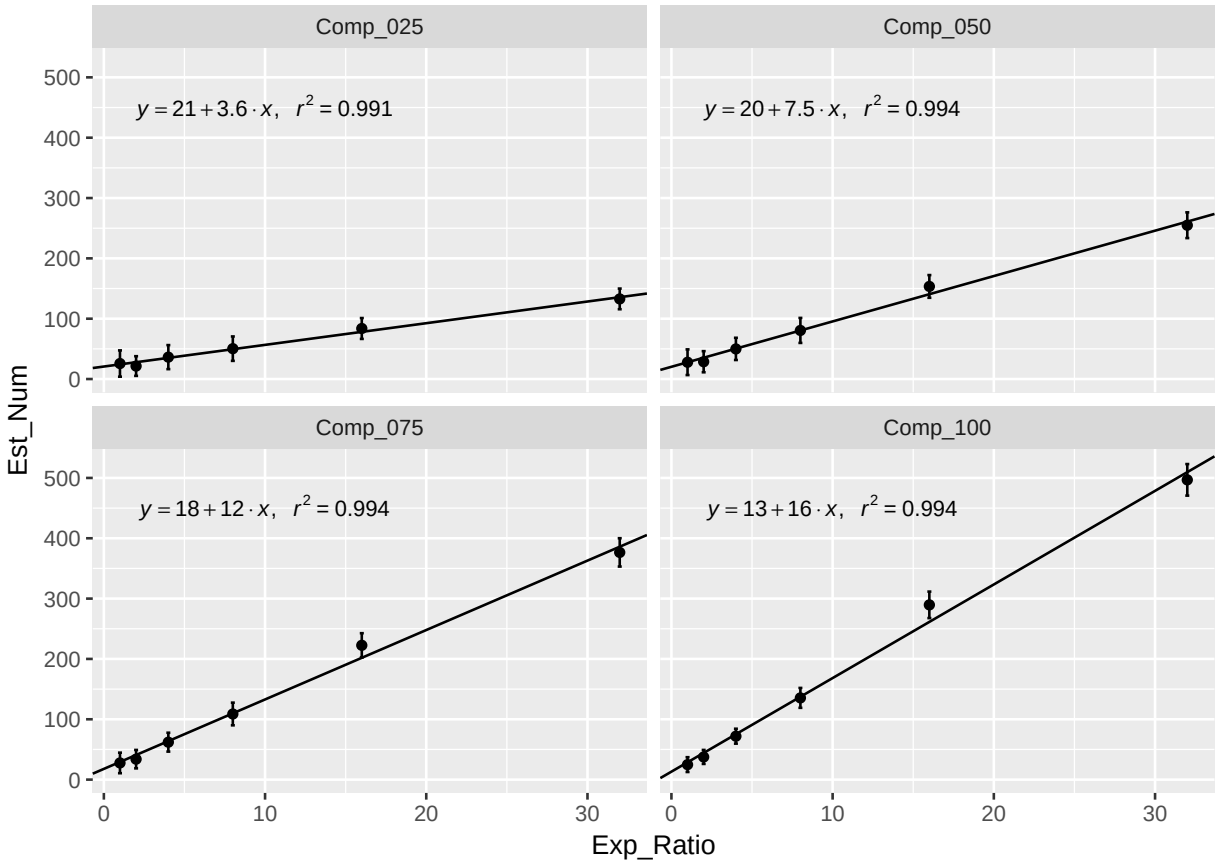

Fig. 3. Correlations of inferred and true relative abundance among six reporter species. Four metagenomes were simulated: Comp_025, _050, _075, and _100, which respectively had 25%, 50%, 75%, and 100% of the sequence reads generated from the six reporter species. The horizontal axis represents the expected ratio (Exp_Ratio) among the reporter species, *i.e.*, *P. vagans : L. plantarum : P. ananatis : B. licheniformis : L. brevis : L. fermentum* $= 1 : 2 : 4 : 8 : 16 : 32$. The vertical axis represents the estimated cell numbers (Est_Num) of the species, which were calculated according to Eq. (1). The error bars are standard deviations.

[10] X. Meng, Q. Wu, L. Wang, D. Wang, L. Chen, and Y. Xu, "Improving flavor metabolism of saccharomyces cerevisiae by mixed culture with bacillus licheniformis for chinese maotai-flavor liquor making," *Journal of industrial microbiology &5:13 PM 8/12/2017 biotechnology*, vol. 42, no. 12, pp. 1601–1608, 2015.

[11] H. Zhang, H. He, X. Yu, Z. Xu, and Z. Zhang, "Employment of near full-length ribosome gene ta-cloning and primer-blast to detect multiple species in a natural complex microbial community using species-specific primers designed with their genome sequences," *Molecular biotechnology*, vol. 58, no. 11, pp. 729–737, Nov 2016.

[12] G.-H. Ahn, J. Moon, S.-Y. Shin, W. Min, N. Han, and J.-H. Seo, "A competitive quantitative polymerase chain reaction method for characterizing the population dynamics during kimchi fermentation." *Journal of Industrial Microbiology and Biotechnology*, vol. 42, no. 1, p. 49, 2014.

[13] L. Bragg and G. W. Tyson, "Metagenomics using next-generation sequencing," *Methods in molecular biology*, vol. 1096, pp. 183–201, 2014.

[14] J. P. Brooks, D. J. Edwards, M. D. Harwich, M. C. Rivera, J. M. Fettweis, M. G. Serrano, R. A. Reris, N. U. Sheth, B. Huang, P. Girerd, V. M. Consortium, J. F. Strauss, K. K. Jefferson, and G. A. Buck, "The truth about metagenomics: quantifying and counteracting bias in 16s rrna studies," *BMC microbiology*, vol. 15, p. 6, March 21 2015.

[15] P. Schloss, "The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16s rrna gene-based studies," *PLoS Comput Biol*, vol. 6, no. 7, p. e1000844, 2010.

[16] P. Schloss, D. Gevers, and S. Westcott, "Reducing the effects of pcr amplification and sequencing artifacts on 16s rrna-based studies," *PLoS ONE*, vol. 6, no. 12, p. e27310, 2011.

[17] D. E. Wood and S. L. Salzberg, "Kraken: ultrafast metagenomic sequence classification using exact alignments," *Genome biology*, vol. 15, no. 3, p. r46, March 03 2014.

[18] N. Segata, L. Waldron, A. Ballarini, V. Narasimhan, O. Jousson, and C. Huttenhower, "Metagenomic microbial community profiling using unique clade-specific marker genes," *Nature methods*, vol. 9, no. 8, pp. 811–814, June 10 2012.

[19] M. S. Lindner and B. Y. Renard, "Metagenomic abundance estimation and diagnostic testing on species level," *Nucleic acids research*, vol. 41, no. 1, p. e10, January 07 2013.

[20] S. M. Srinivasan and C. Guda, "Metaid: a novel method for identification and quantification of metagenomic samples," *BMC genomics*, vol. 14 Suppl 8, p. S4. Epub 2013 Dec 9, 2013.

[21] S. Nayfach and K. S. Pollard, "Toward accurate and quantitative comparative metagenomics," *Cell*, vol. 166, no. 5, pp. 1103–1116, August 25 2016.

[22] J. Tang, Y. Bu, X.-X. Zhang, K. Huang, X. He, L. Ye, Z. Shan, and H. Ren, "Metagenomic analysis of bacterial community composition and antibiotic resistance genes in a wastewater treatment plant and its receiving surface water," pp. 260–269, 2016.