ORIGINAL ARTICLE

Qualitative and Quantitative Analysis for Microbiome Data Matching between Objects

Hee Sang You¹,², Yeon Jeong Ok¹, Song Hee Lee¹,², So Lip Lee¹,², Young Ju Lee¹, Min Ho Lee²,³, Sung Hee Hyun¹,²

¹Department of Biomedical Laboratory Science, School of Medicine, Eulji University, Dajeon, Korea
²Department of Senior Healthcare, BK21 Plus Program, Graduate School, Eulji University, Daejeon, Korea
³Department of Food Science and Service, College of Bio-Convergence, Eulji University, Seongnam, Korea

ARTICLE INFO

Received May 4, 2020
Revised 1st June 26, 2020
Revised 2nd June 28, 2020
Revised 3rd July 13, 2020
Accepted July 14, 2020

ABSTRACT

Although technological advances have allowed the efficient collection of large amounts of microbiome data for microbiological studies, proper analysis tools for such big data are still lacking. Additionally, analyses of microbial communities using poor databases can lead to misleading results. Hence, this study aimed to design an appropriate method for the analysis of big microbial databases. Bacteria were collected from the fingertips and personal belongings (mobile phones and laptop keyboards) of individuals. The genomic DNA was extracted from these bacteria and subjected to next-generation sequencing by targeting the 16S rRNA gene. The accuracy of the bacterial matching percentage between the fingertips and personal belongings was verified using a formula and an environment-related and human-related database. To design appropriate analysis, the bacterial matching accuracy was calculated based on the following three categories: comparison between qualitative and quantitative analysis, comparisons within same-gender participants as well as all participants regardless of gender, and comparison between the use of a human-related bacterial database (hDB) and environment-related bacterial database (eDB). The results showed that qualitative analysis, comparisons within same-gender participants, and the use of hDB provided relatively accurate results. This study provides an analytical method to obtain accurate results when conducting studies involving big microbiological data using human-derived microorganisms.

INTRODUCTION

In recent years, researchers in various fields have been sampling microorganisms for microbiome study. Microorganisms are of two types-cultivable and non-cultivable. Next generation sequencing (NGS) has resulted in a big database of fastidious or non-cultivable organisms through culture-independent profiling [1]. These databases are useful in several fields such as medical, microbial ecology, and metagenomics [2, 3]. Particularly, metagenomics is a fast approach that provides genetic information by extracting and analyzing...
genomic DNA (gDNA) directly from the microbiome in the environment [4]. Although the technology to acquire information on the microbiome has improved, the proper analysis of such large data is still ambiguous. For example, the taxonomy and nomenclature of intestinal microflora are poorly defined within the database of 16S rRNA gene [5], and analysis of microbial communities using such poor taxonomic reference frameworks and tools can lead to erroneous results. Moreover, the phylogenetic profiling method that is applied for the analysis of the microbial community is different depending on the sequencing length and the ngs analysis method [6]. Therefore, the accuracy may depend on the confidence in the results of the ngs analysis. Meanwhile, to increase the accuracy of the results for personal identification by ngs analysis, personal identification using phylogenetic distance and personal identification using microbial strain composition were compared, and it was confirmed that the identification using microbial strain composition was superior [7]. Hence, there is a need to develop efficient analysis methods for large microbial databases.

Some previous studies have carried out microbial big data analysis to compare the similarity in microbial composition between an individual’s body parts and personal belongings [8-10], as an individual continually sheds his or her microbes [11-13], and approximately 30 million bacterial cells per hour [12], leaving a “microbial fingerprint” that has been found to be stable over time [14, 15]. In an attempt to design an appropriate analysis method for big microbial dataset, we harvested bacteria from individuals’ body parts (fingertips) and personal belongings (mobile phone and laptop keyboards), the same as previous studies. The matching ratio of bacteria between an individual’s fingertips and personal belongings was calculated, and the accuracy of matching was compared according to the following three categories: 1) comparison between qualitative and quantitative analysis, 2) comparisons within same-gender participants, and comparison of all participants regardless of gender, and 3) comparison between using a human-related bacterial database (hDB) and using an environment-related bacterial database (eDB). The bacterial matching rate and accuracy of bacterial matching was calculated using the formulas from the numerical results obtained through ngs. All bacteria in hDB and eDB were classified into the respective databases based on published literature. As this analysis uses specific unique bacteria in the sample rather than the general simple similarity analysis by ngs, it can be more reliable for personal identification analysis and can aid in high-level individual differentiation.

MATERIALS AND METHODS

1. Sample collection and gDNA extraction

Four men and four women in their early 20s, who attended the same university were selected as participants in this study. Samples were collected thrice at 2 day intervals between 5~6 pm from each participant’s ten fingertips and personal belongings (mobile phone and laptop keyboards). During the 8 h before sampling, there were no external factors such as washing at all sample sites. The participants’ gender and characteristics were surveyed through a questionnaire (Table S1). Sampling was performed using cotton swabs applied with sterile phosphate-buffered saline solution. The cotton of the swab was cut and stored at −70°C until DNA extraction. The gDNA from the swabs was extracted using the DNeasy PowerSoil kit (Mo Bio Laboratories, Inc., Carlsbad, CA, USA), according to the manufacturer’s protocols. Briefly, the C1 solution from the kit was added to the swab samples with bead tube and incubated at 60°C for 20 min. The bead tube containing the swab sample was physically vortexed for 20 min to ensure lysis of the microbial cell walls and enable an enough DNA extraction. Then, the swab was removed after centrifugation at 13,000 rpm for 1 min. Subsequently, the procedure was performed as per the instructions of the kit, and finally eluted in 80 μL and stored at −70°C until library preparation. All participants
provided informed consent and the study was approved by the Eulji University Institutional Review Board [EUIRB2017-18].

2. NGS (next generation sequencing)

The gDNA concentration was measured using PicoGreen (Invitrogen, Grand Island, NY, USA). The V3–V4 regions of 16S rRNA gene, which enables the identification of many types of microorganisms [16-18], were amplified by primary PCR using universal primers. The primers, included with the Illumina flow cell adapter, are as follows: 341F (5′–TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN GGC WGC A–3′) and 805R (5′–GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA CHV GGG TAT CTA ATC C–3′). The PCR was performed with 2.5 μL DNA sample (5 ng/μL), 5 μL forward primer, 5 μL reverse primer, and 12 μL 2X KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, USA) in a total volume of 25 μL. The reaction conditions comprised an initial denaturation at 95°C for 3 min, followed by 25 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s, and a final extension at 72°C for 5 min. Primary PCR products were amplified by secondary PCR for library preparation. The primer sequences for secondary PCR, included with the Illumina® flow cell adapter, are as follows: a Nextera Index PCR primer (Illumina®, USA) pair (forward: 5′ – AAT GAT ACG GCG ACC ACC GAG ATC TAC AC – [i5] – TCG TCG GCA GCG TC –3’ and reverse: 5′ – CAA GCA GAA GAC GGC ATA CGA GAT – [i7] – GTC TCG TGG GCT CGG – 3’). The PCR consisted of 5 μL sample DNA, 5 μL each of Nextera XT Index primers 1 and 2, 25 μL 2X KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, USA), and 10 μL PCR Grade Water. The reaction conditions comprised an initial denaturation at 95°C for 3 min, followed by 8 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s, and a final extension at 72°C for 5 min. The PCR products (libraries) were checked for quality using TapeStation DNA ScreenTape D1000 (Agilent, USA) and PicoGreen assay. The libraries were sequenced using the Illumina MiSeq platform (300 cycles ×2), according to the manufacturer’s instructions [19]. DNA extracted from the samples was sent to a commercial sequencing facility (Macrogen, Seoul, Korea) for ngs. Library concentration and size QC were as per the standards of the commercial sequencing facility (Data not shown).

3. Operational taxonomic unit (OTU) analysis and taxonomic assignment

After the sequencing was completed, Miseq raw data was extracted as a FASTQ file using MSR (MiSeq Reporter) and the PhiX sequence was removed through BWA (Burrows-Wheeler Alignment Tool). For paired end data for each sample, FLASH (1.2.11) was used to overlay 120~160 bp and select only the high-quality sequences of length 440~460 bp. The obtained sequences were first subjected to CD-HIT-OTU to remove low-quality sequences, ambiguous sequences, and chimera sequences, which are considered as sequencing errors, and then clustered with sequence similarity of 97% or more to form species-level OTUs. The representative sequence of each OTU was subjected to BLASTN (v2.4.0) against the Reference DB (type strain DB registered in StrainInfo among 16S rDNA of EBI), and taxonomic assignment was carried out based on the highest similarity. At this time, if the query coverage of the best hit matching the DB was less than 85% and the identity of the matched area was less than 85%, then taxonomy was not defined. Using Qiime (v1.8) with the OTU information, various microbial community comparison analysis was performed. To confirm the species diversity and uniformity of the microbial community in the environmental sample, Shannon Index and Inversed Simpson Index were obtained, and alpha diversity information was confirmed through Rarefaction curve and Chao1 value. Based on the weighted UniFrac distance, Beta diversity between samples (diversity information between samples in a comparison group) was obtained, and the
Table 1. Qualitative and quantitative comparison of the bacterial matching between fingertips and items (%)

(A) Subjects * items M1 FT M2 FT M3 FT M4 FT F1 FT F2 FT F3 FT F4 FT
M1 MP 38.5 5.4 0 0 2.9 1.5 2.4 1.0
M2 MP 9.1 0 0 0 0 0 0 0
M3 MP 0 0 0 0 0 0 0 0
M4 MP 0 2.4 4.8 4.8 0 2.4 0
F1 MP 3.2 0 0 3.2 3.2 0
F2 MP 4.2 0.6 0 1.2 1.8 0.6 0.6 0
F3 MP 5 0 0 0 1.7 1.7 0
F4 MP 15.2 4.5 0 0 1.5 0
M1 LT 5.8 1.4 0 0 2.9 1.4 2.9 0
M2 LT 2.8 8.3 0 0 0 0 13.9 0
M3 LT 1.6 1.6 0 0 0 1.6 0
M4 LT 3.8 3.8 0 3.8 0 7.5 0
F1 LT 3.3 3.3 3.3 0 4.9 4.9 3.3 0
F2 LT 0.9 1.9 0 0.9 3.8 1.9 1.9 0
F3 LT 1.7 6.9 0 0 8.6 3.4
F4 LT 0 0 0 6.3 0

(B) Subjects * items M1 FT M2 FT M3 FT M4 FT F1 FT F2 FT F3 FT F4 FT
M1 MP 16.7 1.7 0 0 1.2 0.6 0.2 0.8
M2 MP 0 0.4 0 0 0 0 0 0
M3 MP 0 0 0 0 0 0 0 0
M4 MP 0 0.6 1.0 0.7 0 0.02 0
F1 MP 0 0.4 0 0 0.6 0.4 0
F2 MP 1.5 0.1 0 0.2 0.8 0.2 0.1 0
F3 MP 1.1 0 0 0 0.1 0.1 0
F4 MP 5 0.5 0 0 0 0 0.1 0
M1 LT 0.4 0.1 0 0 0.1 0.04 0.1 0
M2 LT 0.1 0.8 0 0 0.1 0.4 0
M3 LT 0.04 0.02 0.2 0 0 0.03 0
M4 LT 0.7 0.3 0 0.9 0.1 0.8 0
F1 LT 0.1 0.1 0.8 0 1.6 0.1 0.1 0
F2 LT 0.4 0.2 0 0.03 0.7 0.5 0.1 0
F3 LT 0.01 0.3 0 0 0.2 0.1 0
F4 LT 0 0 0 0.8 0 0

(C) Subjects * items M1 FT M2 FT M3 FT M4 FT F1 FT F2 FT F3 FT F4 FT
M1 MP 6.3 5.4 1.0 2.4 1.0 1.0 1.5 3.4
M2 MP 0 9.1 9.1 4.6 0 0 4.6 0
M3 MP 10.0 0 0 0 0 10.0 0 0
M4 MP 0 2.4 2.4 7.1 4.8 4.8 0 2.4
F1 MP 0 6.5 3.2 0 3.2 0 3.2 0
F2 MP 2.4 5.4 0 12.1 3.6 1.8 1.8 3.0
F3 MP 5.0 3.3 6.7 0 5.0 1.7 1.7 1.7
F4 MP 15 4.6 1.5 1.5 0 0 0 4.6
M1 LT 10.1 2.9 1.5 8.7 7.3 2.9 2.9 2.9
M2 LT 0 5.6 2.8 0 2.8 0 2.8 0
M3 LT 3.1 4.7 1.6 1.6 1.6 1.6 0 1.6
M4 LT 1.9 5.7 1.9 5.7 7.6 0 0 1.9
F1 LT 3.3 4.9 1.6 6.6 4.9 0 1.6 0
F2 LT 7.6 5.7 0 9.4 5.7 0 0.9 2.8
F3 LT 0 3.5 0 5.2 9.8 0 3.5 1.7
F4 LT 0 12.5 0 0 6.3 0 6.3 0
Table 1. Continued

| Subjects * items | M1 FT | M2 FT | M3 FT | M4 FT | F1 FT | F2 FT | F3 FT | F4 FT |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| M1 MP            | 4.0   | 2.4   | 0.02  | 0.1   | 0.02  | 0.1   | 0.3   | 3.9   |
| M2 MP            | 0     | 0.4   | 0.01  | 0.01  | 0     | 0     | 0.1   | 0     |
| M3 MP            | 0.02  | 0     | 0     | 0     | 0     | 0.004 | 0     | 0     |
| M4 MP            | 0     | 0.6   | 0.6   | 1.2   | 0.4   | 0.7   | 0     | 0.1   |
| F1 MP            | 0     | 0.4   | 0.2   | 0     | 0.5   | 0     | 0     | 0.8   |
| F2 MP            | 0.2   | 1.5   | 0     | 2.3   | 4.0   | 0.5   | 0.1   | 0.2   |
| F3 MP            | 1.3   | 0.2   | 0.4   | 0     | 0.5   | 0.4   | 0.1   | 0.9   |
| F4 MP            | 0.3   | 0.5   | 0.5   | 0.1   | 0     | 0     | 0     | 0.7   |
| M1 LT            | 1.1   | 2.0   | 0.01  | 1.2   | 2.2   | 0.2   | 0.3   | 0.6   |
| M2 LT            | 0     | 0.2   | 0.1   | 0     | 0.04  | 0     | 0     | 0.004 |
| M3 LT            | 0.3   | 0.2   | 0.4   | 0.01  | 0.01  | 0.01  | 0     | 0.2   |
| M4 LT            | 0.7   | 0.1   | 0.03  | 0.04  | 0.04  | 0     | 0     | 0.2   |
| F1 LT            | 0.1   | 0.7   | 0.7   | 0.1   | 0.5   | 0     | 0.008 | 0     |
| F2 LT            | 1.3   | 1.3   | 0     | 3.3   | 0.9   | 0.11  | 0.2   |       |
| F3 LT            | 0     | 0.3   | 0     | 0.2   | 0.4   | 0.2   | 0.01  |       |
| F4 LT            | 0     | 1.0   | 0     | 0     | 0.5   | 0     | 0.5   |       |

(A) Qualitative analysis of the degree of bacterial matching between a fingertips and items of subjects based on the hDB. (B) Quantitative analysis of the degree of bacterial matching between fingertips and items of the subjects based on the hDB. (C) Qualitative analysis of the degree of bacterial matching between fingertips and items of the subject's based on the eDB. (D) Quantitative analysis of the degree of bacterial matching between fingertips and items of the subject's based on the eDB. The higher the number in the cell, the higher the degree of bacteria that matches with the item and fingertip.

Abbreviations: M, Male; F, Female; FT, Fingertip; MP, Mobile phone; LT, Laptop.

The flexible relationship between samples was visualized through principal coordinate analysis (PCoA) and UPGMA tree. The Canonical Correspondence Analysis to confirm the correlation with the sample’s environmental variables was analyzed at the species, genus, and door level using R (v3.1.2). OTU, alpha diversity, and beta diversity (PCoA, UPGMA tree) analyses results are attached as supplementary data (Supplementary Figure 1, Table S2∼S4).

4. Bacterial matching analysis using Venn diagram and simple formula

After taxonomic assignment, the bacterial matching between participants' fingertips and personal belongings was established using Venn diagrams, which were prepared using webtools (http://bioinformatics.psb.ugent.be/webtools/Venn/). Bacterial matching was compared according to the following categories: 1) comparisons between qualitative and quantitative analysis, 2) comparisons within the same-gender participants and comparison of all participants regardless of gender, and 3) comparison between analysis using hDB (Table S5) and eDB (Table S6). For qualitative and quantitative analysis, the ratio of bacterial matching was calculated using simple formulas. Qualitative analysis was based on the ratio of 'unique bacteria' species matched between the fingertips and personal belongings of the participants as follows:

Number of matching 'unique bacteria' species between fingertips and personal belongings \times 100 (%) 

The total number of 'unique bacteria' species within personal belongings

Quantitative analysis was defined as the total percentage (%) of 'unique bacteria' matched between the fingertips and the personal belongings of participants. Quantitative analysis was based on the ratio of 'unique bacteria' species matched between the fingertips and personal belongings of the participants as follows:
Table 2. Comparison of bacterial matching between fingertips and items within subjects who have same gender

(A)

|   | M1 FT | M2 FT | M3 FT | M4 FT | F1 FT | F2 FT | F3 FT | F4 FT |
|---|-------|-------|-------|-------|-------|-------|-------|-------|
| MPa | M1s' MP | M2 s' MP | M4 s' MP | M4 s' MP | F2 s' MP | F1 s' MP | F1 s' MP | F4 s' MP |
| LTa | M1s' LT | M2 s' LT | M3 s' LT | M4 s' LT | F1 s' LT | F1 s' LT | F3 s' LT | F3 s' LT |
| MPb | M1s' MP | M1 s' MP | M4 s' MP | M4 s' MP | F2 s' MP | F1 s' MP | F1 s' MP | F4 s' MP |
| LTb | M4s' LT | M2 s' LT | M3 s' LT | M4 s' LT | F1 s' LT | F2 s' LT | F3 s' LT | F3 s' LT |

(B)

|   | M1 FT | M2 FT | M3 FT | M4 FT | F1 FT | F2 FT | F3 FT | F4 FT |
|---|-------|-------|-------|-------|-------|-------|-------|-------|
| MPa | M3 s' MP | M2 s' MP | M2 s' MP | M4 s' MP | F3 s' MP | F2 s' MP | F2 s' MP | F4 s' MP |
| LTa | M1 s' LT | M4 s' LT | M2 s' LT | M1 s' LT | F3 s' LT | None | F4 s' LT | F2 s' LT |
| MPb | M1 s' MP | M1 s' MP | M4 s' MP | M4 s' MP | F2 s' MP | F2 s' MP | F2F3 s' MP | F3 s' MP |
| LTb | M1 s' LT | M1 s' LT | M3 s' LT | M1 s' LT | F2 s' LT | None | F4 s' LT | F2 s' LT |

The table is showed the items having the highest bacterial matching with subjects’ fingertips based on Table S1. The degree of bacterial matching of items and fingertips is compared within same gender. (A) presents the results of analysis based on a hDB. (B) presents the results of analysis based on eDB. A small letter ‘a’ is qualitative analysis results, and a small letter ‘b’ is quantitative analysis results. Abbreviations: See Table 1.

Results

First, we confirmed the accuracy of bacterial matching based on hDB (Figure 1). Quantitative and qualitative formulas were compared and analyzed for fingers, mobile phones, and laptops, including hDB and eDB. The hDB had higher similarity to the owners of fingers and objects than eDB, and there was no significant difference in qualitative and quantitative comparison (Table 1). The accuracy of the qualitative and quantitative analyses was 62.50% and 56.25%, respectively, when bacterial matching was carried out between the fingertips and the belongings of participants of the same gender (Table 2). The accuracy of the qualitative and quantitative analyses was 43.75% and 37.50%, respectively, when the bacterial matching was calculated regardless of the gender (Table 3). Next,
Table 3. Comparison of bacterial matching between fingertips and items within all subjects regardless of gender

(A)

|     | M1 FT | M2 FT | M3 FT | M4 FT | F1 FT | F2 FT | F3 FT | F4 FT |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|
| MPa | M1s' MP | M2s' MP | M4s' MP | M4s' MP | F1s' MP | F1s' MP | F2s' MP | F4s' MP |
| LTa | M1s' LT | M1s' LT | M2s' LT | M2s' LT | F2s' LT | M1s' LT | M2s' LT | M1s' LT |
| MPb | M1s' MP | M1s' MP | M4s' MP | M4s' MP | F1s' MP | F1s' MP | F2s' MP | F4s' MP |
| LTb | M4s' LT | M2s' LT | F1s' LT | F2s' LT | M2s' LT | M4s' LT | F1s' LT | M1s' LT |

(B)

|     | M1 FT | M2 FT | M3 FT | M4 FT | F1 FT | F2 FT | F3 FT | F4 FT |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|
| MPa | M3s' MP | M2s' MP | M2s' MP | F2s' MP | F3s' MP | M3s' MP | M2s' MP | F4s' MP |
| LTa | M1s' LT | F4s' LT | M2s' LT | F2s' LT | F3s' LT | M1s' LT | F4s' LT | M1s' LT |
| MPb | M1s' MP | M1s' MP | M4s' MP | F2s' MP | M2s' MP | M4s' MP | M1s' MP | M4s' MP |
| LTb | F2s' LT | M1s' LT | F1s' LT | F2s' LT | M2s' LT | F1s' LT | M1s' LT | F4s' LT |

The table showed the items having the highest bacterial matching with subjects’ fingertips based on Table S1. The degree of bacterial matching of items and fingertips is compared within all subjects regardless of gender. (A) presents the results of analysis based on hDB. (B) presents the results of analysis based on eDB. A small letter ‘a’ is qualitative analysis results, and a small letter ‘b’ is quantitative analysis results.

Abbreviations: See Table 1.

The accuracy of bacterial matching was confirmed based on eDB (Figure 2). The accuracy of qualitative and quantitative analyses was 31.25% and 37.50%, respectively, when bacterial matching was carried out between the fingertips and the belongings of the participants of same gender (Table 2), indicating that the results of both analyses were of similar accuracy. In contrast, the accuracy of the qualitative and quantitative analyses was 18.75% and 6.25%, respectively, when the participant’s gender was not considered (Table 3), revealing the low accuracy (<10%) of the quantitative analysis.

Together, these results indicate that the accuracy of qualitative analysis was higher than that of the quantitative analysis. Secondly, the accuracy of bacterial matching using hDB was higher than that of eDB. Finally, the accuracy of bacterial matching within the same gender was higher as compared to when gender was not considered. The qualitative and quantitative ratios of bacterial matching according to the method of analysis is shown in Table 1 and detailed information on the bacteria on the fingertips and personal belongings are provided in Tables S2, S3, and S4. Tables S5 and S6 provide further information on hDB and eDB.

**DISCUSSION**

To obtain reliable results from microbial big data, appropriate analysis is required. However, the definitions of databases regarding bacterial classification and nomenclature used for analysis are weak compared with the big data acquisition technology that is currently being developed [2]. Using poor or imperceptive database can lead to inaccurate results. Most studies on human skin have sampled the skin of the subject and then by ngs and using one or two reference gene data, have derived results of cluster analysis and similarity analysis [20]. This is possible due to the fact that, in general, healthy adults who live in daily life maintain a stable skin microbial community for up to 2 years despite persistent environmental changes [21]. However, this applies only to healthy adults, and not all humans. The skin microbial community can change with respect to diet, medication, and physical condition, and more than 20% of microbial diversity can be identified by these conditions [22]. In addition, genetic factors also influence the microbial diversity. Hence, skin microbial community studies require additional analytical methods [23]. Therefore, this study aimed to design an appropriate and accurate analysis method by
Figure 1. The bacterial matching of fingertips and items based on hDB. The Venn diagram represents number of bacterial species matching between fingertips and personal belongings. The bacterial matching is analyzed based on hDB. (A) shows the number of matched bacteria of the mobile phones of all participants and the fingertips of male participants. (B) shows the number of matched bacteria of the mobile phones of all participants and the fingertips of female participants. (C) shows the number of matched bacteria of the laptop keyboards of all participants and the fingertips of male participants. (D) shows the number of matched bacteria of the laptop keyboards of all participants and the fingertips of female participants.

Abbreviations: hDB, human-related bacterial database; M1∼M4, male participants; F1∼F4, female participants; FT, fingertips; MP, mobile phone; LT, laptop keyboards.
Figure 2. The bacterial matching of fingertips and items based on eDB. The Venn diagram represents the number of bacterial species matching between fingertips and personal belongings. This data is drawn based on eDB. (A) shows the number of matched bacteria of the mobile phones of all participants and the fingertips of male participants. (B) shows the number of matched bacteria of the mobile phones of all participants and the fingertips of female participants. (C) shows the number of matched bacteria of the laptop keyboards of all participants and the fingertips of male participants. (D) shows the number of matched bacteria of the laptop keyboards of all participants and the fingertips of female participants. Abbreviations: eDB, environment-related bacterial database; M1∼M4, male participants; F1∼F4, female participants; FT, fingertips; MP, mobile phone; LT, laptop keyboards.
presenting a new perspective analysis approach that includes a control period and questionnaire for analysis of large-scale microbial database. In most studies, different types of databases are used to reliably identify microorganisms [24, 25], and the results for bacterial analysis and nomenclature differ depending on the reference database. In contrast, as we used only one reference database of NCBI, it has no significant effect in this study. This is because the main purpose of this study was to compare the accuracy according to the three specific categories defined in the Methods section (4).

This study analyzed bacteriological similarity between individuals and personal belongings (environment), referring to previous studies [8-10]. The fingertips were selected as a body part that easily contact with the external environment, and the mobile phone and laptop keyboard were selected as the personal belongings that frequently contact an individual’s fingertips. The common point between previous and this study is that experimental analysis was conducted on microorganisms derived from humans. The comparison of the accuracy of analysis using hDB and eDB separately and using one reference database indicated that the accuracies were different. Notably, the accuracy of bacterial matching was higher in hDB than in eDB. All participants were exposed to the same environment for 5∼6 h each day, including the same floor, as the students belonged to the same university. This is consistent with a high degree of similarity reported among the microbial community of a family living together, wherein many environmental bacteria were shared among the participants [10].

Thus, hDB, which eliminates environmentally related bacteria by bacterial profiling, allows for a more accurate identification of the participant in microbial community studies. Additionally, the similarity of the participants’ fingers, laptops and mobile phones can be analyzed to find the owner of the object (Table 2). As shown in Tables 2 and 3, analysis was conducted to compare the accuracy between personal identification through the formula and personal identification considering gender. With the criteria of gender, the probability of similarity by formula increased by approximately 20% (Table 2, 3). This may suggest that personal identification using not only human-related bacteria in hDB, but also gender-specific bacteria is required. In addition, instead of qualitative analysis of human samples, by adding gender and quantitative analysis methods that show a higher probability similarity, the approach described here can be used without problems in all ngs applications involving microorganisms such as forensic science (Table 3).

In summary, our study confirmed the microbial similarity between individuals and their personal belongings, based on which we propose new formulas to make this match clearer. In addition, we propose a method of using hDB and eDB to analyze the identified microbial community to enable matching analysis with a higher level of accuracy. Based on sample type, it was confirmed that the analysis using hDB for the human sample and eDB for the environmental sample can obtain more effective and accurate results. This not only contributes to various microorganisms-related fields, but also suggests an analytical method for big data that improves accuracy through simple formulas and databases for the similarity analysis. However, our study was limited as it involved only four healthy males and four females in their 20s. Moreover, consideration was not given to those who shared the same places. In future, our method should be validated with a greater number of subjects, studies on subjects exposed to the same environment and those exposed to completely different environments, and by evaluating subjects of different age groups and with disease conditions.

요약

미생물 연구에서 대량의 마이크로바이옴 데이터를 효율적으로 얻는 기술이 발전해왔지만, 마이크로바이옴 빅 데이터를 적절하게 분석하는 도구는 여전히 부족하다. 또한 다양한 데이터 베이스를 사용하여 미생물 군집을 분석하면 잘못된 결과를 초래
Acknowledgements: This paper was provided by Eulji University in 2019 and supported and funded by the Korean National Police Agency. [Project Name: Development of bloodstain age analysis technique using next generation sequencing methods. Project Number: PA-I000001] Conflict of interest: None

Author’s information (Position): You HS1,2, Graduate student; Ok YJ1, Graduate student; Lee SH1,2, Graduate student; Lee SL1,2, Graduate student; Lee YJ1, Undergraduate student; Lee MH2,3, Professor; Hyun SH1,2, Professor.

### SUPPLEMENTARY DATA

Supplementary data can be found with this article online at https://doi.org/10.15324/kjcls.2020.52.3.202.

### REFERENCES

1. Mardis ER. Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. 2008;9: 387-402. https://doi.org/10.1146/annurev.genom.9.081307.164559
2. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. Nucleic Acids Res. 2016;44(D1):D694-D697. https://doi.org/10.1093/nar/gkw1299
3. Jansson JK, Prosser JI. Microbiology: The life beneath our feet. Nature. 2013;494:40-41. https://doi.org/10.1038/49440a
4. Lal R. The new science of metagenomics: fourth domain of life. Indian J Microbiol. 2011;51:245-246. https://doi.org/10.1007/s12088-011-0183-5
5. Henderson G, Yilmaz P, Kumar S, Forster RJ, Kelly WJ, Leahy SC, et al. Improved taxonomic assignment of rumen bacterial 16S rRNA sequences using a revised SILVA taxonomic framework. PeerJ. 2019;7:e7690. https://doi.org/10.7717/peerj.7690
6. Rajan SK, Lindqvist M, Brummer RJ, Schultz I, Repslib D. Phlogenetic microbiota profiling in fecal samples depends on combination of sequencing depth and choice of NGS analysis method. PLoS ONE. 2019;14:E0222171. https://doi.org/10.1371/journal.pone.0222171
7. Woemer AE, Novoski NM, Wendt FR, Ambers A, Wiley R, Schermede SE, et al. Forensic human identification with targeted microbe markers using nearest neighbor classification. Forensic Science International. Genetics. 2019;38:130-139. https://doi.org/10.1016/j.fsigen.2018.10.003
8. Blaser MJ, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Estrada L, et al. Distinct cutaneous bacterial assemblages in a sampling of South American Amerindians and US residents. ISME J. 2013;7:85-95. https://doi.org/10.1111/1751-736X.12081
9. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Forensic identification using skin bacterial communities. Proc Nutl Acad Sci. 2010;107:6477-6481. https://doi.org/10.1073/pnas.1000162107
10. Ravindra S, Hampton-Marcell JT, Gibbons SM, Colares GB, Smith D, Eisen JA, et al. Forensic analysis of the microbiome of phones and shoes. Microbiome. 2015;3:21. https://doi.org/10.1186/s13029-015-0082-9
11. Meadow JF, Altrichter AE, Bateman AC, Steenson J, Brown GZ, Green JL, et al. Humans differ in their personal microbial cloud. PeerJ. 2015;3:E1258. https://doi.org/10.7717/peerj.1258
12. Qian J, Hospodsky D, Yamamoto N, Nazaroff WW, Peccia J. Size resolved emission rates of airborne bacteria and fungi in an occupied classroom. Indoor Air. 2012;22:339-351. https://doi.org/10.1111/j.1600-0668.2012.00769.x
13. Adams RJ, Bhargar S, Faut W, Arens EA, Taylor JW, Lindow SE, et al. Chamber bioaerosol study: outdoor air and human occupants as sources of indoor airborne microbes. PLoS ONE. 2015;10:E0120022. https://doi.org/10.1371/journal.pone.0120022
14. Oh J, Byrd AL, Park M, Kong HH, Segre JA. Temporal stability of the human skin microbiome. Cell. 2016;165:854-866. https://doi.org/10.1016/j.cell.2016.04.008
15. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. Science 2009;326:1694-1697. https://doi.org/10.1126/science.1177486
16. Fadrosh DW, Ma B, Gjater P, Sengamulay N, Ott, Brotman RM, Ravel J. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform Microbiome. 2014;2:6. https://doi.org/10.1186/2049-2618-2-6
17. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6:1621-1642. https://doi.org/10.1038/ismej.2012.8
18. Vasileiadis S, Puglisi E, Arena M, Cappa F, Cocconcelli PS, Trevican M. Soil bacterial diversity screening using single 16S rRNA gene V regions coupled with multi-million read generating sequencing technologies. PLoS ONE. 2012;7:E472671 https://
19. Kim HN, Yun Y, Ryu S, Chang Y, Kwon MJ, Cho J, et al. Correlation between gut microbiota and personality in adults: a cross-sectional study. Brain Behav Immun. 2018;69:374-385. https://doi.org/10.1016/j.bbi.2017.12.012

20. Grice EA, Kong HH, Renaud G, Young AC. NISC Comparative Sequencing Program, Bouffard GG, et al. A diversity profile of the human skin microbiota. Genome Res. 2008;18:1043-1050. https://doi.org/10.1101/gr.075549.107

21. Oh J, Byrd AL, Park M, NISC Comparative Sequencing Program, Kong HH, Segre JA. Temporal stability of the human skin microbiome. Cell. 2016;165:854-566. https://doi.org/10.1016/j.cell.2016.04.008

22. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018;555:210-215. https://doi.org/10.1038/nature25973

23. Si JY, Lee SH, Park JM, Sung JH, Ko GP. Genetic associations and shared environmental effects on the skin microbiome of Korean twins. BMC genomics. 2015;16:992. https://doi.org/10.1186/s12864-015-2131-y

24. Petrosino JF, Highlander S, Luna RA, Gibbs RA, Verslovic J. Metagenomic pyrosequencing and microbial identification. Clin Chem. 2009;55:856-866. https://doi.org/10.1373/clinchem.2008.107565

25. Tringe SG, Mering CV, Kobyashi A, Salamov AA, Chen K, Chang HW, et al. Comparative metagenomics of microbial communities. Science. 2005;308:554-557. https://doi.org/10.1126/science.1107851