J. Microbiol. Biotechnol. (2017), 27(12), 2089–2093
https://doi.org/10.4014/jmb.1709.09027

Review

jmb

Deciphering Diversity Indices for a Better Understanding of Microbial Communities

Bo-Ra Kim[1†], Jiwon Shin[1†], Robin B. Guevarra[1], Jun Hyung Lee[1], Doo Wan Kim[2], Kuk-Hwan Seol[2], Ju-Hoon Lee[3], Hyeun Bum Kim[1*], and Richard E. Isaacson[4*]

[1]*Department of Animal Resources Science, Dankook University, Cheonan 31116, Republic of Korea*
[2]*National Institute of Animal Science, Rural Development Administration, Wanju 55365, Republic of Korea*
[3]*Department of Food Science and Biotechnology, Institute of Life Science and Resources, Kyung Hee University, Youngin 17104, Republic of Korea*
[4]*Department of Veterinary and Biomedical Sciences, University of Minnesota, St. Paul, MN 55108, USA*

Received: September 14, 2017
Revised: September 29, 2017
Accepted: October 12, 2017

First published online
October 14, 2017

**Corresponding authors*
R.E.I.
Phone: +1-612-624-0701;
Fax: +1-612-624-8707;
E-mail: isaac015@umn.edu
H.B.K.
Phone: +82-41-550-3653;
Fax: +82-41-559-7881;
E-mail: hbkim@dankook.ac.kr

†These authors contributed equally to this work.

pISSN 1017-7825, eISSN 1738-8872

The past decades have been a golden era during which great tasks were accomplished in the field of microbiology, including food microbiology. In the past, culture-dependent methods have been the primary choice to investigate bacterial diversity. However, using culture-independent high-throughput sequencing of 16S rRNA genes has greatly facilitated studies exploring the microbial compositions and dynamics associated with health and diseases. These culture-independent DNA-based studies generate large-scale data sets that describe the microbial composition of a certain niche. Consequently, understanding microbial diversity becomes of greater importance when investigating the composition, function, and dynamics of the microbiota associated with health and diseases. Even though there is no general agreement on which diversity index is the best to use, diversity indices have been used to compare the diversity among samples and between treatments with controls. Tools such as the Shannon-Weaver index and Simpson index can be used to describe population diversity in samples. The purpose of this review is to explain the principles of diversity indices, such as Shannon-Weaver and Simpson, to aid general microbiologists in better understanding bacterial communities. In this review, important questions concerning microbial diversity are addressed. Information from this review should facilitate evidence-based strategies to explore microbial communities.

Keywords: Microbiota, microbial diversity, microbial ecology, diversity index

Introduction

The microbiome, the totality of microbes, is found in and on all subjects from plants to animals. For example, in a stable gastrointestinal (GI) ecosystem, all available niches are inhabited by components of the microbiome, the collection of microorganisms that normally occupy the GI tract. Any transient species derived from foreign sources other than the GI ecosystem will pass through the GI tract without colonization [1, 2]. Nevertheless, the GI tract ecosystem is very complex and dynamic. It has been estimated that a total of about 10^{14} bacteria populate the gut microbiome, and that there are 500–1,000 bacterial species present in the GI tract [3–5].

The majority of the bacteria in the GIT are yet to be discovered and are currently "unculturable" using standard methods, although modern molecular techniques have led to the characterization of complex bacterial communities. The development of a culture-independent method based on the PCR amplification, cloning, and sequencing of the 16S rRNA gene has been limited because of high costs and the lack of throughput [5, 6]. However, the recent application of targeted DNA sequence analysis of the 16S rRNA gene coupled with next-generation sequencing technologies now enables us to intensively explore microbial communities of the GI tract and describe overall diversity [7, 8].

With the development of high-throughput DNA sequencing, characterization of microbial populations is advancing at an accelerated pace. The approach to use high-throughput next-generation sequencing in combination with taxonomic classification of 16S rRNA genes maximizes the bacterial species identification with high-resolution power [9–12].

The important features of a bacterial community in a certain niche are characterized by the number of species present and their numerical composition, bacterial diversity. In order to compare the bacterial diversity from samples of microorganisms, a variety of bioinformatics tools have been developed [13–15]. Shannon-Weaver and Simpson diversity indices are commonly used in bacterial diversity measurement based on operational taxonomic units (OTUs). OTUs are inferred to exist based on sequence data, and can be defined at different levels of resolution (phylum, class, order, family, genus, and species). Rarefaction (a statistical technique used to approximate the number of OTUs expected in a random sample of individuals taken from a sample collection) can be used to measure bacterial richness (*i.e.*, relative richness; measurement of OTUs actually observed in samples), whereas ACE (Abundance-based Coverage Estimator) and Chao1 indices are used to estimate richness (estimated richness; measurement of OTUs expected in samples given all the bacterial species that were identified in the samples) [16–18].

In the analysis of microbial community diversity, there is no general agreement on which diversity index is the best to use [19]. However, the uses of Shannon-Weaver and Simpson diversity indices have been recommended to robustly measure microbial diversity [20]. Here in, we describe the estimates of species richness and evenness in the study of structure, function, and evolution of microbial communities. The purpose of this review is to explain principles of diversity indices, such as Shannon-Weaver and Simpson, to aid the general microbiologist to better understand bacterial communities. In this review, important questions concerning microbial diversity are addressed.

Shannon-Weaver and Simpson Diversity Indices

A definition of biodiversity is widely cited as follows: "Biological diversity means the variability among living organisms from the ecological complexes of which organisms are part, and it is defined as species richness and relative species abundance in space and time" [14]. A variety of approaches have been used to quantify biological diversity. Two main factors, richness and evenness, should be taken into account when measuring the diversity of certain samples. A measure of the number of different kinds of organisms present in a particular community is defined as richness; thus, species richness refers to the number of different species present in a certain niche. If more species are present in "A" than "B", "A" is richer than "B". When it comes to species richness, it does not consider the number of individuals of each species present (Figs. 1A and 1B). Nevertheless, diversity depends not only on richness, but also on evenness. Evenness compares the uniformity of the population size of each of the species

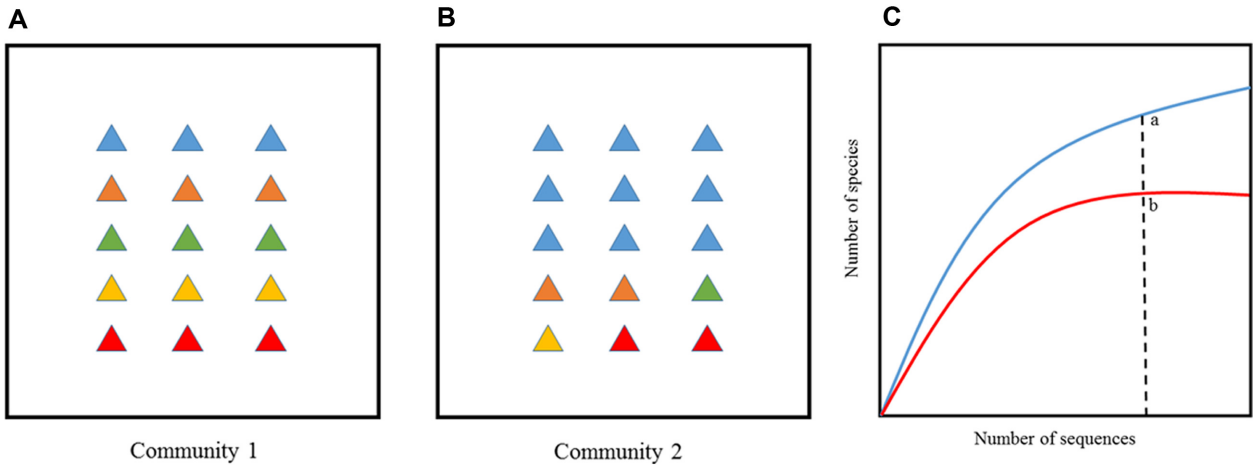

Fig. 1. Species richness, evenness, and rarefaction curve.
Both communities 1 (**A**) and 2 (**B**) have the same species richness, five species each. However, organisms in community 1 (**A**) are more evenly distributed than in community 2 (**B**). With the same sampling efforts, **A** is more diverse than **B** based on the rarefaction curve (**C**). The triangles represent bacterial species, and different species are presented in different colors.

present. A measure of the relative abundance of the different species consisting of a community is evenness (Figs. 1A and 1B). In general, when species richness and evenness increase, diversity does so too. A species diversity index denotes a mathematical measure of species diversity in a community. Shannon-Weaver and Simpson diversity indices have been traditionally used to measure the diversity of communities [15].

Shannon-Weaver and Simpson diversity indices provide more inference about the community composition than simple species richness or evenness (Table 1). Both also contemplate the relative abundances of different species. By considering relative abundances, a diversity index depends on both species richness and the evenness, or equitableness, with which individuals are distributed among the different species. However, both diversity indices have specific biases. The Shannon-Weaver index places a greater weight on species richness, whereas the Simpson index considers species evenness more than species richness in its measurement [14, 15].

In addition, the Shannon-Weaver index measures the average degree of uncertainty in predicting where individual species chosen at random will belong. The value increases as the number of species increases and as the distribution of individuals among the species becomes even [21, 22]. On the other hand, the Simpson index indicates the species dominance and reflects the probability of two individuals that belong to the same species being randomly chosen. It varies from 0 to 1 and the index increases as the diversity decreases [23].

When we compare Shannon-Weaver or Simpson diversity indices between samples with different number of sequences, normalization of the number of sequences in all samples is important before examining microbial community diversity, because these diversity index values increase as the number of sample size increases; hence normalization is crucial to avoid biases in the results [21].

Rarefaction

When microbiota are collected from a certain niche, there is a need to evaluate how well a sample reflects the true diversity of the specific niche, which is synonymous with species richness and relative abundance in time and space [14]. The advent of cutting-edge biological techniques, such as high-throughput sequencing technology, has uncovered

Table 1. Ecological diversity measures commonly used in microbial ecology studies.

Diversity indices/ Parameters	Description	Formula	Reference
Shannon diversity index (H)	Estimator of species richness and species evenness: more weight on species richness	$H = -\sum_{i=1}^{s}(p_i \ln p_i)$ where s is the number of OTUs and p_i is the proportion of the community represented by OTU i.	Lemos *et al*. [21] Magurran [22]
Simpson's index (D)	Estimator of species richness and species evenness: more weight on species evenness	$D = \frac{1}{\sum_{i=1}^{s} p_i^2}$ where s is the total number of species in the community and p_i is the proportion of community represented by OTU i.	Simpson [23] Lemos *et al*. [21]
ACE	Abundance-based coverage estimator of species richness	$S_{ACE} = S_{abund} + \frac{S_{rare}}{C_{ACE}} + \frac{F_1}{C_{ACE}}\gamma^2_{ACE}$ where S_{abund} and S_{rare} are the number of abundant and rare OTUs, respectively, C_{ACE} is the sample abundance coverage estimator, F_1 is the frequency of singletons, and γ^2_{ACE} is the estimated coefficient of variation for rare OTUs. The estimated coefficient of variation (γ^2_{ACE}) is defined as $\gamma^2_{ACE} = max\left[\frac{S_{rare}}{C_{ACE}}\frac{\sum_{i=1}^{10} i(i-1)F_i}{(N_{rare})(N_{rare}-1)} - 1, 0\right]$	Chao and Lee [28] Chao *et al*. [29]
Chao1	Abundance-based estimator of species richness	$S_{Chao1} = S_{obs} + \frac{F_1(F_1-1)}{2(F_2+1)}$ where F_1 and F_2 are the count of singletons and doubletons, respectively, and S_{obs} is the number of observed species.	Chao [16]

a variety of species that were not detected with conventional culture-dependent methods and morphological identification. Nevertheless, it is still impossible to discover all the species of microbial communities. Consequently, microbiologists must depend on samples to disclose the actual diversity of microbial communities. A number of statistical approaches have been developed to compare species richness between samples. Rarefaction curves measure OTUs observed with a given depth of sequencing, and are used to compare observed richness among communities that have been unequally sampled [24].

The simple method to measure species richness is to count the number of species present in the community. Some communities are simple enough to enable a complete count of the species numbers present. However, it is often impossible to count all the species in a community of microorganisms. In addition, a typical problem emerges from comparing samples of different sizes. A richness measurement is affected by a sample size. Therefore, it is difficult to determine immediately which community has higher species richness when we compare samples of different sizes. One way to overcome this problem is to standardize all samples from different communities to a common sample size of the same number of individuals [24, 25].

Rarefaction is a statistical technique to approximate the number of species expected in a random sample of individuals taken from a sample collection. Rarefaction permits direct comparisons of samples of different sizes of their sample sizes. The rarefaction method is dependent upon the shape of the species abundance curve and discovery rate rather than the absolute number of species per sample (Fig. 1C) [25]. Rarefaction informs us if the sample composed of a certain number of individuals would likely have been there [24]. A *t*-test can be used to test if rarefaction curves are significantly different from one another [19]. Moreover, the bootstrapping method can also be used to estimate the precision of rarefaction curves [26].

Chao1 and the Abundance-Based Coverage Estimator

In taxon-based approaches, a practical question is how to estimate the number of microbial species in a given sample since the degree of microbial diversity is often difficult to decipher. Accurate assessment of species richness is useful for the effective analysis of biological communities [27]. Contrary to rarefaction, which compares observed richness among samples, richness estimators evaluate the total richness of a community from a sample [24]. Chao1 and ACE have been developed to estimate richness, and they calculate expected OTUs based on observed OTUs (Table 1) [13, 16, 28, 29].

Chao1 is a nonparametric method for estimating the number of species in a community. The Chao richness estimator was developed by Anne Chao and is based on the concept that rare species infer the most information about the number of missing species. Because the Chao richness estimator gives more weight to the low abundance species, only the singletons and doubletons are used to estimate the number of missing species [16]. Therefore, this index is particularly useful for data sets skewed toward the low-abundance species [24].

The ACE is a nonparametric method for estimating the number of species using sample coverage, which is defined as the sum of the probabilities of the observed species. The ACE method divides observed frequencies into abundant and rare groups. The abundant species are those with more than 10 individuals in the sample, and the rare species are those with fewer than 10 individuals. Only the presence or absence information of abundant species is considered in the ACE method because they would be discovered anyway. Therefore, the exact frequencies for the abundant species are not required in the ACE method. On the other hand, the exact frequencies for the rare species are required because the estimation of the number of missing species is based entirely on these rare species [24, 28–30].

Concluding Remarks

The microbiome has been known to play important roles in the well-being and health of animals. Thus, there have been disparate endeavors to better understand the mechanisms of actions of the microbiome contributing to the overall health of the hosts. With the help of next-generation high-throughput sequencing, immense information on the microbiome has been generated. Subsequently, understanding diversity indices became more necessary to decipher the microbial communities. A review of this literature addresses important questions concerning the microbial diversities, such as Shannon-Weaver and Simpson diversity indices, to aid in a better understanding of bacterial communities. Information from this review should facilitate evidence-based strategies to explore microbial communities.

Acknowledgments

The present study was supported by the Rural Development

Administration (Project No. PJ012279), the Individual Basic Science & Engineering Research Program (No. NRF-2015R1D1A1A01061268) through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, and a research fund (14162MFDS972) from the Ministry of Food and Drug Safety, Republic of Korea.

References

1. Berg RD. 1996. The indigenous gastrointestinal microflora. *Trends Microbiol.* **4:** 430-435.
2. Manson JM, Rauch M, Gilmore MS. 2008. The commensal microbiology of the gastrointestinal tract. *Adv. Exp. Med. Biol.* **635:** 15-28.
3. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, *et al.* 2003. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* **299:** 2074-2076.
4. Sonnenburg JL, Angenent LT, Gordon JI. 2004. Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? *Nat. Immun.* **5:** 569-573.
5. Kim HB, Isaacson RE. 2015. The pig gut microbial diversity: understanding the pig gut microbial ecology through the next generation high throughput sequencing. *Vet. Microbiol.* **177:** 242-251.
6. Isaacson R, Kim HB. 2012. The intestinal microbiome of the pig. *Anim. Health Res. Rev.* **13:** 100-109.
7. Schmalenberger A, Schwieger F, Tebbe CC. 2001. Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. *Appl. Environ. Microbiol.* **67:** 3557-3563.
8. Chakravorty S, Helb D, Burday M, Connell N, Alland D. 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* **69:** 330-339.
9. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, *et al.* 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. USA* **103:** 12115-12120.
10. Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, *et al.* 2007. Microbial population structures in the deep marine biosphere. *Science* **318:** 97-100.
11. Highlander SK. 2012. High throughput sequencing methods for microbiome profiling: application to food animal systems. *Anim. Health Res. Rev.* **13:** 40-53.
12. Sanschagrin S, Yergeau E. 2014. Next-generation sequencing of 16S ribosomal RNA gene amplicons. *J. Vis. Exp.* DOI: 10.3791/51709.
13. Schloss PD, Handelsman J. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71:** 1501-1506.
14. Schloss PD, Handelsman J. 2006. Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Appl. Environ. Microbiol.* **72:** 6773-6779.
15. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et al.* 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75:** 7537-7541.
16. Chao A. 1984. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* **11:** 265-270.
17. Chao A, Bunge J. 2002. Estimating the number of species in a stochastic abundance model. *Biometrics* **58:** 531-539.
18. Chao A, Chazdon RL, Colwell RK, Shen TJ. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* **62:** 361-371.
19. Hughes JB, Bohannan BJM. 2004. Application of ecological diversity statistics in microbial ecology. *Mol. Microb. Ecol. Manual* **7.01:** 1321-1344.
20. Haegeman B, Hamelin J, Moriarty J, Neal P, Dushoff J, Weitz JS. 2013. Robust estimation of microbial diversity in theory and in practice. *ISME J.* **7:** 1092-1101.
21. Lemos LN, Fulthorpe RR, Triplett EW, Roesch LF. 2011. Rethinking microbial diversity analysis in the high throughput sequencing era. *J. Microbiol. Methods* **86:** 42-51.
22. Magurran A. 2004. *Measuring Biological Diversity*. Blackwell Science Ltd, Oxford, United Kingdom.
23. Simpson EH. 1949. Measurement of diversity. *Nature* **163:** 688.
24. Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJ. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67:** 4399-4406.
25. Sanders HL. 1969. Benthic marine diversity and the stability-time hypothesis. *Brookhaven Symp. Biol.* **22:** 71-81.
26. Efron B, Tibshirani R. 1993. *An Introduction to the Bootsrap*. *Chapman & Hall, New York.*
27. Shen TJ, Chao A, Ling C. 2003. Predicting the number of new species in further taxonomic sampling. *Ecology* **84:** 798-804.
28. Chao A, Lee S. 1992. Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* **87:** 210-217.
29. Chao A, Ma M, Yang M. 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* **80:** 193-201.
30. Sornplang P, Piyadeatsoontorn S. 2016. Probiotic isolates from unconventional sources: a review. *J. Anim. Sci. Technol.* **58:** 26.