A New Biodiversity Index and the Corresponding Index of Evenness: A Simple Theoretical Analysis

Sudipto Roy*[1] and Kankana Roy Bhattacharya[2]
[1]*Department of Physics, St. Xavier's College, Kolkata, West Bengal, India*
[2]*Department of Zoology, Raja Peary Mohan College, Uttarpara, Hooghly, West Bengal, India*
**Corresponding author. Email: roy.sudipto@sxccal.edu*

Received on: August 06, 2023 | Accepted on: September 06, 2023 | Published on: September 29, 2023

Abstract

In the present article, we have proposed a new biodiversity index based on the standard deviation of the number of individuals belonging to a species in a collection of biological organisms of different species. Using this index of biodiversity, we have proposed an index for the measurement of evenness of the distribution of individuals among different species in the collection of organisms. Using a hypothetical dataset representing the distribution of individuals among six species, for six different samples, we have calculated the indices defined by us and compared their values with the values of Shannon-Wiener diversity index, Simpson diversity index and their corresponding evenness indices calculated from the same dataset. It is observed that, our new indices undergo much greater changes, compared to the changes of the most commonly used indices, due to any change in the relative proportions of species present in a sample. This observation indicates that any change in the number of organisms of a species, in an area or habitat, is better reflected in the values of these new indices.

Keywords: *Biodiversity index, Evenness index, Species richness, Species abundance, Relative abundance, Shannon-Wiener index, Simpson index, Ecology*

1. Introduction

The word *biodiversity* stands for the variety or diversity of biological species in the entire world or in a particular habitat. It represents variations in natural systems in terms of both number and frequency. The results of evolution for billions of years, governed by natural processes and influenced by humans, is what we observe as biodiversity. We are an integral part of the web of life generated and represented by biodiversity. According to a recent estimate there are 8.7 million eukaryotic species in this planet of which 2.2 million belong to the oceans [1]. Biodiversity has its greatest manifestation in the variety of ecosystems such as those we find in forests, deserts, mountains, wetlands, rivers, lakes and agricultural fields. A simple measure of the variety of species in any collection of living organisms is a parameter called *species richness* which stands for the total number of species found in the sample of interest. The number of individuals of a certain species present in the sample is called the *species abundance* of that species. If the value of this parameter is the same for all species in a sample, the diversity is regarded as the highest for the sample among all samples having the same *species richness*. If one or two species are represented by a huge number of individuals in comparison to other species, the diversity is regarded as low in the sample of interest. Biodiversity does not have a uniform distribution on this planet. The diversity is found to be the highest in the tropics. The terrestrial biodiversity is found to be the largest near the equator, mainly due to the warm climate and high primary productivity [2, 3]. The highest marine biodiversity is observed along the coasts in the Western Pacific and in the mid-latitudinal band in all oceans, where the temperature of the sea surface remains the highest. The diversity of species shows a variation with latitude [4]. The areas where we witness the presence of endemic species (species which are observed only in a particular geographical region of the earth) with high concentrations are often designated as the biodiversity hotspots [5]. Diversity cannot go on increasing with time indefinitely in an area. The rate of increase of *species richness* is likely to become smaller in future [6]. Considering biodiversity to be a representation of the variety of life and the processes it involves, it encompasses the variety of ecosystems, communities, living organisms and their genetic differences [7].

Although *species richness* is a primary estimate of the variety of species in a given sample, it cannot alone be an accurate measure of biodiversity since it does not take into account the relative proportions of species in the sample. It needs to be combined with other measures to formulate a reliable index of diversity [8]. Another less obvious but equally important aspect is the distribution of individuals among different species in a collection of biological organisms (i.e., species equitability or evenness). During the last 50 years, several numerical indices have been proposed to quantify these two properties (diversity and evenness) of a sample. In scientific literature, one comes across lots of studies showing a number of methods or theories to measure biodiversity, based on various sets of data which have been collected by employing various techniques [9-12]. Although all indices are correlated and depend on the sample size (N), no single diversity index can adequately describe the population structure in all situations [13, 14].

In the present study, we propose a new index of biodiversity (denoted by K) and the corresponding index of evenness (denoted by E_K). For a hypothetical set of data regarding the populations of six biological species in six different samples, we have calculated these new indices and compared their values with those of some widely used indices based on the same dataset. An important observation is that the new indices are more sensitive than the indices which are commonly used, to any change in the composition of a sample.

2. Some Commonly Used Indices

One of the most widely used indices to measure biodiversity is Shannon-Wiener diversity index (H'), which is expressed as [15],

$$H' = -\sum_{i=1}^{S} p_i \ln p_i \qquad (1)$$

In equation (1), $p_i = \frac{n_i}{N} = \frac{n_i}{\sum_{i=1}^{S} n_i}$ is known as the relative abundance of the i^{th} species. $S =$ total number of species present in a sample which is known as species richness. $n_i =$ number of organisms of the i^{th} species (known as species abundance) and $N = \sum_{i=1}^{S} n_i =$ sample size (total number of organisms in the sample of all species combined). It can be shown that, when the number of individuals of each species in a sample is the same (i.e., $n_1 = n_2 = n_3 = \cdots = n_S$), H' reaches its maximum value, i.e., $H' = H'_{max} = \ln S$. The concept of H' is based on information theory and it represents the uncertainty about the identity of an individual chosen randomly from the sample [16]. In a system which is highly diverse, an arbitrarily chosen individual can belong to any species, reducing the predictability of its identity [15]. Some modified versions of Shannon-Wiener diversity index have recently been proposed and tested by some researchers [17, 18].

Apart from H', a pair of very commonly used measures of diversity are Simpson's diversity indices (denoted here by D_1 and D_2), which are represented by the following equations [19, 20].

$$D_1 = 1 - \sum_{i=1}^{S} p_i^2 \quad (2)$$

$$D_2 = \frac{1}{\sum_{i=1}^{S} p_i^2} \quad (3)$$

D_1 is the complement of Simpson's original index $\left(\lambda = \sum_{i=1}^{S} p_i^2\right)$. It represents the probability that two randomly chosen individuals are of two different species [21]. D_2 is the inverse of Simpson's original index and closely related with D_1 [19]. D_2 is more widely used among these two indices [16]. When the species have equal abundances (n_i) in a sample, both D_1 and D_2 reach their maximum values, which are, $D_{1max} = 1 - 1/S$ and $D_{2max} = S$ respectively. Here $\sum_{i=1}^{S} p_i^2$ is known as dominance index, which can be regarded as a measure of dominance of one or more species among all species in a sample [22]. Its value is less affected or influenced by the existence of species with smaller values of the relative abundance (p_i) in a sample. Augousti *et al* have recently proposed a new index whose value is close to Simpson index value for communities with large number of individuals and large number of species with equal relative abundance [23]. Zhou *et al* have shown how Simpson's index can be used for the purpose of diversifying multi-aspect search results [24].

Evenness in a sample is a measure of the degree of closeness of the numbers of individuals belonging to different species. It is low when one or a few species dominate in the sample. A sample is said to have a high degree of evenness when the values of p_i (relative abundance) for different species are sufficiently close to one another. To quantify the equitability of a distribution, different diversity indices (H', D_1, D_2) are generally used, as they contain the components of richness and evenness inherently.

An index of evenness, named Pielou's index, is expressed by the following equation [25].

$$E_P = \frac{H'}{H'_{max}} = \frac{H'}{\ln S} \quad (4)$$

Another such index, named Buzas & Gibson's evenness, is defined as [26],

$$E_{BG} = \frac{e^{H'}}{e^{H'_{max}}} = \frac{e^{H'}}{S} \quad (5)$$

Using Simpson's diversity index D_1, one may define the following index (E_{S1}) for evenness.

$$E_{S1} = \frac{D_1}{D_{1max}} = \frac{1-\sum_{i=1}^{S} p_i^2}{1-1/S} \quad (6)$$

In terms of another Simpson's index (i.e., D_2), an evenness index (E_{S2}) is expressed as [20],

$$E_{S2} = \frac{D_2}{D_{2max}} = \frac{1/S}{\sum_{i=1}^{S} p_i^2} \quad (7)$$

For each of the above evenness indices (eqns. 4-7), the maximum value is 1.

For a simple estimate of diversity based on *species richness* (S), one often uses Margalef index, which is expressed as [27],

$$D_{mg} = (S - 1) / \ln(N) \quad (8)$$

Another such index, having a linear relationship with S, is Menhinick index (D_{mn}) [28]. It is given by,

$$D_{mn} = \frac{S}{\sqrt{N}} \quad (9)$$

D_{mg} and D_{mn} cannot properly quantify diversity because they are independent of p_i for different species in a sample. For a fixed combination of values of S and N, there can be several combinations of p_i values in a sample. Thus, for samples with the same values of S and N, the values of D_{mg} (or D_{mn}) are the same.

In an article by Mulya *et al*, one gets a detailed comparative study of some widely used diversity indices [29]. Detailed information regarding the research on biodiversity conservation can be obtained from an article by Jurkus *et al* [30].

3. New Indices of Biodiversity and Evenness Proposed in the Present Study

We propose a new biodiversity index (K), represented by the following equation.

$$K = \frac{S}{1+\sigma_n} \quad (10)$$

In equation (10), σ_n is the standard deviation of the values of n_i (*species abundance*) which stands for the number of individuals of the i^{th} species in a sample. If the species in a sample are present in equal numbers, we have $\sigma_n = 0$, leading to $K = K_{max} = S$.

The standard deviation (σ_n) of the *species abundance* (n_i) values is expressed as,

$$\sigma_n = \sqrt{\frac{1}{S}\sum_{i=1}^{S}(n_i - \bar{n})^2} \quad (11)$$

where $\bar{n}$ is the mean of the n_i values, which is expressed as,

$$\bar{n} = \frac{1}{S}\sum_{j=1}^{S} n_j = \frac{N}{S} \quad (12)$$

Substituting equation (11) into equation (10), we get,

$$K = \frac{S}{1+\sqrt{\frac{1}{S}\sum_{i=1}^{S}(n_i-\bar{n})^2}} \quad (13)$$

According to equation (10), if S is kept fixed, our proposed diversity index (K) decreases as σ_n increases. For a fixed value of the standard deviation (σ_n), K increases linearly with S. These observations regarding K are consistent with the characteristics of the most commonly used indices of biodiversity, as discussed in Section-2 of this article.
Based on our new diversity index (K), we propose the following evenness index (E_K) for a distribution of individuals among different species in a sample.

$$E_K = \frac{K}{K_{max}} = \frac{K}{S} \quad (14)$$

Using equation (10) in equation (14), we get,

$$E_K = \frac{1}{1+\sigma_n} \quad (15)$$

Substituting equation (11) into equation (15), we get,

$$E_K = \frac{1}{1+\sqrt{\frac{1}{S}\sum_{i=1}^{S}(n_i-\bar{n})^2}} \quad (16)$$

Thus, the maximum value of evenness (E_K) is 1 (corresponding to $\sigma_n = 0$), just like the indices expressed by equations (4-7). It varies in the range expressed as, $0 < E_K \leq 1$.

The relation $p_i = n_i/N$ leads to the expression $\sigma_p = \sigma_n/N$, where σ_p is the standard deviation of the p_i values in a sample. In terms of σ_p, our new indices can be represented by the following expressions (eqns. 17, 18).

$$K = \frac{S}{1+N\sigma_p} \quad (17)$$

$$E_K = \frac{1}{1+N\sigma_p} \quad (18)$$

The standard deviation of the p_i values (i.e., σ_p) can be easily calculated using the information regarding the proportions of different species in the sample of interest, using the formula $\sigma_p = \sqrt{\frac{1}{S}\sum_{i=1}^{S}(p_i - \bar{p})^2}$ where $\bar{p}$ denotes the mean of the p_i values, expressed as, $\bar{p} = \frac{1}{S}\sum_{j=1}^{S} p_j$. For a fixed set of values of S and N, K decreases as σ_p increases, as per equation (17).

The ratio $\sigma_n/\bar{n}$ is called the *coefficient of variation* (C_V) in statistical parlance. Thus, we can write, $\sigma_n = \bar{n}C_V = NC_V/S$ since $\bar{n} = N/S$. C_V is a measure of variation in the values of n_i relative to their mean ($\bar{n}$). In terms of C_V, our new indices can be represented by the following expressions (eqns. 19, 20).

$$K = \frac{S}{1+NC_V/S} \quad (19)$$

$$E_K = \frac{1}{1+NC_V/S} \quad (20)$$

According to equations (19) and (20), the new indices (K & E_K) are functions of N (sample size), S (species richness) and C_V (coefficient of variation) for the population of a sample. These expressions are expected to be useful for a statistical analysis based on the values of K and E_K calculated for different communities of organisms in any part of the globe.

4. Results and Discussion

For the evaluation of indices discussed in the present study, we have used a hypothetical dataset (represented by Table 1) regarding the populations of six species, in each of the six different samples.

Species Abundance (n_i)	Sample No.					
	I	II	III	IV	V	VI
n_1	6	7	8	9	10	11
n_2	6	5	4	3	2	1
n_3	6	7	8	9	10	11
n_4	6	5	4	3	2	1
n_5	6	7	8	9	10	11
n_6	6	5	4	3	2	1

Table 1 A hypothetical dataset for 6 samples, each with $\boldsymbol{N = 36}$ and $\boldsymbol{S = 6}$.

The values of various indices of biodiversity, based on the data in Table 1, have been listed in Table 2. The values of several evenness indices, based on the data in Table 1, have been listed in

Diversity Indices	Sample No.					
	I	II	III	IV	V	VI
H'	1.792	1.778	1.735	1.661	1.549	1.385
D_1	0.833	0.829	0.815	0.792	0.759	0.718
D_2	6.000	5.838	5.400	4.800	4.154	3.541
K	6.000	3.000	2.000	1.500	1.200	1.000

Table 2 Values of diversity indices based on the data of Table 1.

Evenness Indices	Sample No.					
	I	II	III	IV	V	VI
E_P	1	0.992	0.968	0.927	0.865	0.773
E_{BG}	1	0.986	0.945	0.877	0.785	0.666
E_{S1}	1	0.994	0.978	0.950	0.911	0.861
E_{S2}	1	0.973	0.900	0.800	0.692	0.590
E_K	1	0.500	0.333	0.250	0.200	0.167

Table 3 Values of evenness indices based on the data of Table 1.

Table 3. Calculations of all these indices have been carried out using their definitions given in Sections-2 & 3 of this article. The total number of

organisms (N) in each of the six samples is 36. Their distribution (among six species) is not the same in any two of these six samples.

As we move across Table 1 from left to right (i.e., from Sample-I to Sample-VI), the uniformity of distribution of individuals among different species decreases, increasing both σ_p and σ_n.

In Table 2, each of the four diversity indices is found to decrease in the direction in which the uniformity of distribution decreases (i.e., in the direction from Sample-I to Sample-VI).

In Table 3, the value of each of the five evenness indices is found to decrease (from 1 to gradually smaller values) in the direction in which the uniformity of distribution decreases (i.e., in the direction from Sample-I to Sample-VI).

The values of the indices proposed by us, for diversity and evenness measurement (i.e., K & E_K respectively), are found to be reduced by a sufficiently greater amount, compared to changes in the other indices, as the distribution of individuals changes from one sample to another in the direction from left to right.

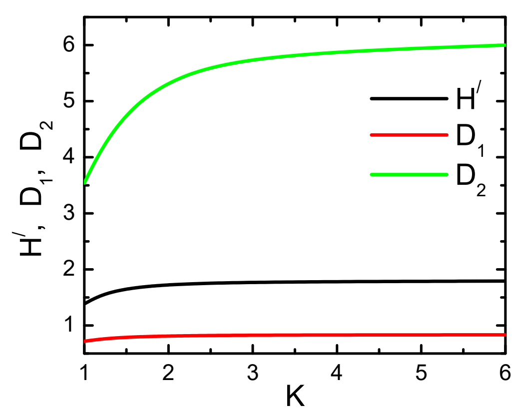

Figure 1 Plots of different biodiversity indices versus the new diversity index (K).

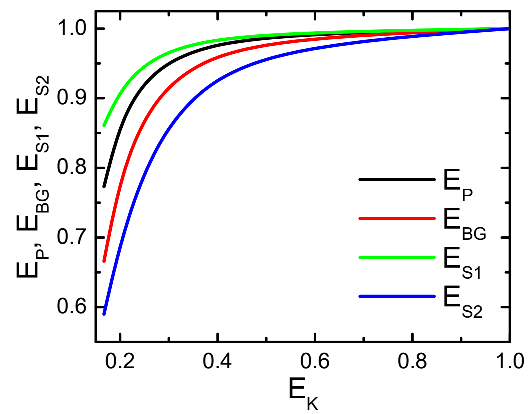

Figure 2 Plots of different evenness indices versus the new evenness index (E_K).

It should be noted in this context that, if we reshuffle the values of n_i along any column in Table 1, the values of the indices (as listed in Tables 2 & 3), corresponding to the sample represented by that column, won't change, in accordance with the mathematical expressions for these indices given in Sections-2 & 3 of this article.

The values of D_{mg} and D_{mn} are 1.395 and 1, respectively, for all samples considered here.

To compare the variations of different indices pictorially, due to changes in sample composition, we have plotted the most commonly used indices graphically against the newly proposed ones.

Figure 1 shows the variations of three biodiversity indices with respect to the new diversity index K. Figure 2 shows the variations of four evenness indices with respect to the new evenness index E_K. These two figures are based on Tables 2 & 3 respectively. It is evident that, the changes in the new indices (K & E_K), due to the changes in species abundance (n_i), are accompanied by smaller changes in the commonly used indices. It is observed that, the rate of change of any of these indices, with respect to a new index, is smaller for higher values of the new index (indicating higher diversity).

To control the sensitivity of our proposed indices to the changes in sample compositions, one may put an arbitrary parameter (say, β) in their expressions such that, $K = \frac{S}{1+\beta \;\; _{n}}$ and $E_K = \frac{1}{1+\beta\sigma_n}$ where $\beta > 0$. By tuning this parameter (β), the sensitivity of these indices (to the changes in σ_n) can be altered. For the results shown in Tables 2 & 3, we have $\beta = 1$. The difference in the values of K and E_K, between Sample-I and Sample-VI, are 5 and 0.833 respectively for $\beta = 1$. For $\beta = 2$, these values are 5.455 and 0.909 respectively. For $\beta = 3$, these values are 5.625 and 0.937 respectively. These results show that, if β is increased, the indices (K & E_K) undergo greater changes as we change the uniformity of the distribution of individuals among different species in a sample.

There are six species in each sample used for the present study. We have no specific reason for choosing this particular number. For the purpose of a proper comparison of values of different indices, we chose to calculate them for samples having the same population (N) and the same number of species (S), but with different proportions (p_i) of individual species. Our objective was to find the effect of variation of p_i values upon different indices of biodiversity and evenness. According to some studies, the *species richness* (S) depends upon the total number of individuals (N) in a sample, especially for small values of N [31, 32]. As a future extension of this work, one may examine the changes of the values of K and E_K as the sample size (N) varies.

5. Conclusion

In the present study, we have defined a new index (K) to measure biodiversity and a corresponding evenness index (E_K) to determine the equitability of the distribution of organisms among different species in a sample. These indices (represented by eqns. 10 and 15 respectively) are sufficiently easy to calculate. The characteristics of the new indices have been determined using a set of hypothetically constructed samples with the same sample size (N) and species richness (S). It is observed that, these new indices are more sensitive to changes in n_i (species abundance) values of the species in a community, in comparison to some commonly used indices (such as Shannon-Wiener diversity index, Simpson's diversity index and their corresponding evenness indices). Graphical depiction of the variations of different indices also conveys the same message. Loss of organisms of any species is expected, therefore, to be better reflected in the values of the new indices proposed here. It has been shown that these indices can be represented in a different way in terms of species richness (S), sample size (N) and coefficient of variation (C_V) which can be easily determined using the population statistics of different species in a sample. These indices can also be expressed as functions of the relative abundance (p_i) values like the commonly used indices.

References

1. Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B. How many species are there on Earth and in the ocean?. *PLoS biology*, 2011, *9*(8), e1001127, 1-8.
2. Gaston, K. J. Global patterns in biodiversity. *Nature*, 2000, *405*(6783), 220-227.
3. Field, R., Hawkins, B. A., Cornell, H. V., Currie, D. J., Diniz-Filho, J. A. F., Guégan, J. F., & Turner, J. R. Spatial species-richness gradients

across scales: a meta-analysis. *Journal of biogeography*, 2009, *36*(1), 132-147.

4. Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., & Worm, B. Global patterns and predictors of marine biodiversity across taxa. *Nature*, 2010, *466*(7310), 1098-1101.
5. Myers, N. Threatened biotas:" hot spots" in tropical forests. *Environmentalist*, 1988, *8*, 187-208.
6. Rabosky, D. L. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecology letters*, 2009, *12*(8), 735-743.
7. Wilson, E. O. The current state of biological diversity. *Biodiversity*, 1998, *521*(1), 3-18.
8. Supriatna, J. Biodiversity Indexes: Value and evaluation purposes. In *E3S Web of Conferences* (Vol. 48, p. 01001). EDP Sciences, 2018.
9. Lakićević, M., & Srđević, B. Measuring biodiversity in forest communities – a role of biodiversity indices. *Contemporary Agriculture*, 2018, *67*(1), 65-70.
10. Kumar, P., Dobriyal, M., Kale, A., Pandey, A. K., Tomar, R. S., & Thounaojam, E. Calculating forest species diversity with information-theory based indices using sentinel-2A sensor's of Mahavir Swami Wildlife Sanctuary. *Plos One*, 2022, *17*(5), e0268018, 1-17.
11. Herrmann, B., Cerbule, K., Brčić, J., Grimaldo, E., Geoffroy, M., Daase, M., & Berge, J. Accounting for uncertainties in biodiversity estimations: A new methodology and its application to the mesopelagic sound scattering layer of the high Arctic. *Frontiers in Ecology and Evolution*, 2022, *10*, 775759.
12. Roswell, M., Dushoff, J., & Winfree, R. A conceptual guide to measuring species diversity. *Oikos*, 2021, *130*(3), 321-338.
13. Buzas, M. A., & Hayek, L. A. C. Biodiversity resolution: an integrated approach. *Biodiversity Letters*, 1996, 40-43.
14. Buzas, M. A., & Hayek, L. A. C. SHE analysis for biofacies identification. *Journal of Foraminiferal Research*, 1998.
15. Shannon, C. E. A mathematical theory of communication. *The Bell system technical journal*, 1948, *27*(3), 379-423.
16. Magurran, A. E. Measuring biological diversity. *Current Biology*, 2021, *31*(19), R1174-R1177.
17. Omayio, D., Mzungu, E., & Kakamega, K. Modification of shannon-wiener diversity index towards quantitative estimation of environmental wellness and biodiversity levels under a non-comparative Scenario. *Journal of Environment and Earth Science*, 2019, *9*(9), 46-57.
18. Konopiński, M. K. Shannon diversity index: a call to replace the original Shannon's formula with unbiased estimator in the population genetics studies. *PeerJ*, 2020, *8*, e9391.
19. Simpson, E. H. Measurement of diversity. *Nature*, 1949, 163(4148), 688-688.
20. Morris, E. K., Caruso, T., Buscot, F., Fischer, M., Hancock, C., Maier, T. S., ... & Rillig, M. C. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and evolution*, 2014. *4*(18), 3514-3524.
21. McCune, B., Grace, J. B., & Urban, D. L. Analysis of ecological communities: Gleneden Beach. *Oregon, MJM Software Design*, 2002
22. Odum, E. P., & Barrett, G. W. *Fundamentals of ecology*, Vol. 3, 1971, 5. Philadelphia: Saunders.
23. Augousti, A. T., Atkins, N., Ben-Naim, A., Bignall, S., Hunter, G., Tunnicliffe, M., & Radosz, A. A new diversity index. *Physical Biology*, 2021, *18*(6), 066004.
24. Zhou, J., Agichtein, E., & Kallumadi, S. Diversifying Multi-aspect Search Results Using Simpson's Diversity Index. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, October 2020, 2345-2348.

25. Pielou, E. C. The measurement of diversity in different types of biological collections. *Journal of theoretical biology*, 1966, 13, 131-144.
26. Buzas, M. A., & Gibson, T. G. Species diversity: benthonic foraminifera in western North Atlantic. *Science*, 1969, 163(3862), 72-75.
27. Margalef, R. Temporal succession and spatial heterogeneity in natural phytoplankton. 1958.
28. Menhinick, E. F. A comparison of some species-individuals diversity indices applied to samples of field insects. *Ecology*, 1964, *45*(4), 859-861.
29. Mulya, H., Santosa, Y., & Hilwan, I. Comparison of four species diversity indices in mangrove community. *Biodiversitas Journal of Biological Diversity*, 2021, *22*(9).
30. Jurkus, E., Povilanskas, R., & Taminskas, J. Current trends and issues in research on biodiversity conservation and tourism sustainability. *Sustainability*, 2022, *14*(6), 3342.
31. Soetaert, K., & Heip, C. Sample-size dependence of diversity indices and the determination of sufficient sample size in a high-diversity deep-sea environment. *Marine Ecology Progress Series*, 1990, 305-307.
32. Buzas, M. A., and Hayek, L. A. C. On richness and evenness within and between communities. *Paleobiology*, 2005, *31*(2), 199-220.