Local taxonomic spectra in plants, animals, fungi and terrestrial protists show common mathematical patterns

D. V. Leontyev*, I. I. Yatsiuk**, T. Y. Markina*, L. P. Kharchenko*, E. V. Tverdokhleb*, I. O. Fedya***, Y. A. Yatsiuk**

*H. S. Skovoroda Kharkiv National Pedagogical University, Kharkiv, Ukraine
**Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia
***Kharkiv State Academy of Physical Culture, Kharkiv, Ukraine

Keywords: Agaricomycetes; Aves; biota; biodiversity; Bryophyta; Coleoptera; distribution fitting; Eumycetozoa; fauna; flora; Insecta; Magnoliophyta; Myxogastrea; Mxyomycetes; national park; nature conservation territories; species diversity.

Introduction

Taxonomic spectrum is a traditional term for the relation between supraspecific taxa by the number of included species (or other taxa) within a natural area (Scheiner, 2013). The description of taxonomic spectra with the naming of leading taxa serves as a part of a standard description of natural area (Scheiner, 2013). The description of taxonomic spectra with praspecific taxa by the number of included species (or other taxa) within a natural area is typical for the species abundance curves, based on collections made from small areas. At the same time, in the genera–families–orders row the similarity to the lognormal distribution increases. The central values and variability vary considerably between different taxonomic groups and ranks, however, without affecting the type of distribution. The number of orders in all taxa except Bryophyta has reached the saturation and coincides with the curve of the estimated number of species according to the Chao1 coefficient. For families and especially genera the correspondence with estimated number of species is much lower. Our results do not confirm the assumption that hollow-curve distributions of taxonomic spectra result from the artificial fragmentation of taxa. These distributions neither depend on the insufficient knowledge about the species composition at the locality, nor reflect the size of the studied area. The presence of such distributions in both local and global biota of different groups may be explained by the common features of their evolution, especially by the existence of relict orphan groups. The fact that in Hornilska Forests the kurtosis and skewness of distributions decreases in the genera–families–orders row can therefore be explained by the relatively low percentage of the high-rank orphan taxa in the local biota. This may be a common feature of communities studied at small geographical scale, since orphan taxa often demonstrate a high level of endemism. Comparative studies of local communities from different climate zones may help to understand how universal are the patterns, described herein.

The separation of taxa based on the molecular-phylogenetic data has passed the knee of the curve before 1950-s and now is very close to the asymptote of saturation for all ranks, except genera (Mora et al., 2011). It can be expected that the available data on the scope and delimitation of supraspecific taxa are representative, and their analysis will help to identify the biological patterns existing in nature (Desnoues et al., 2017). However, such an analysis is hampered by a number of historical and subjective factors, related to the classification of individual groups. The rank of taxon depends on the history of its study (for example, when sister branches already have a certain status, it determines the status of the new taxon), the presence of easily recognizable morphological features (Leontyev & Felflov, 2012), individual views of researchers, etc. (Bertrand et al., 2006). Different supraspecific taxa of a certain rank, though being perfectly monophyletic per se, are still incomparable if they are not sister taxa or were not established based on time-calibrated phylogenies. For example, the genus of bright-spored myxomycetes (lucisporemyxomycetes) cannot be considered as a biological analogue of the genus of dark-spored myxomycetes (Columbellomyxomycetes) even at the level of the variability of marker genes (Leontyev & Schnittler, 2017; Borg Dahl et al., 2018), and the order of vertebrates can differ from the order of invertebrates by the age up to 400 myr (Holt & Jonson, 2014).

Since taxonomic spectra can be seen as mathematical distributions (Bacaro et al., 2007; Scheiner, 2013), some of their properties have been already investigated. Richness, diversity, dominance and evenness estima-
tutors for such spectra have repeatedly been reported, and the distributions were compared with each other using rank correlation coefficients (Leontyev, 2007). However, the distribution fitting for taxonomic spectra still issues many challenges. It is believed that most of the species abundance spectra (the ratio between species and number of specimens) may be described by one of four classical types of distribution: logarithmic, log-normal, exponential and broken stick (Magurnan, 2004). Many other models have been proposed with different underlying ecological, mathematical or evolutionary assumptions (Matthews & Whitaker, 2014).

As for the distributions of species or other taxa within higher-rank taxa, they are known to correspond to so-called hollow-curves distributions, or HCDs (Dial & Marzluff, 1989). These distributions usually include one or several leading taxa and a long tail of outsiders. The curvature of the distribution function graph usually depends on the taxonomic rank, increasing from phylum to genus (Dial & Marzluff, 1989; Mora et al., 2011). With all that, the origin of these HCDs remains unclear. Some authors consider them as an artifact of extensive division of taxa, which occurred during the second half of 20th century. Others explain HCDs to be a result of the diversity of life history traits (body structure, rate of change of generations, etc.), which can stimulate the diversification of individual taxa or prevent it (Clayton, 1972; Dial & Marzluff, 1989; Matthews & Whitaker, 2014). It remains unknown whether the taxonomic spectra of different groups have different types of distributions, and to what extent the features of the distribution depend on the rank or on the attribution of the taxon to a particular phylogenetic group. A comprehensive assessment of the mathematical properties of the taxonomic spectrum, the identification of its inherent characteristics remains an important task.

The understanding of the taxonomic spectra reported in individual studies is greatly reduced by the incompleteness of primary data. The methods of detection of organisms e.g. season, duration and type of route (Leontyev et al., 2012; Leontyev et al., 2014) etc. Therefore, it is important to study taxonomic spectra using large checklists that integrate studies of plants, animals, fungi and protists have been systematically carried out for over a century is the Homilsha Forests National Park. The data on following taxa were included: (1) plants: Viridiplanta: Magnoliophyta (Vlashchenko et al., 2010), Bryophyta (Barskov, 2008); (2) animals: Metazoa: Coleoptera (Puchkov et al., 2010; Puchkov, 2018; Skrylnik & Bielavite, 2020), Aves (Vlashchenko et al., 2010); (3) fungi: Agaricomycetes (Prylutskyi et al., 2017); (4) protists: Amoebozoa: Myxomycetes (Prylutskyi et al., 2017). The systematic lists of organisms, taken from the literature sources, have been revised according to the current phylogenetic classifications of Magnoliophyta (Angiosperm Phylogeny Group, 2016), Coleoptera (Löbl & Smetana, 2003, 2004, 2006, 2007, 2008, 2010, 2011, 2013), Aves (del Hoyo & Collar, 2018), Ascomycota (Wijayawardeni et al., 2020) and Myxomycetes (Leontyev et al., 2019). After that, the number of species in each order, family and genus was determined for each taxonomic group. Since invertebrates are represented in our study exclusively by the order Coleoptera, only family and genus spectra were studied for this group. In total, the dataset comprised 2,349 species which occur in the Homilsha Forests, belonging to 1,121 genera, 331 families and 91 orders. The number of taxa included in the analysis for each taxonomic group is shown in Table 1.

The fitdistrplus package for R (https://cran.r-project.org/web/packages/fitdistplus/fitdistplus.pdf) was used to calculate the mathematical characteristics of taxonomic spectra, including range, median, arithmetic mean, standard deviation, kurtosis and skewness. Fitting theoretical distributions to, including Fisher’s log-series, lognormal, gamma, geometric, weibull and broken stick (Mattheus & Whitaker, 2014) to our data was done with sads package for R (https://cran.r-project.org/web/packages/sads/index.html). The completeness of the taxonomic spectra, and, accordingly, their suitability for further analyses was analysed using Chaol estimator with R package vegan (https://cran.r-project.org/web/packages/vegan/vegan.pdf).

Results

The species accumulation curves indicate (Fig. 1) that the number of orders in all taxa found in the Homilsha Forests, except Bryophyta, have reached saturation and merge with the curve of predicted species number, estimated by the Chaol coefficient. The curves for the family spectra have passed the knee and show a tendency to approach the plateau, but remain significantly lower than the values predicted by Chaol. At the genus level curves show considerable steepness; for most taxa their dynamics is close to linear.

Taxonomic spectra of plants, animals, fungi and protists show quite different levels of variability and central values (Table 2). The median number of species was on average 7.1 for orders, 4.1 for families and 1.5 for genera. The standard deviation of the number of species averaged 31.4 for orders, 18.0 for families and 2.3 for genera. The values of the skewness coefficient averaged 2.4 for orders, 2.9 for families and 3.2 for genera, while the value of the kurtosis coefficient was 9.7 for orders, 15.8 for families and 19.8 for families.

### Materials and methods

The data for the taxonomic spectra of the analysed groups were taken from published checklists and annual reports of the Homilsha Forests National Park. The data on following taxa were included: (1) plants: Viridiplanta: Magnoliophyta (Vlashchenko et al., 2010), Bryophyta (Barskov, 2008); (2) animals: Metazoa: Coleoptera (Puchkov et al., 2010; Puchkov, 2018; Skrylnik & Bielavite, 2020), Aves (Vlashchenko et al., 2010); (3) fungi: Agaricomycetes (Prylutskyi et al., 2017); (4) protists: Amoebozoa: Myxomycetes (Prylutskyi et al., 2017). The systematic lists of organisms, taken from the literature sources, have been revised according to the current phylogenetic classifications of Magnoliophyta (Angiosperm Phylogeny Group, 2016), Coleoptera (Löbl & Smetana, 2003, 2004, 2006, 2007, 2008, 2010, 2011, 2013), Aves (del Hoyo & Collar, 2018), Ascomycota (Wijayawardeni et al., 2020) and Myxomycetes (Leontyev et al., 2019). After that, the number of species in each order, family and genus was determined for each taxonomic group. Since invertebrates are represented in our study exclusively by the order Coleoptera, only family and genus spectra were studied for this group. In total, the dataset comprised 2,349 species which occur in the Homilsha Forests, belonging to 1,121 genera, 331 families and 91 orders. The number of taxa included in the analysis for each taxonomic group is shown in Table 1.

The fitdistrplus package for R was used to calculate the mathematical characteristics of taxonomic spectra, including range, median, arithmetic mean, standard deviation, kurtosis and skewness. Fitting theoretical distributions to, including Fisher’s log-series, lognormal, gamma, geometric, weibull and broken stick (Mattheus & Whitaker, 2014) to our data was done with the sads package for R (https://cran.r-project.org/web/packages/sads/index.html). The completeness of the taxonomic spectra, and, accordingly, their suitability for further analyses was assessed using Chaol estimator with R package vegan (https://cran.r-project.org/web/packages/vegan/vegan.pdf).

### Results

The species accumulation curves indicate (Fig. 1) that the number of orders in all taxa found in the Homilsha Forests, except Bryophyta, have reached saturation and merge with the curve of predicted species number, estimated by the Chaol coefficient. The curves for the family spectra have passed the knee and show a tendency to approach the plateau, but remain significantly lower than the values predicted by Chaol. At the genus level curves show considerable steepness; for most taxa their dynamics is close to linear.

Taxonomic spectra of plants, animals, fungi and protists show quite different levels of variability and central values (Table 2). The median number of species was on average 7.1 for orders, 4.1 for families and 1.5 for genera. The standard deviation of the number of species averaged 31.4 for orders, 18.0 for families and 2.3 for genera. The values of the skewness coefficient averaged 2.4 for orders, 2.9 for families and 3.2 for genera, while the value of the kurtosis coefficient was 9.7 for orders, 15.8 for families and 19.8 for families.

### Table 1

| Taxonomic group       | Species | Genera | Families | Orders |
|-----------------------|---------|--------|----------|--------|
| Bryophyta             | 102     | 60     | 30       | 13     |
| Magnoliophyta         | 798     | 256    | 55       | 12     |
| Coleoptera            | 456     | 144    | 59       | 11     |
| Aves                  | 266     | 226    | 76       | 10     |
| Agaricomycetes        | 576     | 263    | 12       | 5      |
| Myxomycetes           | 151     | 34     | 8        | 2      |

### Table 2

| Taxonomic characteristics | Orders | Families | Genera |
|---------------------------|--------|----------|--------|
| Range                     | 43.0   | 6.71     | 10.95  |
| Median                    | 2.0    | 1.46     | 4.0    |
| Mean                      | 7.85   | 2.56     | 4.44   |
| Standard deviation        | 11.57  | 3.0     | 4.44   |
| Skewness                  | 2.65   | 3.48     | 1.76   |
| Kurtosis                  | 10.95  | 3.48     | 7.28   |

Note: Br – Bryophyta, Ma – Magnoliophyta, Co – Coleoptera, Ag – Agaricomycetes, Av – Aves, My – Myxomycetes.
Visual assessment of the rank-abundance plots shows the similarity of studied taxonomic spectra to the HCDs at all levels (Fig. 2). Testing the correspondence of taxonomic spectra to null models according to the Akaike information criterion (AIC) has shown that the spectra of most taxa at all ranks are closely approximated by the log-series distribution model (Table 3). At the same time, in the genera–families–orders row the distribution becomes closer to the lognormal model. It is well seen in Agaricomycetes, Aves, Coleoptera and Magnoliophyta; in the latter, at the order level the distribution is even better approximated by the lognormal distribution than by the log-series one (Fig. 3). All three spectra of Myxomycetes have a peculiar pattern: at the level of genera the distribution is close to geometric, while at the level of families and orders it is best approximated by the broken stick model (Table 3). The geometric distribution is also observed in Bryophyta at the family level (Table 3).

### Table 3
Delta Akaike information criterion values (dAIC) showing correspondence of taxonomic spectra to the distribution models

| Distribution type   | Agaricomycetes | Coleoptera | Aves | Bryophyta | Magnoliophyta | Myxomycetes |
|---------------------|----------------|------------|------|-----------|---------------|-------------|
|                     | O  | F | G | O  | F | G | O  | F | G | O  | F | G | O  | F | G |
| Broken stick        | 16.0 | 16.7 | 216.8 | 44.0 | 216.8 | 14.2 | 26.8 | 105.9 | 8.1 | 7.4 | 45.7 | 0.6 | 76.8 | 314.1 | 0.0* | 0.0* | 6.0 |
| Gamma               | 9.5 | 20.2 | 145.7 | 36.2 | 145.7 | 12.4 | 30.3 | 55.5 | 10.8 | 8.8 | 21.5 | 2.7 | 67.5 | 238.7 | 6.5 | 5.6 | 7.9 |
| Geometric           | 13.6 | 7.6 | 362.2 | 32.4 | 36.2 | 12.0 | 14.1 | 8.9 | 7.2 | 0.0* | 2.7 | 0.7 | 63.3 | 68.2 | 2.4 | 1.7 | 0.0* |
| Lognormal           | 3.5 | 4.8 | 33.7 | 5.5 | 33.7 | 5.0 | 15.3 | 17.8 | 7.7 | 5.0 | 4.3 | 0.0* | 21.0 | 68.0 | 7.3 | 7.3 | 6.0 |
| Log-series          | 0.0* | 0.0* | 0.0* | 0.0* | 0.0* | 0.0* | 0.0* | 0.0* | 0.0* | 0.0* | 0.1 | 0.0* | 1.5 | 0.0* | 0.0* | 5.5 | 3.7 | 2.7 |
| Weibull             | 7.2 | 19.3 | 193.0 | 28.1 | 193.0 | 10.3 | 30.3 | 76.8 | 10.1 | 9.9 | 33.2 | 2.0 | 56.6 | 292.5 | 6.5 | 5.6 | 8.6 |

*Note:* asterisks (*) indicate the best distribution model among tested for the rank and the taxonomic group; O – orders, F – families, G – genera.

Fig. 1. Rarefaction curves of observed (plain line) and estimated by Chao1 (jagged line) number of supraspecific taxa (orders, families and genera) in Hornilsha Forests: a – Agaricomycetes; b – Aves; c – Bryophyta; d – Magnoliophyta; e – Myxomycetes; f – Coleoptera.
Discussion

The drastic changes in taxonomy that have occurred during the recent decades theoretically should have made taxonomic spectra less similar to HDCs distributions, if this similarity was only an artifact of artificial taxon splitting. However, the data we obtained do not confirm this assumption. Despite the use of phylogenetically based classifications in our study, at the levels of genera, and sometimes at the family and order levels as well, the distribution curves have extremely high values of skewness and kurtosis. Such values correspond to the difference between taxa of the same rank by the number of included species in 2.0–2.5 orders of magnitude. We also did not observe any influence of the state of knowledge about the local diversity of the group on the distribution evenness. For example, Aves and Myxomycetes show similar distribution properties, although the first group is one of the most well-studied in the world (Mora et al., 2011), while the second one may include 10 or more undescribed cryptic species within one morphospecies (Shchepin et al., 2016; Shchepin et al., 2019; Lloyd et al., 2019). So we can assume that there are objective factors that help to maintain the hollow-curve distributions of taxonomic spectra. Interestingly, such patterns can be found even at the highest taxonomic levels worldwide. This may explain the phenomenon of micro-kingdoms, i.e. eukaryotic supergroups, whose rank is often equal to kingdom and higher, but the scope does not exceed one or two species (Brown et al., 2018; Lax et al., 2018; Adl et al., 2019).

Fig. 2. Hollow-curve distributions at the levels of genera (a), families (b) and orders (c) for the different taxonomic groups in Homilsha Forests

The comparison of distributions at different ranks show that the indexes of skewness and kurtosis decrease in the genera–families–orders row. In other words, the share of monotypic and other small taxa is higher among genera, than among orders. However, taking into consideration the above-mentioned existence of monotypic superkingdoms, we cannot expect that if the rank of a taxon is higher, the proportion of monotypic subtaxa is smaller. Orphan taxa in most cases are the relictual basal groups, so their presence can be interpreted as a characteristic of the “reluctuality” of the whole group. This characteristic, however, is not identical to the evolutionary age of taxon. Orphan basal groups are known even in relatively young taxa like Aves and Magnoliophyta (Angiosperm Phlogeny Group, 2016; del Hoyo et al., 2018). However, they are not necessarily present in every local biota. This can be the source of the difference in the skewness value between Aves and Magnoliophyta, because the first group is represented in the Homilsha Forests by several orphan orders, and the second is not. Further study is needed to understand, whether there are any general patterns in the distribution of orphan taxa, but it is obvious that relictual groups are often endemics, and this may reduce their representation in local communities, especially outside the world biodiversity hot spots. At the same time, we must take into account the still existing trend to limit the number of higher taxa for practical reasons (Bertrannd et al., 2006). Small higher-rank groups are reluctantly created. This also can contribute to the curvatures reducing in the genera–families–orders row.

Among the studied types of distribution, the log-series model appears the most appropriate for the studied taxonomic spectra. This model is also typical for the species abundance spectra studied at small territories (Antão et al., 2021). Therefore, both taxonomic spectra and species abundance spectra show similar distribution patterns. However, taking into consideration the groups with the distribution different from the log-series model (Myxomycetes, Bryophyta), the limited territory cannot be the sufficient explanation for the patterns, described above. The state of knowledge about the local diversity of the group cannot plausibly explain it either, because the deviations from the log-series distribution do not appear in all ranks of a certain group. Additionally, non-log-series distributions are observed even in groups with very well-studied species composition (Magnoliophyta). Finally, the deviation from the log-series distribution cannot be explained by the conservative taxonomic traditions, because taxa that have shown high compliance with this model, as well as those that deviate from it, are represented by two types of groups, those with a traditionally conservative system of ranks (Aves, Magnoliophyta), and those in which the system of higher taxa has been radically revised on the basis of phylogenetic data in recent decades (Agaricomycetes, Myxomycetes).
Fig. 3. Distribution fitting for taxonomic spectra of Agaricomycetes (a), Aves (b) and Bryophyta (c) at the levels of genera, families and orders: grey columns show the observed number of taxa within abundance classes; coloured lines show estimated distribution models.

Fig. 4. Distribution fitting for taxonomic spectra of Magnoliophyta (a), Myxomycetes (b) and Coleoptera (c) at the levels of genera, families and orders: grey columns show the observed number of taxa within abundance classes; coloured lines show estimated distribution models.
Some properties of taxonomic distributions in the local communities are consistent with the general patterns found in the global biota at different taxonomic ranks. In particular, this is the case of the rarefaction curves approaching the plateau, which get more prominent as the rank increases, as shown on Figure 1 (Mora et al., 2011). This means that phylogenetic relationships at the level of large lineages of the Tree of Life are now successfully reflected in the classification of living things (at least, for the macroorganisms), but the structure of genera in many groups still remains artificial and generally less understood.

Conclusion

Our data suggest the existence of common mathematical properties of local taxonomic spectra that are independent of the taxonomic position or rank of the taxon. At the same time, both factors have an impact on specific quantitative characteristics of distributions. For example, in the orders-families-genres row the predicted level of knowledge about the number of taxa is systematically decreasing. The values of kurtosis and skewness, and the correspondence to the log-series distribution model, also show a variation, but without a clear trend. The hollow-curve distribution of taxonomic spectra is more likely not an artifact of taxonomic research, but reflects some natural factors, first of all the presence of orphan groups, which occupy a basal position in the phylogenetic tree. The kurtosis of the taxonomic spectrum largely depends on the presence of such taxa in the local biota, but the factors that affect their distribution remain unclear. Further studies, involving comparable datasets from different local areas, preferably from different climate zones, may help to understand how universal are the patterns, described here, and how the geographical factors affect the distribution of orphan taxa that probably has a fundamental impact on the structure of taxonomic spectra.

References

Adl, S. M., Bass, D., Lane, C. E., Lukeš, J., Schoch, C. L., Smrakov, A., Agatha, S., Berner, C., Brown, M. W., Burki, F., Cárdenas, P., Cepicka, I., Chistovakova, L., Campo, J., Dartmouth, M., Edvardsen, B., Egli, Y., Guillou, L., Hääm, V., Heiss, A. A., Hopkins, M., James, T. Y., Karpov, S., Kürs, E., Kotášek, M., Kadryavtsev, A., Lahr, D. J., Lena, E., Le Gall, L., Lynn, D. H., Mann, D. G., Mässan, A., Monjuk, R., Mitchell, E. A. D., Morrow, C., Park, J. S., Pavlovski, J. W., Powell, M. J., Richter, D. J., Rueckert, S., Shadwalk, L., Shimo, S., Spiegel, F. W., Trimella, C., Gortz, E., Youssef, N., Zlatogursky, V., & Zhang, Q. (2019). Revisions to the classification, nomenclature, and diversity of Eukaryotes. Journal of Eukaryotic Microbiology, 66(1), 4-119.

Bacaro, G., Ricotta, C., & Mazzoleni, S. (2007). Measuring beta-diversity from fic quantitative characteristics of distributions. For example, in the orders–families–genus row the predicted level of knowledge about the number of taxa is systematically decreasing. The values of kurtosis and skewness, and the correspondence to the log-series distribution model, also show a variation, but without a clear trend. The hollow-curve distribution of taxonomic spectra is more likely not an artifact of taxonomic research, but reflects some natural factors, first of all the presence of orphan groups, which occupy a basal position in the phylogenetic tree. The kurtosis of the taxonomic spectrum largely depends on the presence of such taxa in the local biota, but the factors that affect their distribution remain unclear. Further studies, involving comparable datasets from different local areas, preferably from different climate zones, may help to understand how universal are the patterns, described here, and how the geographical factors affect the distribution of orphan taxa that probably has a fundamental impact on the structure of taxonomic spectra.

Some properties of taxonomic distributions in the local communities are consistent with the general patterns found in the global biota at different taxonomic ranks. In particular, this is the case of the rarefaction curves approaching the plateau, which get more prominent as the rank increases, as shown on Figure 1 (Mora et al., 2011). This means that phylogenetic relationships at the level of large lineages of the Tree of Life are now successfully reflected in the classification of living things (at least, for the macroorganisms), but the structure of genera in many groups still remains artificial and generally less understood.

Conclusion

Our data suggest the existence of common mathematical properties of local taxonomic spectra that are independent of the taxonomic position or rank of the taxon. At the same time, both factors have an impact on specific quantitative characteristics of distributions. For example, in the orders-families-genus row the predicted level of knowledge about the number of taxa is systematically decreasing. The values of kurtosis and skewness, and the correspondence to the log-series distribution model, also show a variation, but without a clear trend. The hollow-curve distribution of taxonomic spectra is more likely not an artifact of taxonomic research, but reflects some natural factors, first of all the presence of orphan groups, which occupy a basal position in the phylogenetic tree. The kurtosis of the taxonomic spectrum largely depends on the presence of such taxa in the local biota, but the factors that affect their distribution remain unclear. Further studies, involving comparable datasets from different local areas, preferably from different climate zones, may help to understand how universal are the patterns, described here, and how the geographical factors affect the distribution of orphan taxa that probably has a fundamental impact on the structure of taxonomic spectra.
