Acta Zoologica Academiae Scientiarum Hungaricae 57 (1), pp. 85–94, 2011

SPECIES ACCUMULATION CURVES AND SIMILARITY TRAITS OF A SPECIES-RICH FLY (DIPTERA) COMMUNITY

IZSÁK, J.[1] and PAPP, L.[2]

[1]*Department of Systematic Zoology and Ecology, Eötvös Loránd University*
H-1117 Budapest, Pázmány Péter sétány 1/A, Hungary; E-mail: ijanos@elte.hu
[2]*Department of Zoology, Hungarian Natural History Museum and*
Animal Ecology Research Group of the Hungarian Academy of Sciences
H-1088 Budapest, Baross utca 13, Hungary; E-mail: lpapp@mail.zoo.nhmus.hu

Here we report some statistical properties of a sample of 92 680 flies collected in Hungarian pastures between 2002 and 2005. We applied descriptive statistical methods, incl. species accumulation curves to describe the relationship between species number and sample effort. Although these curves show an essential slowing down of the increase in the number of fly species, they do not appear to have reached saturation. We discuss the importance of the specification of the sampling effort. To analyse variability conditions of the species structure of the sub-collections, we applied the normalised version of the 'expected species shared' parametric similarity indices. We applied these indices with significantly different parameter values. As known, the parameter influences the sensitivity to changes in small species abundances. The similarity between sub-collections is lower for sub-collections originate from different collection years. With a higher index parameter value the index indicates larger variability in the species structure of the sub-collections. Ignoring advanced ecological reasoning, we offer reference data regarding quantitative traits of other diptera statistics.

Key words: Diptera fauna, neutral community, sampling effort, descriptive statistics, similarity indices

INTRODUCTION

Statistical data and analyses on large samples of dipterans are sporadic. NOVOTNY *et al.* (2005) and CLARKE *et al.* (2004) performed analyses of species accumulation curves of tephritid flies, but based on presence-absence relations. Formerly we analysed the species diversity of drosophilids collected on apple-baits (IZSÁK & PAPP 1994) and in brook valleys (IZSÁK & PAPP 1995), species number estimations and species abundance distribution relating to Diptera collected on sheep-run droppings (PAPP *et al.* 1997) and similarity conditions of those samples (PAPP & IZSÁK 1999). As for the distribution of fly species abundances, LAURENCE (1955) found a poor fit of both the lognormal and logseries distributions in cow pat fly collections of the *Limosina* species, probably as a consequence of low total abundances and because the flies were captured during a single year only. ARGEMÍ *et al.*(1999) reported a poor fit of log-series distribution to abundances in a collection of Drosophilidae.

In the present publication we report on the one hand data on species accumulation curves (SACs) relating to a large sample of dipterans from cow pats in Hungarian pastures between 2002 and 2005. In the multi-species community the species populations were living in the same area and habitat and in the same time period (see e.g. MAGURRAN 2004); in addition, the relations between the co-occurring species can be neglected (cf. PAPP 1971); that is, in our multi-species community the relationship between species can be considered neutral. Although this condition does not assure that the community itself is neutral, and does not play a role in the discussion of the results, it may play a role in further considerations. Faunistic results yielded by the collections have already been published elsewhere (PAPP 2007). We do not go into advanced ecological discussion regarding the finding about the SACs. At the same time the elementary traits of these curves and the sample variability (see below) can serve as valuable faunistic and methodological data when evaluating similar faunistic works.

SACs may inform on the relationship between sampling effort and number of registered species. According to the type of the sampling effort, there are two kinds of the accumulation curves (COLWELL *et al.* 2004). Sample-based SACs show the dynamics of the increase of species number in the sample by increasing the sampling effort, taking the latter in a strict sense. In this case, sampling effort can be identified by the length of observation time or the size of the sample area; specifically the number of sample quadrates or the number of the groups of tens (see below). Another possibility is to construct an individual-based SAC in which case one identifies the unit sampling effort with adding a fix number of new individuals to the sample. Motivated by the possibility to predict the total (including the unknown) number of species in the community, several SAC models have been published (SCHEINER 2003, COLWELL *et al*. 2004). NOVOTNY *et al*. (2005) presented a sample-based SAC related to fruit fly (Diptera: Tephritidae) collections with sample effort quantified as the weight of the fruits; furthermore, they also calculated individual-based SAC. CLARKE *et al.* (2004) also provided examples of SACs applied to fruit fly communities. PETERSEN & MEIER (2003) estimated species-richness by analysing large sample of Diptera, collection sizes were tallied from the pinned collections preserved in museums.

Here we construct SACs to characterise the increase of the species number by adding new sub-samples to the collection. Postulating the homogeneity of the population in time and space, the curves can be taken for quasi individual-based SACs. At all, in the present case the SACs should be considered sample based. The published SACs make possible to compare the richness parameter of the present collection and other collections at different sample sizes.

To get information about the variability of the species structure of some sub-collections we carried out similarity studies. Concretely, we applied two measures to investigate the difference in the species structure of sub-collections originating from the same and from different years. Another aim was to detect and demonstrate in the present collection of Diptera the larger structural difference between sub-collections originating from different annual years. We presumed that the effect of the role of collection year factor is more clearly expressed when using a similarity/dissimilarity measure more sensitive to changes in the small species frequencies (those of the rare species). All these can one relate to similar findings on other insect collections. We think that both the properties of the SACs and the similarity traits can one recognise only from concrete Diptera statistics even if advanced theories are known on these issues in general.

MATERIALS AND METHODS

Flies were collected between 2002 and 2005 on cow pats in pastures near Nagyiván (Új-kút and Mérges-kút, in a distance of less than 5 km from each other), a village located in the Hortobágy National Park (HNP) in Hungary. These fields have been grazed continuously through the past decades; a period sufficiently long for all potential fly species to invade this area and reach a relative stability. Taking into account our and others' previous experience with diurnal and seasonal variations of population densities (ARGEMÍ *et al.* 1999, ARGEMÍ *et al.* 2000), we collected flies from the middle of July to the middle of August between 10 and 16 hours. The cow pats were dropped ca. 12–24 hours before the collection of flies. The combination of samples collected through some weeks compensates the fluctuations and successional changes within a few days or weeks. Further on, a combination of within a year samples can reduce variations due to large-scale meteorological and similar yearly changes. For details on the sampling device and sampling method see PAPP (2007). We must stress also here, that the individual adult fly samples are not results of estimations but counting. The rough data are available at http://www.nhmus.hu/~lpapp. The annual numbers of individuals and species are summarised in Table 1.

We identified a total of 92,680 individuals representing 106 species, indicating a considerably species-rich community. According to our previous studies, the species pool of fly species inhabiting the pasture of HNP may consist of 270–280 species belonging to 26 dipterous families (PAPP 1992). The species *Coproica lugubris* was super-dominant in all collections. The fly community represents a single guild, as the species concerned are characteristic of the cow pats. We did not separate species with dung eating larvae from those having predatory larvae. Similarly, species that have larvae developing in habitats other than cow pats were not excluded, provided that the adults occur on cow pats

Table 1. Basic data of the collections.

	2002	2003	2004	2005	2003–2004	2003–2005
Total of individuals	1,146	21,009	29,525	41,000	50,534	91,534
Species	37	80	80	83	99	105
Elementary samples	15	400	300	300	700	1,000

frequently. For types of insect specificity to habitat patches and food resources see e.g. NOVOTNÝ and BASSET (2000).

In order to gain sufficiently detailed data, we registered the species and their frequencies in small sub-collections. These are the so-called ‘elementary samples’. These are meant to be the sets of individuals captured on a single day and single cow pat or, in a few cases, two small pats. The number of elementary samples in the collections is given in Table 1. Elementary samples were arranged randomly within a year. The series of the elementary samples taken between 2003 and 2005 were randomly combined to form 100 ‘groups of ten’. All these samples consist of ten elementary samples. These groups of ten were used in the further analysis. We have to acknowledge that the number of individuals in the elementary groups and consequently also in the groups of ten differed considerably.

SACs were constructed separately for the years 2003, 2004, 2005, and for the 2003–2005 samples combined. The separate treatment of samples makes possible to explore the variability of findings. On the other hand, through the combination of the samples we gained an approximately total scenery of the occurring flies. The unit increase in sampling effort was the addition of a new group of ten to the sample. Presuming that the groups of ten in a particular year contain roughly similar number of individuals, the sample-based sampling effort corresponds to a kind of quasi individual-based sampling effort. However, when choosing another annual year, the mean number of individuals in the samples of ten may differ markedly.

To make the SACs of different years comparable, we applied a simple method. Let us denote the number of groups of ten and consequently the first co-ordinate of the endpoint of SAC in case of collection A by g_A. Let us denote the number of groups of ten in collection B by g_B. (This latter coincides with the first coordinate, g_B of the endpoint of the SAC belonging to B.) Then, the first transformed co-ordinate g'_A of the endpoint of the SAC belonging to B should be defined by the equation $n_B / g'_B = n_A / g_A$, that is $g'_B = n_B \ / (n_A / g_A)$, where n_A and n_B is the number of individuals in A and B. The interval $(0,g'_B)$ will be divided into approximately g'_B equal sized intervals, the endpoint of each being a first co-ordinate of a point of the transformed SAC belonging to collection B. Taking the 2003 curve for the basis, we obtained the curves applying this procedure in Fig. 3. Using a similar procedure one can obtain a transformed SAC belonging to the 2003–2005 collection.

The choice between time-based, area-based or individual-based SAC depends on the approach. It is noteworthy that some relations, such as the intersection of the curves depend on this transformation.

To reveal similarity characteristics, we combined the first 50 randomly arranged elementary samples, the second 50 elementary samples and so on, gaining finally a total of 20 combined samples, each containing a randomised set of 50 elementary samples. These were the ‘groups of 50’. We studied the similarities among these groups (see below). The applied parametric similarity indices NESS(m) with possible parameters $m = 1,2, \ldots$ are members of the normalised ESS(m) index family. The formula of the latter index family, denoting also the populations concerned is

$$ESS(m,A,B) = \sum_i (1-(1p_{iA})^m)(1-(1p_{iB})^m),$$

where p_{iA} and p_{iB} is the occurrence probability of the ith species in population A and B.

The normalising factor is $(ESS(m,A,A) + ESS(m,B,B)) / 2$. The ESS($m$) (expected species shared) index expresses the expected number of common species in two random sets of m individuals selected randomly and with replacement from each of the two collections to be compared (SMITH & GRASSLE 1977, SMITH 1989). NESS(1) is identical to the Morisita index. A larger m makes the index more sensitive to abundance differences in the range of small abundances. Jackknifed estimates J(NESS(m)) were computed by a subroutine in the program DIVERSI2.2 (IZSÁK 2005), freely available from the first author. The estimation formula allows all positive integer m parameters not exceeding the half size of the smaller collection. We used NESS similarity indices previously by analyses on Diptera collections (PAPP & IZSÁK 1999). For investigating between-site similarity of fruit fly faunas we used NESS(20) index (see WOLDA 1996).

RESULTS AND DISCUSSION

Species accumulation

Sample-based SACs for three collection years are shown in Figure 1. One can establish significant differences among the SACs in the three years. Due to the lower density of individuals, the 2003 curve is less steep as compared to the other ones. Characteristically, the 2003 and 2004 curves are intersecting.

One can observe an essential slowing down of the increase in species number after the accumulation of about 15 groups of ten. That is, in the studied years a further increase in collection size would have not resulted in an essential widening of the species set already known.

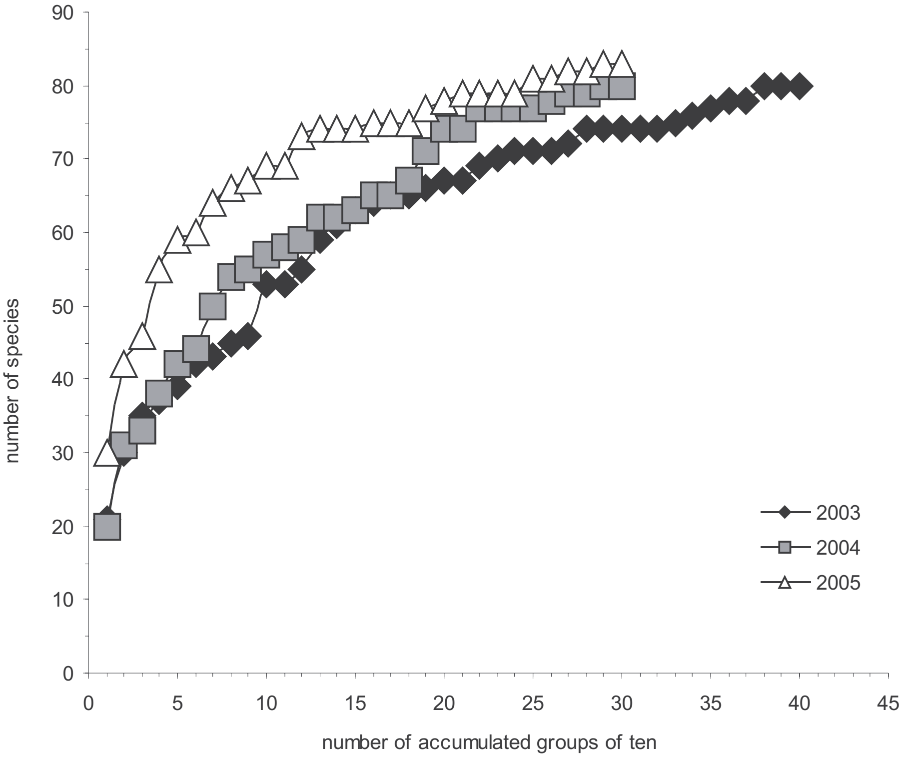

Fig. 1. Sample-based species accumulation curves without normalisation

The SAC representing the 2003–2005 collections combined is shown in Figure 2. The first part of this curve is identical with the 2003 curve, see Figure 1, however, this curve demonstrates a rather continuous increase of species with further increase of the sampling effort. It is noteworthy that arriving at the first group of ten in 2004 and 2005, respectively, the increase in species number did not gain new impetus. At the same time, regarding the final part of the curve, the increase in the number of species does not seem to come to an end in a reasonable time.

In Figure 3 the quasi individual-based SACs are shown. By inspecting these transformed curves, we may make some observations. The SACs belonging to collection 2003 and 2005 are in a long run very similar.

Despite the normalisation of the sampling effort, after a period the SAC of 2004 indicates a relatively slower increase in the number of species. Moreover, the characteristic crossing of the SACs of 2003 and 2004 is missing with the transformed SACs due to the considerably steeper increase of the 2003 curve in this case.

Similarity of collections

In Figure 4 similarities between the first and *k*th group of 50, $k = 1,2,\ldots, 20$ measured by the J(NESS(1)) and J(NESS(100)) index are plotted against *k*. It

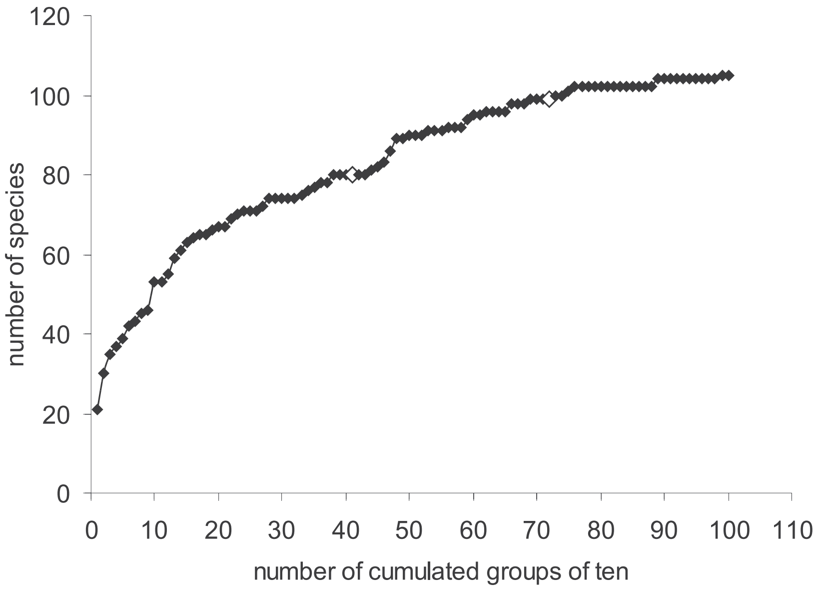

Fig. 2. Sample-based curve for 2003–2005. Two points, corresponding to species numbers after adding the first group of ten to the sample in 2004 and 2005 are denoted by empty symbol

should be noted that similarity is in a sense the opposite of variability. Note that k is a dummy variable. This means that taking into account the random arrangement of the elementary samples in a single year the serial number k of the groups of 50 in a fix year is of no importance. Then, k informs only on the origin of the second group from any of the 2003, 2004, 2005 collections. This is determined essentially by the relations $k \leq 8$, $9 \leq k \leq 14$, $15 \leq k$. The results show that a significant decrease of NESS(100) similarity can be observed when the 'first' group of 50 in 2003 is compared with a group originating from the 2004 collection. A further decrease in similarity can be detected by comparing the 'first' group with those from 2005. It is instructive that the frequently used NESS(1) index, which measures almost exclusively abundance differences of the dominant and subdominant species, does not indicate a significant change in the structural similarity.

Finally, we compared the neighbouring groups of 50. We compared the first group of 50 with the second one, and in general, the kth group of 50 with the $(k + 1)$th one, $k = 1,2, \ldots, 19$. The computed similarity values are plotted against k in Figure 5. Taking into account that groups of 50 in a year form a random series of possible groups of 50, one can expect that structural similarity decreases only when any of the groups of 50 in a certain year is compared to any one group from the next year. This condition is fulfilled for $k = 8$ and $k = 14$.

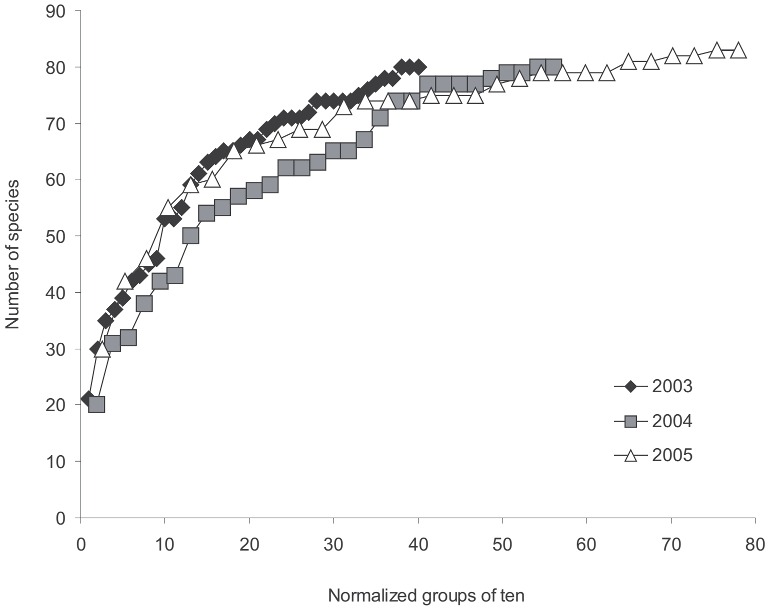

Fig. 3. Quasi individual-based species accumulation curves after normalisation

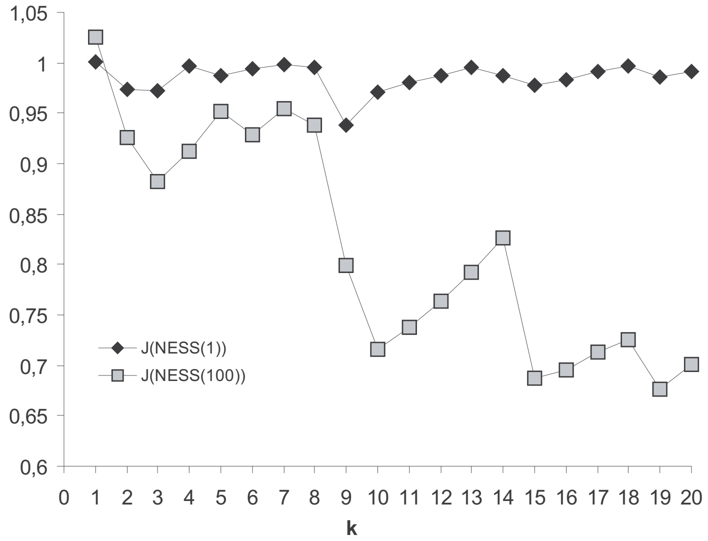

Fig. 4. Jackknifed NESS indices relating to the first and *k*th group of 50, $k = 1,2,\ldots, 20$

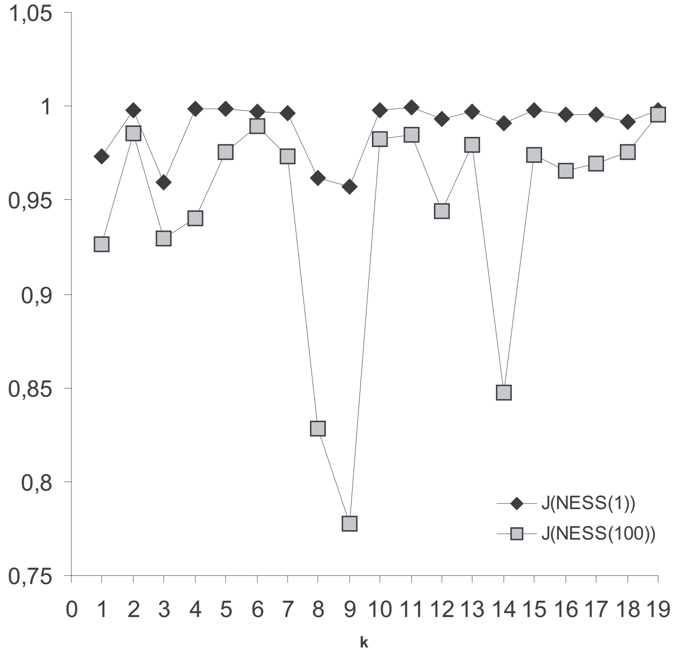

Fig. 5. Jackknifed NESS indices relating to the *k*th and $(k + 1)$th group of 50, $k = 1,2,\ldots, 19$. In case of $k = 8$ and $k = 14$ samples from different years are compared

The 8th similarity value expresses the similarity between the 'last' sample of 50 in 2003 and the 'first' one in 2004. The 14th similarity value expresses the similarity between the 'last' sample of 50 in 2004 and the 'first' one in 2005. Indeed, both similarity values are relatively small. This can be observed principally with the NESS(100) index. The small similarity between the 'first' and 'second' group of 50 in 2004 (see the similarity at $k = 9$) is an exception. According to the results the parametric NESS indices are applicable to analyse the similarity between entomological collections in more details.

In summary, we collected flies from a multi-species fly community in the period 2002–2005 in Hungary. To obtain some basic statistical information on the collections, we constructed SACs and performed similarity studies on sub-collections. We established a considerable variation of SACs in years 2003, 2004 and 2005. The SAC belonging to the largest combined sample suggests that the increase in the number of species does not cease even if the collection size is as large as 92 680 individuals. Moreover, we detected the differences in the SACs both using the original unit-based and quasi individual-based SACs. The applied parametric similarity indices indicate a smaller similarity of sub-collections originating from different years. This applies primarily to the similarity index NESS(100), that is considerably sensitive to small changes in rare species abundances.

*

Acknowledgements – The study was supported by the Hungarian Scientific Research Fund (OTKA), No. T 043168 for JI and LP. Our thanks are due to Mrs ANIKÓ KOROKNAI for her assistance in informatics. We are grateful to Mrs ZSUZSÁNNA BAJZA-PAPP for her help with the sample collection and to the assistants of the Diptera Collection, HNHM, Mrs JUDIT GALAMBOS, Mrs ZSUZSA PETROVICS and Mrs ESZTER ZAHORÁNSZKY-VIDA for their careful and precise work in mounting, labelling, etc.

REFERENCES

ARGEMÍ, M., MONCLÚS, M., MESTRES, F. & SERRA, L. (1999) Comparative analysis of a community of drosophilids (Drosophilidae; Diptera) sampled in two periods widely separated in time. *Journal of Zoological Systematics and Evolutionary Research* **37**: 203–210.

ARGEMÍ, M., BÄCHLI, G., MESTRES, F. & SERRA, L. (2000) Analysis of the diurnal activity patterns of six species of Drosophila (Diptera, Drosophilidae) at Bordils (North East of Spain). *Mitteilungen der schweizerischen entomologischen Gesellschaft* **73**: 337–348.

CLARKE, A. R., BALAGAWI, S., CLIFFORD, B., DREW, R. A. I., LEBLANC, L., MARARUAI, A., MCGUIRE, D., PUTULAN, D., ROMIG, T., SAR, S. & TENAKANAI, D. (2004) Distribution and biogeography of Bactrocera and Dacus species (Diptera: Tephritidae). *Australian Journal of Entomology* **43**: 148–156.

COLWELL, R. K., MAO, CH. X. & CHANG, J. (2004) Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology* **85**: 2717–2727.

IZSÁK, J. (2005) Diversity and related statistics: a new version of the DIVERSI program. *Community Ecology* **6**: 249–252.

IZSÁK, J. & PAPP, L. (1994) Numerical properties of jackknifed diversity indices tested on loose sets of coenological samples (Diptera, Drosophilidae). *Coenoses* **9**: 59–67.

IZSÁK, J. & PAPP, L. (1995) Application of the quadratic entropy indices for diversity studies of drosophilid assemblages. *Environmental and Ecological Statistics* **2**: 213–224.

LAURENCE, B. R. (1955) The ecology of some British Sphaeroceridae (Borboridae, Diptera). *Journal of Animal Ecology* **24**: 187–199.

MAGURRAN, A. E. (2004) *Measuring biological diversity.* Blackwell Science, Oxford.

NOVOTNÝ, V. & BASSET, Y. (2000) Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. *Oikos* **89**: 564–572.

NOVOTNY, V., CLARKE, A. R., DREW, R. A. I., BALAGAWI, S. & CLIFFORD, B. (2005) Host specialization and species richness of fruit flies (Diptera: Tephritidae) in a New Guinea rain forest. *Journal of Tropical Ecology* **21**: 67–77.

PAPP, L. (1971) Ecological and production biological data on the significance of flies breeding in cattle droppings. *Acta Zoologica Academiae Scientiarum Hungaricae* **17**: 91–105.

PAPP, L. (1992) Fly communities in pasture dung: some results and problems (Diptera). *Acta Zoologica Academiae Scientiarum Hungaricae* **38**: 75–88.

PAPP, L. (2007) A study of the cow pat Diptera on the Hortobágy, Hungary. *Folia entomologica hungarica* **68**: 123–135.

PAPP, L., IZSÁK, J. & ÁDÁM, L. (1997) Dipterous assemblages of sheep-run droppings: number of species observed, estimated and generated by simulation. *Acta Zoologica Academiae Scientiarum Hungaricae* **43**: 191–205.

PAPP, L. & IZSÁK, J. (1999) The 'expected species shared' measure to compare composition of dipterous aerial plankton and sheep dung fly communities. *Acta Zoologica Academiae Scientiarum Hungaricae* **45**: 115–123.

PETERSEN, F. T. & MEIER, R. (2003) Testing species-richness estimation methods on single-sample collection data using the Danish Diptera. *Biodiversity and Conservation* **12**: 667–686.

SCHEINER, S. M. (2003) Six types of species-area curves. *Global Ecology and Biogeography* **12**: 441–447.

SMITH, W. & GRASSLE, J. F. (1977) Sampling properties of a family of diversity measures. *Biometrics* **33**: 283–292.

SMITH, W. (1989) ANOVA-like similarity analysis using expected species shared. *Biometrics* **45**: 873–881.

WOLDA, H. (1996) Between-site similarity in species composition of a number of Panamian insect groups. *Miscellania Zoològica* **19**: 39–50.

Revised version received June2, 2010, accepted September 10, 2010, published February 28, 2011