

ON THE NEED TO RENEW THE TAXONOMIC SYSTEM OF THE CANDONINAE (NON-MARINE OSTRACODA, CRUSTACEA). REFLEXIONS FROM AN ANALYSIS OF DATA USING THE YULE PROCESS

DAN L. DANIELOPOL⁽¹⁾, ANGEL BALTANÁS⁽²⁾, URSULA MOROCUTTI^(3,4), FERDINAND ÖSTERREICHER⁽³⁾

⁽¹⁾Commission for the Stratigraphical and Palaeontological Research of Austria, Austrian Academy of Sciences c/o Institute of Earth Sciences (Geology and Palaeontology), Karl-Franzens-Universität Graz, Heinrichstrasse 26, A-8010 – Graz, Austria, e-mail: ddanielo@oeaw.ac.at

⁽²⁾Department of Ecology, Universidad Autónoma de Madrid, E-28049 – Madrid, Spain, e-mail: angel.baltanas@uam.es

⁽³⁾Institute of Mathematics, Universität Salzburg, A-5020 – Salzburg, Austria, e-mail: ferdinand.oesterreicher@sbg.ac.at

⁽⁴⁾Present: Am Steingarten 6, A-4970 – Eitzing (OÖ), Austria, e-mail: ursulam@gmx.at

Abstract. The history of life is the chronicle of diversification. Each taxonomic group originates from a single ancestor and diversifies into a number of species. The evolutionary process, with its two major components (speciation and extinction), is the force working behind. Several models, i.e. explicit mathematical statements of a hypothesis, have been developed in order to inspect the likely output of the evolutionary process and to compare it with actual diversity patterns observed in Nature. The British statistician George Udny Yule produced one of the earliest attempts of such models, now known as the Yule Process, which leads to the Yule-Simon Distribution. His model and the aforementioned probability distribution belong, as well as the Zeta Distribution used in this contribution, to the large family of Power Laws. It focuses on the distribution of the size (= number of species) of genera within a taxonomic group. Here, the data of the non-marine ostracod family Candoninae are discussed within the framework of the Yule Process. Beyond explaining the mathematical basis of the model and how to estimate the parameter of both distributions, the fit or lack-of-fit of the Yule Process (the latter used as a neutral model) is further considered as a test for the suitability of the taxonomy of the group currently accepted.

Biological/evolutionary factors (like the existence of large endemic flocks of species in ancient lakes) combined with socio-psychological factors (like the principle of Authority) are here reviewed as potential sources of taxonomic misrepresentation. Special attention is devoted to monotypic genera within the Candoninae; their relevance for understanding current taxonomic patterns is largely discussed. The utility of the Yule Process, namely, the Yule-Simon Distribution, as reference background against which to compare taxonomic patterns in the Candoninae, opens new possibilities for the analysis of the evolutionary and epistemological problems of ostracod studies.

It is concluded, that the taxonomic system currently accepted for that subfamily must be renewed including: the re-examination of the diagnostic criteria used a restatement of the taxonomic meaning of the species-rich lineages endemic to ancient lakes, the increased cooperation between neontologists and palaeontologists to produce common-based taxonomic criteria, the addition of new sources of taxonomic information (morphometric, molecular, etc.) and the adoption of ideas and/or methods specific to phylogenetic systematics.

Key words: Ostracoda, Candoninae, Taxonomy, Yule Process, Yule-Simon Distribution, Zeta Distribution

1. INTRODUCTION

The present contribution is a study of the taxonomic diversity of the most species-rich subfamily within ostracods occurring in non-marine aquatic and semi-terrestrial habitats, the Candoninae (Crustacea, Ostracoda). This huge diversity can be appreciated from the information Martens *et al.* (2008) published as an extended review on the world Recent

ostracods which live in non-marine environments. The authors mentioned that, until the present day, about 2000 species, included in approximately 200 genera, are documented. From this large number of species, about 25 % belong to the family Candonidae.

The aim of the present project and of our enquiry is two-fold. First, we analyse the relationship of diversity to taxonomic structure in the group; *i.e.* we assess the frequency

distribution of lower rank units (species) into higher rank units (genera), to describe them in the light of several evolution-based models. Second, we explore the likely causes that might explain whether those diversity patterns truly reflect the evolutionary history of the group, or are somewhat distorted by current and past taxonomic practices.

The problem of the comparative documentation of organismic diversity represented during the last decades a major topic in evolutionary ecology and palaeoecology. The rich literature issued is *inter alia* reviewed by Cockburn (1991) and/or in the volume edited by Allmon and Bottjer (2001). An important aspect of the way organismic diversity was documented is also the analysis of genera and/or species abundances published in various compendia and processed with mathematical algorithms and computer simulations. The reliability of such documentation depends *inter alia* also on the various statistical methods proposed for computation. One of us (A. B.) discussed some of the methods used for species richness estimators (Baltanás, 1992). Here, we will expand this interest by discussing alternative hypotheses on the origin of the unequal distribution of species within various genera belonging to the subfamily Candoninae. We start from the idea that the current knowledge of the group's taxonomic diversity has a dual origin; actual biological processes which lead to the present evolutionary pattern, but also the way ostracodologists perceived the morphological differences between ostracod taxa and how they described and placed the taxa in a taxonomic system. This approach is appealing because it compels us to identify the kind of biological processes which participate in the development of the present diversity of genera and their species, in the first place; second, to appreciate the importance of the socio-psychological impact of the specialists who contributed to the present-day taxonomic system in use for Recent Candoninae.

For the development of our project, we use the already well-established tradition of analysing and/or comparing empirical data with information derived from evolutionary-based mathematical models. Many different groups of organisms display a similar taxonomic structure with few genera rich in species and a high number of genera including only one species or two. Such a diversity pattern is not restricted to organismic assemblages, but also happens in non-biological collections of events/objects, from earthquakes to people's personal fortunes (*cf.* Newman, 2005). We call such distributions "long-tailed distributions" and they can be described mathematically by so-called Power Laws. Accordingly, we will first check if the taxonomic diversity of the Candoninae fit a power-law distribution by comparing empirical data with the Zeta and the Yule-Simon Distributions. The Candoninae could follow a Power Law because, as we will see, they are characterised, not only by few genera with a high number of species, but also by a high frequency of monotypic genera. Beside the mathematical demonstration of the stochastic process that generates such a Power Law, there are also

a series of biological hypotheses that explain their empirical occurrence in nature. Therefore, it is interesting to compare the taxonomic diversity of Candoninae data at the lower level of genera and species with a mathematical model (we will specially deal with the Yule Process (or Yule algorithm) which leads to a so-called Yule-Simon Distribution), because it helps to find answers to both biological and socio-psychological questions we address. Furthermore, it will help us to suggest the way the taxonomic system of Candoninae should be renewed.

2. MATERIALS AND METHODS

2.1. OSTRACOD DATA

For the present investigation we use three sets of data:

(1) 293 Candoninae species grouped in 34 genera (*cf.* Table 1) that Baltanás and Danielopol (B&D) prepared for their study on the valve diversity of Candoninae (MS in preparation). We use these data for the demonstration of the way we compute Zeta and Yule-Simon Distributions. They were first used by Morocutti (2010) in her Masters Thesis and they should be seen here as a subset of the data published by Martens and Savatnalinton (2011) (see below).

(2) 512 Candoninae species grouped in 37 genera (*cf.* Table 2) extracted from the world data list of non-marine ostracods published by Martens and Savatnalinton (2011). These form the core data on which we discuss the topics mentioned in the introduction.

(3) For comparative purposes, we use also from the taxonomic list of Martens and Savatnalinton (2011) information on the other two subfamilies of the Candonidae, namely, the S. F. Cyclopyridinae, with 122 species distributed in 6 genera, and the S. F. Paracypridinae, with 28 species belonging to 8 genera (*cf.* Tables 4 and 5).

2.2. METHODS: POWER LAWS, WITH SPECIAL EMPHASIS ON ZETA AND YULE – SIMON DISTRIBUTIONS

Power Laws are statistical relationships between two variables, one increasing as a power of some attribute of the other variable ($Y = a \cdot X^b$). Power Laws are well known by morphologists and evolutionary biologists since Huxley (1932) who described allometry as the pattern of covariation among morphological traits; although it soon became evident that Power Laws also fit relationships involving ecological and physiological traits (Peters, 1983; Brown and West, 2000; Marquet *et al.*, 2007). For the special case of the relationship between genera and the number of species, Nee *et al.* (1996) write the general power-law expression as $G = A \cdot N^{-d}$, where G is the number of genera with N species, A is a constant and the exponent d of the Power Law has a fixed dimension, generally larger than 1. Moreover, we have to point out that mathematically a Power Law is a probability distribution defined on the set \mathbb{N} of natural numbers, so that the probability $P_\alpha(k)$ have the exact or approximate form

$$P_{\alpha}(k) \approx \frac{c(\alpha)}{k^{\alpha}}, \quad k \in \mathbb{N},$$

where α is a positive parameter and $\alpha \rightarrow c(\alpha)$ a positive function, such that the series

$$\sum_{k=1}^{\infty} P_{\alpha}(k) = 1$$

For this equality to be satisfied, the typical parameter is $\alpha > 1$.

As mentioned above, in this contribution we consider two special families of Power Laws, namely the family of Zeta Distributions and the family of Yule-Simon Distributions.

Definition 1: A probability distribution of the form

$$P_{\alpha}(k) = \frac{1}{\zeta(\rho+1) \cdot k^{\rho+1}}, \quad k \in \mathbb{N}$$

is called Zeta Distribution with parameter $\rho \in (0, \infty)$. The function $\zeta : (0, \infty) \rightarrow (0, \infty)$ is the Riemann - Zeta function.

Remark: The Zipf Distribution which has the form

$$P_i(k) = \begin{cases} \frac{c}{k^1} & \text{for } k \in \{1, 2, \dots, n\} \\ 0 & \text{for } k > n \end{cases}$$

for a fixed number n and c such that

$$\sum_{k=1}^n \frac{c}{k^1} = 1$$

is a member of a "Power Law" – perhaps its most well-known one – but obviously is not a Zeta Distribution in a strict sense.

Definition 2: A probability distribution of the form

$$P_{\rho}(k) = \frac{\rho \cdot \Gamma(\rho+1) \Gamma(k)}{\Gamma(\rho+1+k)} = \frac{\rho \cdot (k-1)!}{(\rho+1) \cdot \dots \cdot (\rho+k)}, \quad k \in \mathbb{N}$$

is called Yule-Simon Distribution with parameter $\rho \in (0, \infty)$. Γ function: $(0, \infty) \rightarrow (0, \infty)$ is Euler's Gamma-function.

Statistics: For our applications we assume that the species are ordered in such a way that species 1 has the highest number of genera, species 2 the second-most, ..., etc. (in case of ties their order is irrelevant). Let us assume that we have n species and let x_i be the number of genera of species i , $i \in \{1, \dots, n\}$. Furthermore, let

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{and} \quad \tilde{x}_n = \sqrt[n]{\prod_{i=1}^n x_i}$$

be the arithmetic mean and the geometric mean of the frequencies x_i , $i \in \{1, \dots, n\}$, respectively.

In case of a Yule-Simon Distribution the appropriate estimate for the parameter ρ is

$$\hat{\rho}_n = 1 + \frac{1}{\bar{x}_n - 1}$$

In case of a Zeta-Distribution the appropriate estimate for the parameter ρ is

$$\frac{-\zeta'(\rho+1)}{\zeta(\rho+1)} = \ln \tilde{x}_n$$

This solution can be obtained by linear interpolation from a suitable statistical table, e.g. from Johnson *et al.* (1992, page 468).

The mathematical distribution which comes out from the algorithm developed by Yule (1925) was later rediscovered by Simon (1955). This latter author showed in detail the stochastic process of the model and called it also Yule Distribution and/or Yule Process. Simon (*op. cit.*) pointed to the possible applications of this statistic distribution in various domains of activity, like linguistics, sociology and urbanism. A series of publications deals nowadays with this topic and their authors decided to honour also H. Simon for the mathematical contribution to this algorithm and its distribution, therefore it is called frequently Yule-Simon Distribution. Therefore, as an example, one should see Tambovtsev and Martindale (2007), who deal with such a distribution of phonemes in linguistics, and/or Tonelli *et al.* (2010), who propose improved solutions to the original Yule algorithm.

The origin of Yule's Process is based on Willis and Yule (1922) attempt to explain the evolution of organismal diversity through time and space in a predictable way. Their biological model assumes that speciation occurs independently and at random (homogeneous pure birth process) with no extinction explicitly included in it. Yule (1925) produced a detailed description of his mathematical model showing that genera will increase their number of species with time in geometric proportion to their previous values at stochastically constant time rates. Correspondingly, the speciose groups will become more and more species-rich with time. Additionally, monotypic genera will also occur with a given fixed probability (Reed and Hughes, 2002). The output of this stochastic process presupposes that old phylogenetic lineages will be represented nowadays by genera with high number of species, while those genera with fewer species represent young evolutionary groups. In a metaphoric way, this is the type of process where "the rich gets richer" (Newman, 2005). As mentioned above, another property of the Yule Process is the long-tail distribution of its entities (in the case of taxonomic groups entities are genera containing species as elements *cf.* Tonelli *et al.*, 2010). Therefore, the distribution curve of entities is graphically represented by "a hollow curve" with two prominent sections, the "head" with few dominant (i.e. markedly species rich) genera and the "tail" with many monotypic genera (i.e. represented by a single species).

The evolutionary theory using the Yule Process, known also under the name “Age and Area”, following the name of the book published by J.C. Willis (1922), stimulated at that time a strong debate (*cf. inter alia* Pearson, 1923; Gleason, 1924; Willis, 1926). It received little confirmatory support later on (*cf.* Wright, 1941; Miller, 1997). Nowadays, this statistical distribution continues to be discussed within the context of evolutionary biological topics, like those dealing with the reconstruction of molecular phylogenies (Nee *et al.* 1992; Aldous, 2001; Aldous *et al.*, 2008) or with the origin of adaptive radiations (Freckleton and Harvey; 2006; Simkin and Roychowdhury, 2011). In these latter publications, scientists discussed the Yule Process as a referential model from which data can be tested and, eventually, falsified, hence offering the basis for alternative explanations of empirical data or for proposing new theoretical models.

3. RESULTS

3.1. THE TRIAL DATA-SET: FITTING ACTUAL DATA TO THE ZETA AND THE YULE-SIMON DISTRIBUTIONS

Table 1 presents the list of the 34 genera and 293 species of Candoninae belonging to the B&D data set. One should note that some subspecies of already described Candoninae species were for this investigation elevated at species rank. Alternatively, the subgenus rank was not retained in the present study.

For this approach, we plotted the Candoninae genera (Fig. 1) in decreasing ranks, following the number of species per genus. We expressed further the same data as a dominance curve where each genus is represented by the percentage from the total number of species (Fig. 2).

A “hollow curve” is obtained with a left branch more or less in a vertical position (“the head”), due to the strong dominance of the first five genera each of them having ≥ 14 species collecting 67% from the total ostracod set (Table 1). A long horizontal branch (“the tail”) is formed by many monotypic genera. This latter class represents the modal generic group with the highest frequency (13 genera from 34), characteristic for a Yule-Simon Process (*cf.* Yule, 1925). The same data are plotted on a double logarithmic scale (Fig. 3) showing that the genera lay closely to the linear regression line with a negative slope of 1.51.

The coefficient of determination r^2 obtained from the linear regression between the dominance of genera (expressed as number of species) and their rank order is 0.96. This latter information corroborates the idea that the data follow a power-law distribution following the rule of thumb proposed by Newman (2005). In order to verify if our Candoninae set really conforms to a power-law distribution we apply below a Goodness of fit test for adjusting the data to a Zeta Distribution.

Table 1 The trial data set of the S. F. Candoninae

S-F. Candoninae Kaufmann 1900		
Rank	Genus Name	N. Species
1	<i>Candona</i> Baird, 1845	80
2	<i>Pseudocandona</i> Kaufmann, 1900	51
3	<i>Areacandona</i> Karanovic, 2005	25
4	<i>Fabaeformiscandona</i> Krstic, 1972	23
5	<i>Candonopsis</i> Vávra, 1891	17
6	<i>Baicalocandona</i> Mazepova, 1976	13
7	<i>Mixtacandona</i> Klie, 1938	13
8	<i>Humphreyscandona</i> Karanovic and Marmonier, 2003	10
9	<i>Leicacandona</i> Karanovic, 2007	8
10	<i>Cryptocandona</i> Kaufmann, 1900	7
11	<i>Deminutiocandona</i> Karanovic, 2003	7
12	<i>Pilbaracandona</i> Karanovic and Marmonier, 2003	6
13	<i>Schellencandona</i> Meisch, 1996	3
14	<i>Caribecandona</i> Broodbakker, 1983	3
15	<i>Marococandona</i> Marmonier, Boulal and Idbennacer, 2005	2
16	<i>Notacandona</i> Karanovic and Marmonier, 2003	2
17	<i>Origocandona</i> Karanovic, 2005	2
18	<i>Caaporacandona</i> Pinto, Rocha and Martens, 2005	2
19	<i>Acandona</i> Karanovic, 2003	2
20	<i>Kencandona</i> Karanovic, 2007	2
21	<i>Trajangcandona</i> Karanovic, 1999	2
22	<i>Paracandona</i> Hartwig, 1899	1
23	<i>Phreatocandona</i> Danielopol, 1973	1
24	<i>Meridiescandona</i> Karanovic, 2003	1
25	<i>Pierrecandona</i> Karanovic, 2007	1
26	<i>Danielocandona</i> Broodbakker, 1983	1
27	<i>Namibcypris</i> Martens, 1992	1
28	<i>Nannocandona</i> Ekman, 1914	1
29	<i>Terrestriocypris</i> Danielopol and Betsch, 1980	1
30	<i>Indocandona</i> Gupta, 1984	1
31	<i>Meischcandona</i> Karanovic, 2001	1
32	<i>Latinopsis</i> Karanovic and Datry, 2009	1
33	<i>Amphitritecandona</i> Karanovic, 2007	1
34	<i>Pioneeracandonopsis</i> Karanovic, 2005	1
Total	34	293

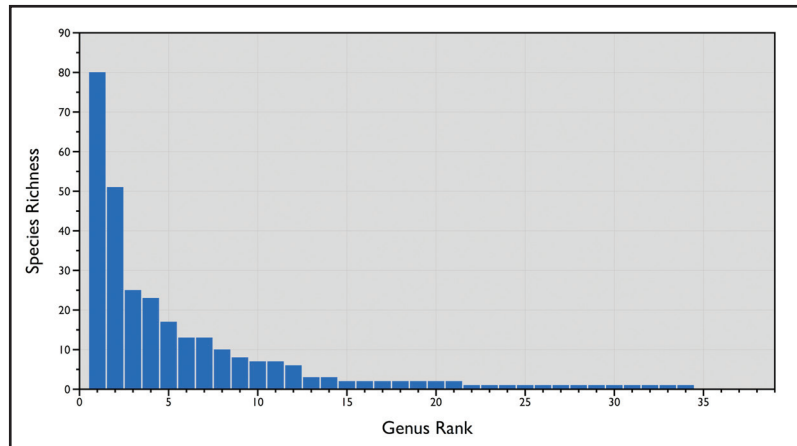


Fig. 1 Histogram of the Candoninae genera (cf. Table 1) ranked from left to right in decreasing order following their species richness

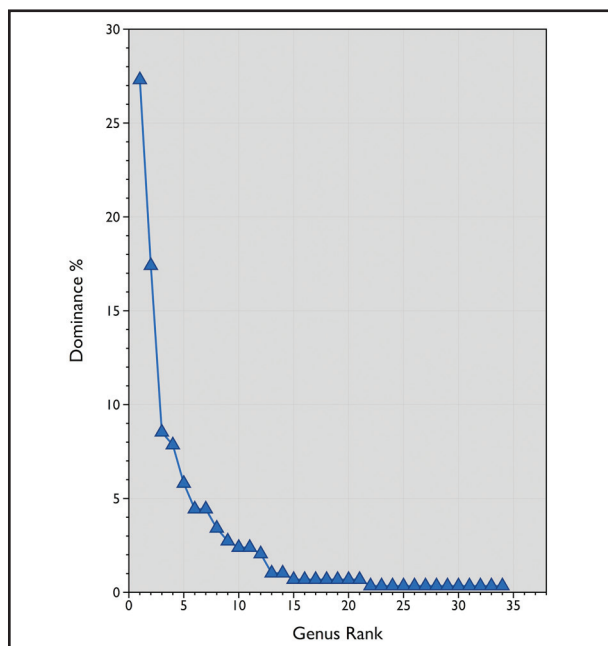


Fig. 2 Dominance of Candoninae genera (cf. Table 1) ranked in decreasing order, following their percentage contribution to the total species richness

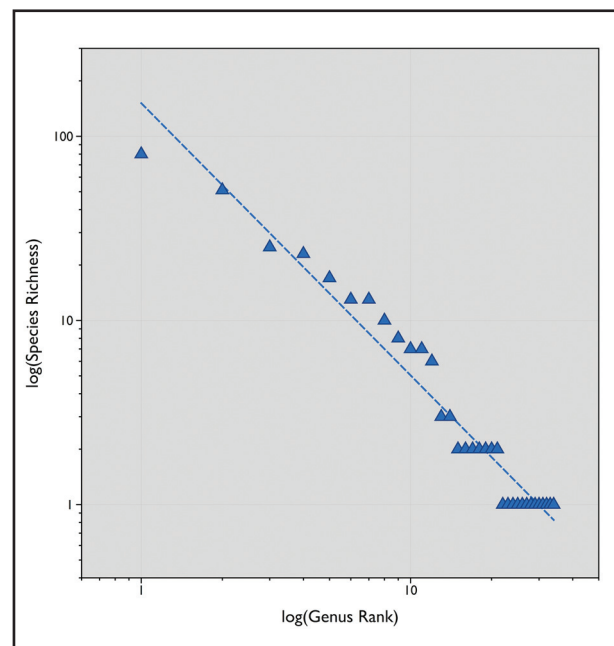


Fig. 3 Dominance of Candoninae genera (data set in Table 1) ranked in decreasing order, on double logarithmic scale, with its regression line with negative slope

Given the data pairs (k, x_k) , $k \in \{1, 2, \dots, 80\}$, derived from Table 1 the corresponding geometric mean is

$$\tilde{x}_{34} = \sqrt[34]{1^{13} \cdot 2^7 \cdot 3^2 \cdot 6^1 \cdot 7^2 \cdot 8^1 \cdot 10^1 \cdot 13^2 \cdot 17^1 \cdot 23^1 \cdot 25^1 \cdot 51^1 \cdot 80^1} = 3.2188$$

and therefore,

$$\ln(\tilde{x}_{34}) = 1.169$$

The solution of the equation

$$\frac{-\zeta'(\rho+1)}{\zeta(\rho+1)} = \ln \tilde{x}_n$$

is obtained by linear interpolation from the following part of the table mentioned before:

ρ	$\frac{-\zeta'(\rho+1)}{\zeta(\rho+1)}$
0.6	1.186
0.7	0.961

The estimated value for the parameter r of the Zeta Distribution is $\hat{\rho} = 0.60756$.

In the following table, we compare the observed number $b(k)$ of genera for species k with the expected number $e_{\hat{\rho}}(k)$, given a Zeta Distribution with parameter $\hat{\rho}$, $k \in \{1, 2, \dots, 13\}$.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	≥ 14
$b(k)$	13	7	2	0	0	1	2	1	0	1	0	0	2	5
$e_{\hat{\rho}}(k)$	15.00	4.92	2.79	2.57	1.62	0.84	0.66	0.53	0.44	0.37	0.32	0.28	0.24	5.08

Now we are going to apply a Chi-square test: Observing the rule of thumb that – under the null-hypothesis – the expected number $e(C)$ of items is $e(C) \geq 5$ for every cell C , we obtain the following 4 cells

$$C_1 = \{1\}, \quad C_2 = \{2, 3\}, \quad C_3 = \{4, 5, 6, 7, 8, 9\}, \quad C_4 = \{10, \dots\}$$

The corresponding observed $\chi^2 = 1.3212$.

The $(1 - \alpha)$ quartiles of χ^2 distribution with 2 degrees of freedom are for $\alpha \in \{0.5, 0.2, 0.1, 0.05, 0.01\}$

α	0.5	0.2	0.1	0.05	0.01
$\chi^2_{2, 1-\alpha}$	1.386	3.219	4.605	5.991	9.210

Since the observed χ^2 value 1.3212 is even smaller than the median 1.386 of the χ^2_2 -distribution, we have a pretty good fit. Therefore it is reasonable to assume that our data stem from a Zeta-distributed population.

j	1	2	3	4	5	6	7	8	9	10	11	12	13	≥ 14
$b(j)$	13	7	2	0	0	1	2	1	0	1	0	0	2	5
$e_{\hat{\rho}}(j)$	18.05	5.76	2.79	1.63	1.06	0.75	0.55	0.42	0.33	0.27	0.22	0.19	0.16	1.82

Now we are going to apply a Chi-square test: With the reasoning explained in the case of a Zeta Distribution we obtain the following 3 cells

$$C_1 = \{1\}, \quad C_2 = \{2\}, \quad C_3 = \{3, \dots\}$$

The observed χ^2 -value turns out to be $\chi^2 = 3.0317$.

The $(1 - \alpha)$ -quartiles of the χ^2 distribution with 1 (=3-2) degree of freedom are for $\alpha \in \{0.5, 0.2, 0.1, 0.05, 0.01\}$

α	0.5	0.2	0.1	0.05	0.01
$\chi^2_{1, 1-\alpha}$	0.455	1.642	2.706	3.841	6.635

With an observed χ^2 value of 3.0317 (compared with the median 0.455 and the expected value 1) of the χ^2_1 -distribution,

We proceed now with a Goodness of fit test for adjusting a Yule-Simon Distribution to our data:

Given the data pairs $(k, x_k), k \in \{1, 2, \dots, 80\}$, derived from Table 1 and observing that we have 293 species and 34 genera, the mean value of species per genus is

$$\bar{x}_{34} = \frac{293}{34} = 8.6176$$

Hence the estimated value for the parameter ρ of the Yule-Simon Distribution is

$$\hat{\rho}_{34} = 1 + \frac{1}{\bar{x}_{34} - 1} = 1 + \frac{1}{\frac{293}{34} - 1} = 1 + \frac{34}{259} = 1.1313$$

In the following table, we compare the observed number $b(k)$ of genera for species k with the expected number $e_{\hat{\rho}_{34}}(k)$ given a Yule-Simon-distribution with parameter $\hat{\rho}_{34}$, $k \in \{1, 2, \dots, 13\}$.

the fit of a Yule-Simon Distribution to the data is by far not as good as in the case of a Zeta Distribution. However, since $3.0317 < 3.841 = \chi^2_{1, 1-0.05}$ the null-hypothesis, that our data stem from a Yule-Simon distributed population, is not to be rejected for the level $\alpha = 0.05$ of significance. The observed level (P-value) is 0.08165.

3.2. THE CANDONINAE DATA SET (MARTENS AND SAVATENALINTON, 2011)

The data of Martens and Savatnalinton (M & S) for the S. F. Candoninae, is presented in Table 2 and Figures 4 - 6. There are 512 species classified in 37 genera.

Fig. 4 Histogram of the Candoninae genera (cf. Table 2) ranked from left to right in decreasing order following their species richness

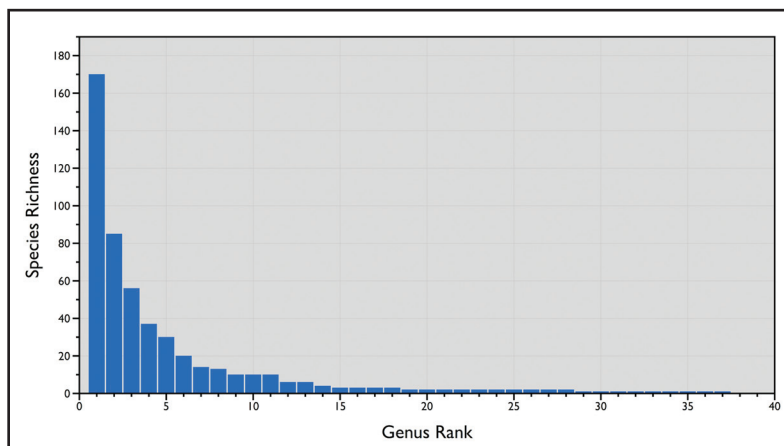


Table 2 The full data set of the S. F. Candoninae
(from Martens and Savatnalinton, 2011)

S-F. Candoninae Kaufmann 1900		
Rank	Genus Name	N. Species
1	<i>Candona</i> Baird, 1845	170
2	<i>Pseudocandona</i> Kaufmann, 1900	85
3	<i>Fabaeformiscandona</i> Krstic, 1972	56
4	<i>Candonopsis</i> Vávra, 1891	37
5	<i>Areacandona</i> Karanovic, 2005	30
6	<i>Mixtacandona</i> Klie, 1938	20
7	<i>Cryptocandona</i> Kaufmann, 1900	14
8	<i>Baicalocandona</i> Mazepova, 1976	13
9	<i>Humphreycandona</i> Karanovic and Marmonier, 2003	10
10	<i>Leicacandona</i> Karanovic, 2007	10
11	<i>Deminutiocandona</i> Karanovic, 2003	10
12	<i>Pilbaracandona</i> Karanovic and Marmonier, 2003	6
13	<i>Schellencandona</i> Meisch, 1996	6
14	<i>Latinopsis</i> Karanovic and Datry, 2009	4
15	<i>Caribecandona</i> Broodbakker, 1983	3
16	<i>Meridiescandona</i> Karanovic, 2003	3
17	<i>Notacandona</i> Karanovic and Marmonier, 2003	3
18	<i>Origocandona</i> Karanovic, 2005	3
19	<i>Caaporacandona</i> Pinto, Rocha and Martens, 2005	2
20	<i>Marococandona</i> Marmonier, Boulal and Idbennacer, 2005	2
21	<i>Acandona</i> Karanovic, 2003	2
22	<i>Kencandona</i> Karanovic, 2007	2
23	<i>Trajanacandona</i> Karanovic, 1999	2
24	<i>Cubacandona</i> Broodbakker, 1983	2
25	<i>Nannocandona</i> Ekman, 1914	2
26	<i>Terrestriocypris</i> Schornikov, 1969	2
27	<i>Indocandona</i> Gupta, 1984	2
28	<i>Amphitritecandona</i> Karanovic, 2007	2
29	<i>Paracandona</i> Hartwig, 1899	1
30	<i>Phreatocandona</i> Danielopol, 1973	1
31	<i>Pierrecandona</i> Karanovic, 2007	1
32	<i>Alatocandona</i> Carbonnel, 1969	1
33	<i>Danielocandona</i> Broodbakker, 1983	1
34	<i>Namibcypris</i> Martens, 1992	1
35	<i>Terrestriacandona</i> Danielopol and Betsch, 1980	1
36	<i>Meischcandona</i> Karanovic, 2001	1
37	<i>Pioneeracandonopsis</i> Karanovic, 2005	1
Total	37	512

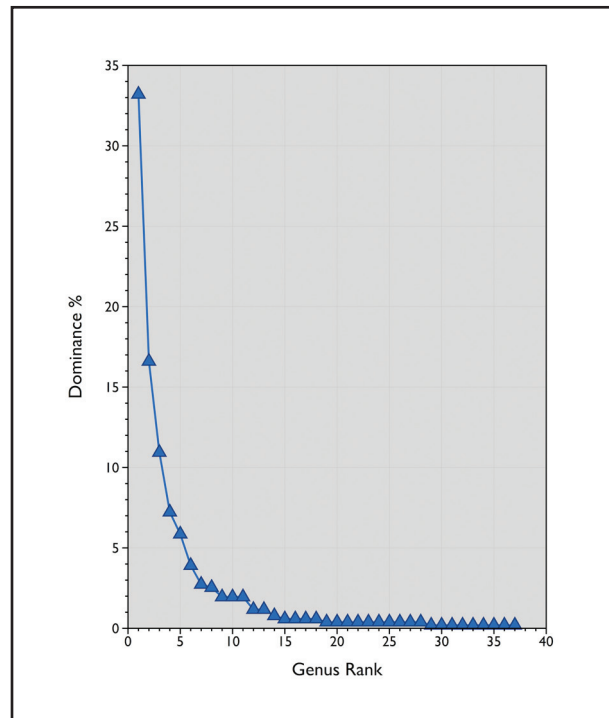


Fig. 5 Dominance of Candoninae genera (cf. Table 2) ranked in decreasing order, following their percentage contribution to the total species richness

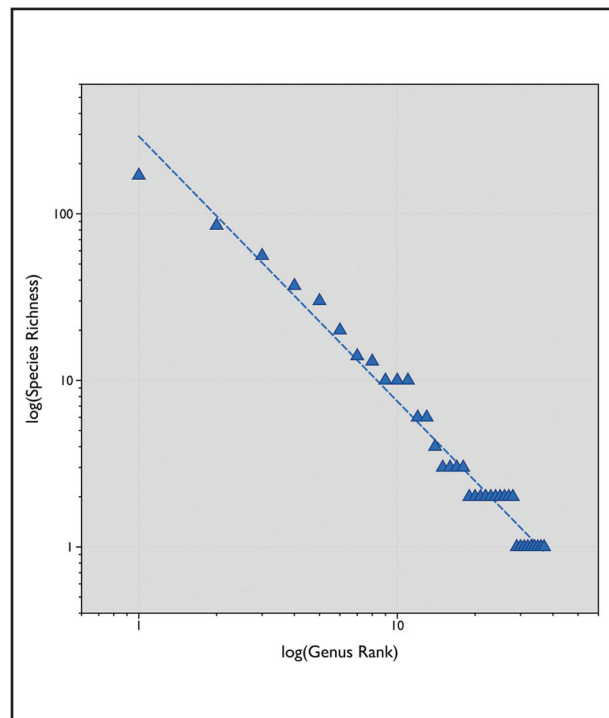


Fig. 6 Dominance of Candoninae genera (cf. Table 2) ranked in decreasing order, on double logarithmic scale, with its regression line with negative slope

The most frequent supraspecific taxa in the tail are genera containing two species (Fig. 4). There are 7 dominant genera (i.e. “primordial” in the terminology of Yule, 1925), with at least 14 species, which cumulate 80.5% of the total species (cf. Table 2). Especially the first genus, *Candona*, is very dominant (cf. Fig. 5).

A Goodness of fit test for adjusting a Zeta Distribution to the data set of Table 2 shows (cf. Table 3) that they are only in limited accordance with a Power Law. However, when the data are double logarithmically plotted (Fig. 6), the position

of the genera lay close to the regression line which has a negative slope of 1.61 and the coefficient of determination r^2 for the linear regression is 0.98. The M & S data set differs from the previous one by a stronger dominance of the first ranked genus (*Candona*) and by the shorter tail of the monotypic genera.

The data set from Table 2 will be also checked for statistically fit to a Yule-Simon Distribution. Below the observed number $b(k)$ of genera for species k with the expected number $e_{\rho_{34}}(k)$.

k	1	2	3	4	5	6	7	8	9	10	11	12	13	≥ 14
$b(k)$	9	10	4	1	0	2	0	0	0	3	0	0	1	7
$e_{\rho_{34}}(k)$	19.19	6.29	3.06	1.81	1.19	0.84	0.62	0.48	0.38	0.31	0.26	0.22	0.18	2.22

In the above data we note large differences between the observed frequency of monotypic genera and their expected value. The latter is for $k = 1$ more than the double of the observed one (19.19 to 9): The same is visible for the “primordial” (i.e. dominant) genera and their expected value, 7 and 2.22, respectively.

The results presented in Table 3 show that they fail to follow a Yule-Simon Distribution. This is due to the rather high number of species-rich genera and to the strong dominance of the first genus, *Candona*; also, to the lower number of monotypic genera in the tail as compared to the series of genera with 2 species (cf. Figs. 3 and 4) leads to the poorer adjustment of the data to the theoretical distribution.

In order to see what is the importance of the endemic species of Baikal lake (Ba), where *Candona* and *Pseudocan-*

dona species are very abundant, we removed first the Baikal Candoninae from the data mentioned in Table 2 and we added one new monotypic genus, *Bicornucandona* recently published (Külköylüoglu *et al.*, 2011). We obtain, in this case, a new data set with 37 genera and 418 species (Fig. 7).

The monotypic genera equal the genera with 2 species namely, 10 to 10 (Fig. 7). On one side the approximation to the Zeta Distribution is improved (cf. Table 3) but on the other side the new data set does not fit a theoretical Yule-Simon Distribution. Even if we notice a lowering of the dominance of the genus *Candona* (Fig. 8), still large differences between the observed and the calculated data for the two extreme classes, those of monotypic and those of the “primordial” genera remain, respectively 10 and 19.36, in the former case, and 7 and 2.13, in the latter one.

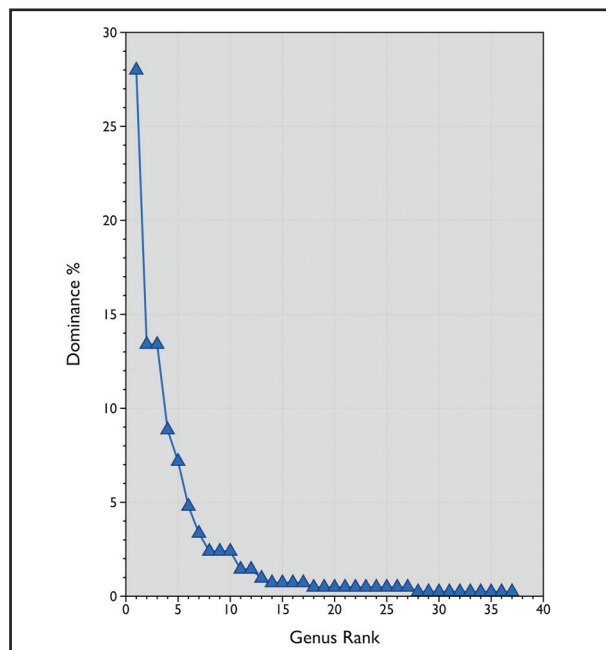


Fig. 7 Dominance of Candoninae genera when the Baikal species are eliminated and *Bicornucandona* included

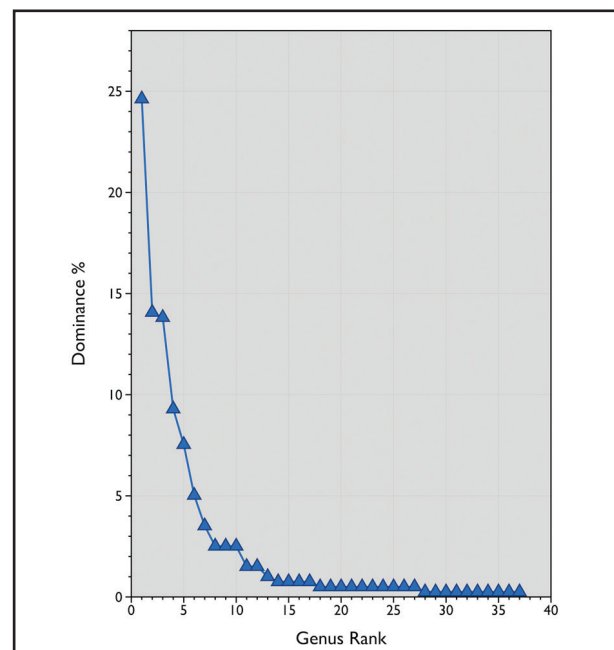


Fig. 8 Dominance of Candoninae genera when the Baikal and Ohrid species are eliminated but *Bicornucandona* included

Table 3 Adjustment of the data to the Zeta and the Yule-Simon Distributions (*cf.* text)

Data set	Zeta distribution				Yule-Simon distribution			
	Estimated value \hat{p}	χ^2 – value	df	P – value	Estimated value \hat{p}	χ^2 – value	df	P – value
G-34 / Sp-293	0.60756	1.3212	2	0.51654	1.1313	3.0317	1	0.08165
G-37 / Sp-512	0.53107	7.6744	3	0.05324	1.0779	18.023	2	0.00012
G-37 / Sp-418	0.55950	6.5897	3	0.08619	1.0971	15.138	2	0.00052
G-37 / Sp-398	0.56116	6.6036	3	0.08567	1.1025	15.325	2	0.00047

We continue by eliminating also the endemic Ohrid Candoninae (Oh), but we maintain the monotypic genus *Bicornucandona*. The new data set has now the same number of genera (37), but 398 species. The dominance of the genus *Candona* continues to decrease (Fig. 8), but the approximation to the Yule-Simon Distribution does not improve (*cf.* Table 3).

3.3. COMPARATIVE DATA

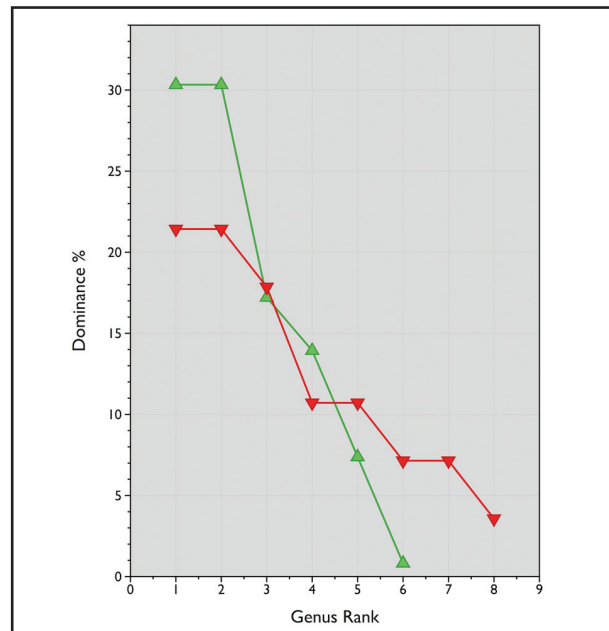
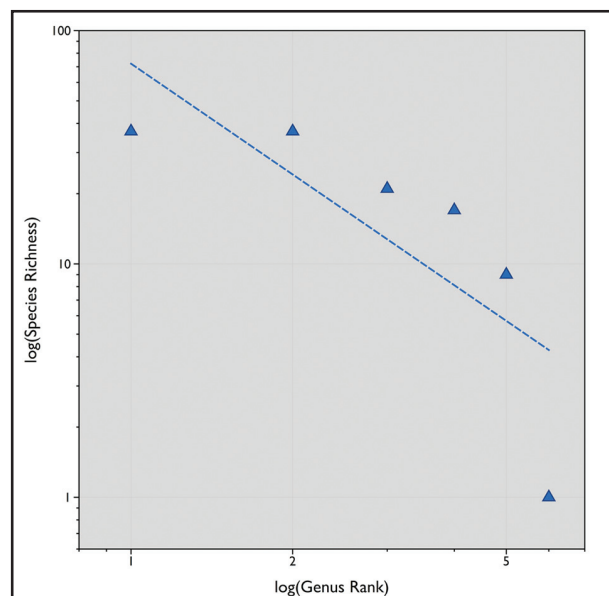
Tables 4 and 5 and Figure 9 present the data for the S. F. Cyclopyridinae (green triangles) and the S. F. Paracypridinae (inverted red triangles) from Martens and Savatnalinton (2011). Both data sets do not display the “hollow” pattern of the Candoninae curves, respectively there is no tail represented by monotypic genera in these curves (Fig. 9). On a double logarithmic plot (Fig. 10), the 6 Cyclopyridinae genera are poorly aligned to the regression line. The coefficient of determination r^2 is only 0.5929 and the total frequency pattern of Cyclopyridinae differs significantly from a Zeta Distribution.

Table 4 The data set of S. F. Cyclopyridinae (Martens and Savatnalinton, 2011)

Rank	Genus	Species
1	<i>Cypria</i> Zenker, 1854	37
2	<i>Physocypria</i> Vávra, 1897	37
3	<i>Cyclopypris</i> Brady & Norman, 1889	21
4	<i>Mecynocypria</i> Rome, 1962	17
5	<i>Allocypria</i> Rome, 1962	9
6	<i>Candocypria</i> Furtos, 1933	1
Total	6	122

Table 5 The data set of S. F. Paracypridinae (Martens and Savatnalinton, 2011)

Rank	Genus	Species
1	<i>Paracypria</i> Sars, 1910	6
2	<i>Dolerocypria</i> Tressler, 1937	6
3	<i>Mungava</i> Harding, 1962	5
4	<i>Thalassocypria</i> Hartmann, 1957	3
5	<i>Mangalocypria</i> Wouters, 1998	3
6	<i>Renaudocypria</i> McKenzie, 1980	2
7	<i>Hansacypris</i> Wouters, 1984	2
8	<i>Pontoparta</i> Vávra, 1901	1
Total	8	28

**Fig. 9** Dominance of Cyclopyridinae (green) and Paracypridinae (red) genera (*cf.* Tables 4 and 5), ranked in decreasing order following their species richness**Fig. 10** Dominance of Cyclopyridinae genera (*cf.* Table 4) ranked in decreasing order, on double natural logarithmic scale, with its regression line with negative slope

4. DISCUSSION: ON THE CAUSES OF TAXONOMIC DIVERSITY FOR SELECTED CANDONINAE GENERA

4.1. INTRODUCTORY NOTES

The taxonomic diversity of the Candoninae genera in the data set of Martens and Savatnalinton (2011) fits a Zeta Distribution but not a Yule-Simon Distribution. This situation needs discussion. First, we have to explain why we consider the Yule-Simon model appropriate for investigation of taxonomic diversity of ostracods.

Yule (1925) developed a power-law algorithm which mathematically described an evolutionary process of taxonomic diversification of organisms as a pure birth model running in a simultaneous way at the level of species (as speciation process) and at the genus level (it can be called geniation process). In this scenario, species originate, randomly, at a constant rate, during the time through micromutations from genera which already other species have. This is similar to the Polya's Urn Problem (cf. Chung *et al.*, 2003). New genera form from a given species through macromutations also randomly and at a constant rate. The end product of the evolutionary Yule process, during a given time interval, is an organismic assemblage which occupies a large space and where one finds a reduced number of species-rich genera to which many monotypic genera are added. Because this is a stochastic process running at constant rates one can visualise this diversification process as a neutral evolutionary model (Nee, 2006). It can be used as a reference (or theoretical standard) for comparing the empirical size distributions of genera within a given organismic assemblage. Nee (2005) underlined the interest for understanding macroecological processes through neutral models similar with those used in molecular biology. For the case of the Yule-Simon model, because the process is supposed to run under constant time sequences, one can compare the rate of diversification for various genera during a given time scale (Aldous, 2002; Nee, 2001). It can also be used for predicting the theoretical structure of the taxonomic diversity of a given assemblage. In the precise case of the ostracod fauna, it can help to formulate hypotheses for the origin of the observed strong dominance of genera, like *Candona* and *Pseudocandona* or for the high frequency of monotypic genera.

The data set of Martens and Savatnalinton (2011), when compared to the Yule-Simon theoretical distribution for the frequency of what we defined as "primordial" or species-rich genera, show a clear deviation. There are more genera than expected and some of the genera are very species rich, especially *Candona* and *Pseudocandona* (cf. Table 2, Figs 4 and 5). Similar deviations were observed also by Yule (1925) in one of his data sets and by Kornai (1999), who used the Zipf's law for data analysis. Statisticians like Simon (1955) and/or Gan *et al.* (2006), noted that Power Laws are statistical phenomena which do not require any other explanation for a possible "deeper" causality. Maruvka *et al.* (2011) suggested that the

lack of fit to the theoretical model should imply that other causes than random events could generate the empirical distribution. This approach will be developed below for the data set of Martens and Savatnalinton (2011).

Yule (1925) expressed the idea that the most ancient taxa at the generic level will be the most species-rich, as compared with genera which have few species, an idea already mentioned by Willis and Yule (1922). We explored this hypothesis considering the Candoninae and we noted that it does not really fit our data. This is in line with the observations of other specialists (review in Minelli *et al.*, 1991).

Below, we will discuss in depth two aspects from which we will derive our proposal for a review of the taxonomic system of Candoninae. First, we look for the possible causes explaining the species-rich genera with a special emphasis on the genera *Candona* Baird and *Pseudocandona* Kaufmann. Is their species richness the outcome of a real evolutionary process or a socio-psychological by-product of taxonomists? Second, we will explore the origin of monotypic genera of Candoninae as well as their potential interest for further taxonomic research of this group.

4.2. POSSIBLE CAUSES FOR THE SPECIES RICHNESS OF GENERA *CANDONA* AND *PSEUDOCANDONA*

4.2.1. The developed morphological traits related to the mating process

The comparative morphology of the various groups of the Candoninae offers us an idea about their evolutionary position within the subfamily. The genera *Cryptocandona* Kaufmann and *Paracandona* Hartwig have a more complete chaetotaxy. For instance, there are setae on the 7th limbs that occur also in the related subfamily Cyclopyridinae, e.g. d2 and the f setae. The male's clasping organs, as well as the inner lobes of the hemipenis belonging to representatives of these genera, are poorly developed like in the case of Cyclopyridinae. The opposite is the case for the representatives of *Candona* and *Pseudocandona*. Species of these latter genera have a more reduced chaetotaxy, e.g. the setae mentioned above for the 7th limb are absent, or those of the exopodite of the 5th limb. The male's copulatory organ has a well developed M process, specialized for sensorial communication during the matting process. Information on this topic was published by Danielopol (1969, 1980, 1982), Horne *et al.* (1998), Namiotko *et al.* (2005), Iepure *et al.* (2008).

The phylogenetic analysis of Candoninae made by Karanovic (2007) points out to a basal place on the phylogenetic tree of *Cryptocandona* and *Paracandona*, while *Candona* and *Pseudocandona* are in a higher hierarchical position. These details point to the primitive position of the former group as compared to *Candona* and *Pseudocandona*.

The most ancient Candoninae group is documented with fossils from Middle Jurassic (Bajocian), i.e. the genus *Septacandona* in Europe (Cabrál and Colin, 2002). Both *Candona* and *Paracandona* were recorded as fossils in the Upper Cre-

taceous. Sczechura (1978) mentioned *Candona altanulaensis* Schzechura and Blaszyk in the Campanian from Nemegt Basin in Mongolia and representatives of the genus *Paracandona* (*P. gardannensis* Babinot and *P. krsticae* Babinot) are known from the Upper Cretaceous deposits in Southern France (Babinot, 1980). The oldest *Cryptocandona* species is known from the Upper Miocene in Slovakia (Pipik, unpublished) and *Pseudocandona* representatives are also known since the Tertiary at the passage between the Oligocene and the Miocene (Triebe, 1963).

We now turn to the species richness of the four genera documented with Recent species in Martens and Savatentalint, (2011) and here, Table 2. *Candona* and *Pseudocandona* are the most species-rich genera, while *Cryptocandona* and *Paracandona* have a reduced number of species. This discrepancy is due, in our opinion, not to the difference in antiquity of these groups as it would be requested by the Yule's evolutionary scenario (cf. Yule, 1925) but by differences in the development of sexual organs and the sexual behaviour. In our opinion, the more evolved sexual organs related to a developed mating behaviour for representatives of *Candona* and *Pseudocandona* (cf. publications mentioned above) played a key role for the production of higher number of species as compared to those of the more primitive genera *Cryptocandona* and *Paracandona* here discussed.

4.2.2. The contribution of the endemic species from lake Baikal and lake Ohrid

Long-lived lakes, like the Baikal in Siberia and the Ohrid in the Balkans, are famous by their endemic genera (cf. Martens 1997). Within these latter, species flocks formed through explosive speciation (Martens and Schön, 1999; Turner, 1999). Their origins can be viewed, either as adaptive, or non-adaptive radiations (*sensu* Guyer and Slowinski, 1993; Schluter, 2000).

In the case of the ostracod Candoninae of lake Baikal, Mazepova (1990) showed that the genera *Candona*, *Pseudocandona* and *Baicalocandona* constitute three large groups with endemic species. To *Candona* belong 53 species and subspecies (these latter taxa are here treated as species). In the genus *Pseudocandona* we find 28 species. The endemic *Baicalocandona* has 13 species. For lake Ohrid we have only one species flock with 19 species of *Candona*, described by W. Klie and F. Mikulić in various publications during the first part of the 20th century. There is only one endemic *Pseudocandona* described by Petkovski (1969). If we go back to the inventory of Martens and Savatentalint, (2011), we notice that the contribution of the *Candona* from these two lakes form 42.3 % from the total species catalogued and *Pseudocandona* of the Baikal represents 30 %.

We hypothesised that the discrepancy between the theoretical Yule-Simon distribution and the empirical data could be partly due to the strong species contribution of these endemic taxa from the two lakes. The results from Table 3

showing that the new data sets do not fit the Yule-Simon Distribution compel us to look for alternative explanations. One possibility would be that the large flocks of *Candona* and *Pseudocandona* of the Baikal are inflated with taxa which do not belong to these genera.

The similarity of the Zenker organ of *C. inaequalis* baikalensis Bronshtein with those of *Ps. belgica* Klie suggested to Danielopol (1982) the necessary transfer of the baikalian taxon from the genus *Candona* to *Pseudocandona*. Also, the criterion of the antennal dimorphic t setae, which was used by Bronshtein (1947) and Mazepova (1990), is not valid anymore. Therefore, a taxonomic revision of the taxa which are now assigned to the two baikalian species flocks is necessary. We foresee the removal from these two genera of taxa which will define new smaller genera, with closer phylogenetic affinities.

The endemic *Candona* of lake Ohrid constitute another problem: the carapace shape of some species is similar to fossil taxa from the Miocene and Pliocene lakes of the Balkans and/or of Central Europe. One should compare species of Mikulić (1961) with those of Krstić (1972) and/or Pipik and Bodergat (2005, 2006). For instance, the valves of the living species *Candona parvula* Mikulić from the Ohrid lake display shape similarity with the Miocene *Candona imaginaria* Pipik and Bodergat (2005) from the long-lived palaeolake Turiec in Slovakia. Hence, maybe, we have to split the *Candona* flock and regroup the various species within genera which have both Recent and fossil taxa. All these reorganisations could diminish the dominance of the two genera *Candona* and *Pseudocandona* and therefore one could obtain a better fit for the whole Candoninae genera to a Yule-Simon Distribution.

4.2.3. The impact of socio-psychological practice of taxonomists

A brief examination of the literature about the taxonomy of non-marine ostracods, during the last 150 years, suggests that the practitioners of this research direction were in many cases guided by what one could call the Principle of Authority or Principle of Faith (van Fraassen, 2000). It means that, in many cases, ostracod taxonomists, accepted and further followed diagnostic descriptions or systematic decisions for whole groups published by established carcinologists in monographic studies. This is due to the fact that every specialist needs clear guidelines for ostracod identification and, therefore, many of us followed the taxonomic monographs of well-known ostracodologists (cf. Kaufmann, 1900; Klie, 1938; Bronshtein, (1947). For non-marine ostracods, present-day European practitioners use the Meisch (2000) monograph with priority.

Considering the systematics of the genera *Candona* and *Pseudocandona*, one can see that about 25% of the number of species was described by one specialist (cf. Mazepova, 1990). In this latter synthetic work, Mazepova followed the Russian authority Z. S. Bronshtein. The above epistemic attitude was adopted by other ostracodologists too. For instance,

H. Löffler in the compendium "Limnofauna Europaea" (cf. Löffler and Danielopol, 1978) followed mainly the systematics used in W. Klie's monograph (Klie, 1938). This latter was the most important identification tool for the limnic ostracods in Central Europe, before the advent in 2000, of the excellent Meisch's book.

Ostracodologists should note that the problem exemplified in this section has been recognised for some time by specialists dealing with other organismic groups. Walters (1986) complained that taxonomists dealing with angiosperms do not explain commonly the identification criteria they use. Moreover, they do not explain what are their philosophical conception for defining species and genera. Therefore, some of the "hollow curves" published for angiosperms could be biased due to invalid taxa. Minelli *et al.* (1991) and, more recently, Gaston (1996) acknowledged the necessity to carefully evaluate the potential taxonomic bias due to subjective evaluation of systematists, if one needs to discuss topics dealing with taxonomic diversity.

In conclusion, we consider that species-rich genera of Candoninae like *Candona* and *Pseudocandona* should be taxonomically reviewed following multiple criteria, like those related to evolutionary and/or ecological processes, or those related to socio-psychological aspects of the practitioner taxonomists.

4.3. THE MONOTYPIC GENERA OF THE CANDONINAE: WHY ARE THEY INTERESTING?

Our comparative study on the three subfamilies of Candonidae, using the taxonomic data from Martens and Savatnalinton (2011), shows that only the Candoninae present, in their ranked order distribution, a long tail of monotypic genera. Their record can be explained as due either: (1) to a random sampling effect, (2) to a macroevolutionary process, (3) or to the subjective decision of specialists who deal with the systematics of the Candoninae group.

Considering the first possibility, we immediately compared the frequency of empirical data with the expected values calculated for the Yule-Simon Distribution. If this should occur by a random effect, one would expect that the empirical value for the frequency of the monotypic genera would be closer to the theoretical one calculated for the Yule-Simon Distribution. We saw previously that there is a large difference, the expected value is more than the double of the observed one (19.19 to 9). Therefore, we explore a biological explanation for the origin of monotypic genera inspired from a hypothesis proposed for other ostracod groups by Martens *et al.* (2000). These latter authors consider that some of the endemic genera in the long-lived lakes Baikal and Tanganyika, belonging to *Cytherissa* and the *Cyprideis* flocks could originate rapidly by large morphological and/or genetical changes that they called quantum geniation. For such rapid and disruptive evolution leading to new genera, Wouters and Martens (2001, p. 125) speak about a "saltatory origin", a process which

is different from the gradual species differentiation. Working with methods common to comparative morphology and systematics, we do not arrive at such a view but there is an indisputable situation: we have a series of ostracod species with a peculiar combination of characters which are so different from the known common species that prompted taxonomists to define new genera. Such an example is *Phreatocandona motasi* (Danielopol, 1978), which, when it was examined for the exopodite of the antenna, it was so different from the other Candoninae, known at the time, that a new genus was erected for a unique species. Going one step further, we have to note that such similar situations occur more frequently in the case of Candoninae as compared to the Cyclocypridinae (cf. Martens and Savatnalinton, 2011). Taking into consideration that both groups have been known and investigated for many years and by many generations of ostracodologists in the same way one can hypothesise that the former subfamily has a higher biological propensity to diversify than the latter one. Interestingly enough, Foote (2011), comparing the differences between the taxonomic diversity of marine bivalve and gastropod groups from the near New Zealand coast, was confronted with a similar problem. The endemic bivalve genera have a higher proportion of monotypics than do endemic gastropods. Foote for this case favours a biological explanation, that is to say, the evolvability of the endemic bivalves at the site investigated is higher than those of the gastropods. The alternative explanation discussed by Foote (2011) is that the difference in the degree taxonomic diversity is due to the practice of taxonomists. This latter alternative is with caution rejected because the intensity of taxonomic studies for both groups appears more or less similar.

Still, this does not solve the question of a biological origin for the monotypic genera of the mentioned data set of Candoninae. Therefore we looked to the history of the sampled fauna mentioned in the published literature. Thus, going from the list of Martens and Savatnalinton (2011) to the original description of the species we noted that most of the monotypic genera were based on material from geographical areas and/or ecological habitats which remained over time poorly investigated. Notorious examples are the species originating from subterranean or semi-terrestrial habitats, domains that only during the last decennials zoologists investigated in an intensive and systematic way. Take the case of the genera *Phreatocandona* Danielopol, 1973, *Terrestri-candona* Danielopol and Betsch, 1980 and/or *Pierrecandona* Karanovic, 2007. The same applies for the aquatic fauna of less well explored geographic areas in Africa, in South America, or Australia and South and East Asia. For instance *Namib-cypris* Martens, 1992, was found in a spring fed by groundwater in Namibia, *Danielocandona* Broodbakker, 1983 stems from groundwater habitats of Venezuela, *Meischcandona* Karanovic, 2001 originates from Mali.

Even in Europe, we are far from getting a comprehensive view of the subterranean Candoninae fauna (Stoch and Galassi, 2010). As an example of this situation we mentioned

above the unexpected discovery of *Phreatocandona motasi* in Romania. One has still to accept that we will always remain with an incomplete inventory of the subterranean aquatic fauna, even in Europe, where a long tradition for subterranean ecology exists. However, with the time our information, even on groundwater ostracods, is improving and new species are added, to the monotypic genera. As an example for *Phreatocandona*, the discovery of a second species (still undescribed) found in a spring fed by groundwater in Slovenia was recently mentioned (Mori *et al.*, 2011). The same situation applies to the groundwater fauna of Australia, practically unknown, until recently. Through extensive research of W. F. Humphreys and colleagues, it became clear that this continent harbours the world's most diverse subterranean aquatic fauna (Humphreys, 2004; Humphreys *et al.*, 2004). Karanovic (2007) presented for the North-western part of Australia in a monographic form the diverse Candoninae fauna.

Therefore, we consider that, beside the individual expertise and subjective decisions of taxonomists to erect new genera for particular species the context of the state of knowledge and the intensity with which various geographic areas and/or ecological domains are explored should be also considered as an explanation for the "long tail" of the Candoninae distribution here discussed.

Finally, one should also ask why to compare the theoretical frequency of monotypic genera expressed by the Yule-Simon Distribution with empirical data. For us the theoretical distribution represents a null model for a given moment in the state of art of our knowledge on the taxonomic diversity of Candoninae. Therefore in the case of the data set of Martens and Savatzen (2011) where we have 9 monotypic genera, instead of 19 predicted, one expects to discover and further to describe new monotypic genera. This is really the case! Take for instance, the recent publication of Külköylüoglu *et al.* (2011) for the genus *Bicornucandona*. Another new Candoninae genus (yet unpublished) comes from subterranean aquatic habitats in Far-East Asia (pers. comm. R. J. Smith). Therefore, inserting mathematical views, like the Yule Process, in the systematics of Candoninae, can be for the future a real advantage in order to make our descriptions and/or previsions more reliable or more objective.

5. CONCLUSION: THE TAXONOMIC SYSTEM OF THE SUBFAMILY CANDONINAE NEEDS TO BE RENEWED

The mathematical contribution presented above is intellectually stimulating, not only for the Candoninae project at which two of us (B. and D.) are currently working, but also for a larger audience interested in evolutionary and epistemological problems of ostracods. Incorporating information derived from the mathematical treatment of the above data sets we became aware about the necessity to rethink the taxonomic system of Candoninae. We now understand better that the image that we have about the taxonomical diversity of Can-

doninae is the result of biological processes combined with the personal perception of various specialists who developed the taxonomy of this group. In order to improve this system we should concentrate on the way we describe ostracods and construct taxonomies.

Inserting the mathematical information provided by U. M. and F. Ö. allow us to better view several hot-lines of research for the improvement of the taxonomic system of Candoninae:

(1) A critical re-examination (and, if possible, improvement) of the diagnostic criteria based on morphological traits for a series of Candoninae genera with living representatives. Such an idea was already foreseen by various colleagues. For instance, a solution to the taxonomical conundrum of the "*Pseudocandona-Typhlocypris*" complex of Karanovic (2005) is necessary.

(2) A special systematic examination and taxonomic re-evaluation of the *Candona* and *Pseudocandona* complex of species from the Baikal and the Ohrid lakes, become, considering the above results, a more stringent than either project.

(3) A redirection of neontological data toward the palaeontologists in order to make these latter clear that their taxonomic system for fossil candonines is very difficult to comprehend and use. Hence, the necessary cooperation between paleontologists and neontologists for the reconstruction of various phylogentic lineages of Candoninae where taxa are related also through natural ancestor-descendent relationships.

(4) The necessary incorporation in our perception about ostracods of new aspects of description. Below several possibilities:

(a) As ostracods are crustaceans the description of the ostracod limbs have to follow the general nomenclature of these latter (*cf.* Horne, 2005; Boxshall *et al.*, 2010; Namiotko *et al.*, 2011).

(b) The ostracod carapace offers additionally interesting information when it is analysed with geometric-morphometric methods and within new research directions like the developmental trajectories (*cf.* Danielopol *et al.*, 2008; Baltanás and Danielopol, 2011).

(c) The exploration of the way molecular biological techniques combined with geometric morphometrics of valves can improve the taxonomic system of selected Candoninae groups. Research in progress by Prof. T. Namiotko and Dr. S. Iepure (pers. comm. to D.L. D., 02. 02. 2010) appears as a promising complementary project to the general goal, here discussed.

(d) Rethinking the origin and antiquity of the Candoninae fauna of long-lived lakes, like Ohrid and Baikal, as well as from the palaeo-lake Pannon. Several ostracod-working groups are nowadays interested by this question and more cooperative research would be beneficial for faster scientific progress! Once we will revise the taxo-

nomic system of the dominant candonine genera *Candona* and *Pseudocandona* it will be possible to better appreciate the evolutionary radiation of Candoninae in the long-lived lakes, like the Ohrid, the Baikal or the palaeo-lake Pannon.

(e) It appears necessary to achieve a kind of harmonisation between the concepts on which species are described. The topic was discussed during the 7th European Ostracodologists's Meeting in Graz, July 2011 (cf. Horne *et al.*, 2011). This approach needs to be incorporated in the proposed framework for the new systematics of the S. F. Candoninae.

(5) Finally, a new taxonomic system of Candoninae should reflect phylogenetic relationships between the various taxa and the systematics should be more impregnated with ideas and methods specific to phylogenetic systematics.

We will close this essay mentioning that knowledge on ostracods will continue to be obtained by a combination of objective and subjective approaches. For us objectivity means that we describe features of ostracods which should be recognisable and, if necessary, reproducible. Subjectivity means that we look at the data and interpret them with our personal interests and knowledge. From the philosophical point of view this attitude is a mild constructivism.

One should note that we will never arrive to impose a unique view on the problems we investigate, therefore the ethical proposal made by the well-known Austrian scientist Heinz von Foerster "Act always so as to increase the numbers of choices" (von Foerster, 1995) appears as a necessary decision. It opens the door to many possible avenues for rejuvenating the research on ostracods and in the special case of the Candonidae it shows that the renewal of its taxonomic

system can be achieved in different ways, depending on the creativity of the scientists and their special interests.

6. ACKNOWLEDGEMENTS

The present contribution is dedicated to Professor Werner E. Piller for his 60th birthday. Discussion with Professor W. E. Piller, as well as with other colleagues on these topics over time was much stimulating. The following colleagues are here acknowledged: S. Belmecheri (Paris), A. Cohen (San Rafael, CA), J. P. Colin (Cestas), M. Foote (Chicago), W. Geiger (Vienna), M. Gross (Graz), S. Iepure (Cluj-Napoca), W. F. Humphreys (Perth), D. J. Horne (London), K. Martens (Brussels), P. Marmonier (Lyon), Cl. Meisch (Luxembourg), T. Namiotko (Gdansk), R. Olteanu (Bucharest), R. Pipík (Banská Bystrica), B. Sket (Ljubljana), R. J. Smith (Oroshimo), I. Tabacaru (Bucharest), C. Vasiliu (Bucharest), U. von Grafenstein (Gif sur Yvette).

This contribution is also an output of earlier projects, namely those of D.L. D. supported by the Austrian Science Fund (projects P. 17738-B03 and P. I35-B06) and that of A.B. in Spain (grant no. CGL2010-21039, Ministry of Science and Innovation).

POSTSCRIPT. After we completed the MS of this essay, we received the publication of L. N. Joppa, D.L. Roberts and S. L. Pimm, "The population ecology and social behaviour of taxonomists" published by *Trends in Ecology and Evolution*. (Vol. 26, 11: 551-553). The authors mention a possible decrease in the number of "expected undescribed species", noticed for several animal groups, which could be related to a trend for increasing specialisation of taxonomists over time. We consider that our contribution should be seen as a companion of the mentioned publication, in order to open a brighter discussion on the sociology of taxonomic practice related to biological diversity.

REFERENCES

- ALDOUS, D. J., 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Statist. Sci.*, 16: 23-34.
- ALDOUS, D. J., KRIKUN, M., POPOVIC, L., 2008. Stochastic models for phylogenetic trees of higher-order taxa. *J. Math. Biol.*, 56: 525-557.
- ALLMON, W. D., BOTTJER, D. J. (Eds.), 2001. *Evolutionary Paleocology*: 1-357, Columbia University Press, New York.
- BABINOT, J-F. (1980). Les Ostracodes du Crétacé Supérieur de Provence. Systématique, biostratigraphie, paléoécologie, paléogéographie. *Trav. Lab. Géol. Hist. and Paléontol. Univ. Provence*, 10: 1-634.
- BALTANÁS, A., 1992. On the use of some methods for the estimation of species richness. *Oikos*, 65: 484-492.
- BALTANÁS, A., DANIELOPOL, D.L., 2011. Geometric Morphometrics and its use in ostracod research. *Joannea Geol. Paläont.*, 11: 235-272.
- BRONSHTEIN, Z.S., 1947. *Ostracodes des eaux douces*. In: *Faune de l'URSS. Crustacés* In.: E. N. Pavlovsky and A. A. Stackelberg (Eds.). Institut Zoologique de l'Académie des Sciences de l'URSS, N.S. Moscou-Leningrad, 31, 2, 1: 1-339. (In Russian, with French title).
- BROWN, J. H., WEST, G.B. (Eds.), 2000. *Scaling in Biology*: 352p, Oxford University Press, New York.
- CABRAL, M.C., COLIN, J.P., 2002. Taxonomie et paléocologie de nouveaux ostracodes limniques Candonidae dans l'Oxfordien (Jurassique supérieur) du Portugal. *Geodiversitas*, 24: 61-76.
- CHUNG, F.R.K., HANDJANI, S., JUNGREIS, D., 2003. Generalisations of Polya's urn problem. *Annals of Combinatorics*, 7: 141-154.
- COCKBURN, A., 1991. *An introduction to evolutionary ecology*: 1-370, Blackwell Sci. Publications, Oxford.

- DANIELOPOL, D.L., 1969. Recherches sur la morphologie de l'organe copulateur mâle chez quelques Ostracodes du genre *Candona* (Fam. Cyprididae Baird). In: J.W. Neale (Ed.) *The taxonomy, morphology and ecology of Recent Ostracoda*: 135-153, Oliver & Boyd, Edinburg.
- DANIELOPOL, D.L., 1978. Über Herkunft und Morphologie der süßwasserhypogäischen Candoninae (Crustacea, Ostracoda). *Sitz. Ber. Österr. Akad. Wiss., Math.-Nat. Kl., Abt. I*, 187: 1-162p.
- DANIELOPOL, D.L., 1980. An essay to assess the age of the freshwater interstitial ostracods of Europe. *Bijdr. Dierk.*, 50: 243-291.
- DANIELOPOL, D.L., 1982. Nouvelles données sur les Candoninae (Ostracoda) hypogées de Roumanie et Yougoslavie. *Bull. Mus. Nat. Paris.*, 4^e sér. Section A, 3-4 : 369-396.
- DANIELOPOL D.L., BALTANÁS, A., NAMIOTKO, T., GEIGER, W., PICHLER, M., REINA M., ROIDMAYR, G., 2008. Developmental trajectories in geographically separated populations of non-marine ostracods: morphometric applications for palaeoecological studies. *Senckenbergiana lethaea*, 88: 183-193.
- FOOTE, M., 2011. Evolutionary dynamics of taxonomic structure. *Biol. Lett.* doi:10.1098/rsbl.2011.0521.
- FRECKLETON, R. P., HARVEY, P. H. (2006). Detecting non-Brownian trait evolution in adaptive radiations. *PLoS Biol.*, 4: e373. doi:10.1371/journal.pbio.0040373.
- GAN, L., LI, D., SONG, S., 2006. Is the Zipf law spurious in explaining city-size distributions? *Economics Lett.*, 92: 256-262.
- GASTON, K.J., 1996. Species richness: measure and measurement. In: K. J. Gaston (Ed.) *Biodiversity; a biology of numbers and difference*: 77-113, Blackwell Science London.
- GLEASON, H.A., 1924. Age and area from the view-point of phytogeography. *Am. J. Bot.*, 11: 541-546.
- GUYER, C., SLOWINSKI, J. B. (1993). Adaptive radiation and the topology of large phylogenies. *Evolution*, 47: 253-263.
- HORNE, D.J., 2005. Homology and homeomorphy in ostracod limbs. In: N. Ikeya, A. Tsukagoshi and D. J. Horne (Eds.) *Evolution and diversity of Ostracoda*, *Hydrobiologia*, 538: 55-80.
- HORNE, D. J., DANIELOPOL, D. L., MARTENS, K., 1998. Reproductive behaviour. In: K. Martens (Ed.) *Sex and parthenogenesis. Evolutionary ecology of reproductive modes in non-marine ostracods*: 157-196, Backhuis Publ., Leiden.
- HORNE, D.J., BUNBURY, J., WHITTAKER, J.E., 2011. Taxonomic harmonization and calibration of nonmarine ostracods for palaeoclimate applications: the case of *Candona acutula* Delorme 1967. *Joannea Geol. Paläont.* 11:76-79.
- HUMPHREYS, W.F., 2004. Diversity patterns in Australia. In: D. C. Culver, W.B. White (Eds.), *Encyclopedia of caves*: 183-196. Elsevier-Academic Press, San Diego.
- HUMPHREYS, W.F., WATTS, W.F., BRADBURY, C.H.S., 2004. Emerging knowledge of diversity, distribution and origins of some Australian stygofauna. In: J. Gibert (Ed.) *Symposium on world subterranean biodiversity*: 58-60, Univ. Cl. Bernard Lyon1, Villeurbanne.
- HUXLEY, J.S., 1932. *Problems in relative growth*: 1-276, Methuen, London.
- IEPURE, S., NAMIOTKO, T., DANIELOPOL, D.L., 2008. Morphological diversity and microevolutionary aspects of the lineage *Cryptocandona vavrai* Kaufmann, 1900 (Ostracoda, Candoninae). *Ann. Limnol-Int. J. Lim.*, 44:151-166.
- JOHNSON, N.L., KOTZ, S., KEMP, A.W., 1992. *Univariate discrete distributions*. 2nd Edition: 564p, John Wiley, New York.
- KARANOVIC, I., 2005. On the genus *Typhlocypris* Vojdovsky, 1882 (Crustacea, Ostracoda, Candoninae) with description of two new species. *Syst. and Biodiv.*, 3: 375-406.
- KARANOVIC, I., 2007. *Candoninae (Ostracoda) from the Pilbara region in Western Australia. Crustacea Monogr.*, 7: 433p, Brill, Leiden.
- KAUFMANN, A., 1900. Cypriden und Darwinuliden aus der Schweiz. *Rev. Suisse Zool.*, 8: 209-423.
- KLIE, W., 1938. *Ostracoda, Muschelkrebse. Die Tierwelt Deutschlands und der angrenzenden Meeressteile nach ihren Merkmalen und nach ihrer Lebensweise*: 1-230. Gustav Fischer Verlag. Jena.
- KRSTIĆ, N., 1972. Genus *Candona* (Ostracoda) from Congerian Beds of southern Pannonian Basin. *Monographs of Serbian Academy of Sciences and Arts; Section Natural Sciences – Mathematics*, 39: 1-145p.
- KÜLKÖYLÜOĞLU, O., GIBSON, R., DIAZ, P., COLIN J-P., 2011. *Bicornucandona* gen.nov., sp. nov. (Crustacea, Ostracoda) from Finegan Springs (Texas, U.S.A.). *Zootaxa*, 3059:47-58.
- LÖFFLER, H., DANIELOPOL, D.L., 1978. Ostracoda (Cyprididae). In J. Illies (Ed.) *Limnofauna Europaea*: 196-208, Gustav Fischer Verlag, Stuttgart.
- MARQUET, P.A., ABADES, S.R., LABRA, F.A., 2007. Biodiversity power laws. In D. Storch, P. A. Marquet and J. Brown (Eds.) *Scaling biodiversity*: 441-461, Cambridge University Press, New York.
- MARTENS, K., 1997. Speciation in ancient lakes. *Trends. Ecol. Evol.*, 12: 177-182.
- MARTENS, K., SAVATENALINTON, S., 2011. A subjective checklist of the Recent, free-living, non-marine Ostracoda (Crustacea). *Zootaxa*: 2855: 1-79p.
- MARTENS, K., SCHÖN, I., 1999. Crustacean biodiversity in ancient lakes: a review. *Crustaceana*, 72: 899-910.
- MARTENS, K., WOUTERS, K., MAZEPOVA, G., 1999. Geniation and the genus concept in ancient lakes. *Verh. Internat. Verein. Limnol.*, 27: 2640-2641.
- MARTENS, K., SCHÖN, I., MEISCH, C.L., HORNE, D.J., 2008. Global diversity of ostracods (Ostracoda, Crustacea) in freshwater. *Hydrobiologia*, 595: 185-193.
- MARUVKA, Y.E., KESSLER, D.A., SHNERB, N.M., 2011. The birth-death-mutation process: a new paradigm for fat tail distributions. *PLoS 6* (11): e26480.doi:10.1371/journal.pone.0026480
- MAZEPOVA, G.F., 1990. *Rakushkovye rachki (Ostracoda) Baykala*: 1-472, Nauka Sibirskoe Otdelenie, Novosibirsk.
- MEISCH, C.L., 2000. *Freshwater Ostracoda of Western and Central Europe*. In J. Schwoerbel and P. Zwick (Eds.), *Süßwasserfauna von Mitteleuropa 8/3*: 1-522, Spektrum Akademischer Verlag, Heidelberg.
- MIKULIĆ, F., 1961. Nove Candona vrste iz Ohridskog Jezera. *Bull. Mus. Hist. Nat. Belgrade*, 17 B: 87-107.

- MILLER, A.I., 1997. A new look at age and area: the geographic and environmental expansion of genera during the Ordovician Radiation. *Paleobiology*, 23: 410-419.
- MINELLI, A., FUSCO, G., SARTORI, S., 1991. Self-similarity in biological classifications. *BioSystems*, 26: 89-97.
- MORI, N., OZ, B., KANDUČ, T., KOČMAN, D., 2011. Ostracoda in South-eastern Alps (Slovenia): assemblages in ground waters, springs and adjacent spring brooks. *Joannea Geol. Paläont.*, 11: 147-150.
- MOROCUTTI, U., 2010. *Power laws – Wahrscheinlichkeitstheoretische Modelle und statistische Anwendungen*. Master Thesis: 1-77, Univ. of Salzburg, Inst. of Mathematics, Salzburg.
- NAMIOTKO, T., DANIELOPOL, D.L., BALTANÁS, A., 2011. Soft body morphology, dissection and slide-preparation of Ostracoda: a primer. *Joannea Geol. Paläont.* 11: 327-344.
- NAMIOTKO, T., MARMONIER, P., DANIELOPOL, D.L., 2005. *Cryptocandona kieferi* (Crustacea, Ostracoda): redescription, morphological variability, geographical distribution. *Vie et Milieu*, 55: 91-108.
- NEE, S., 2001. Inferring speciation rates from phylogenies. *Evolution*, 55: 661-668.
- NEE, S., 2005. The neutral theory of biodiversity: do the numbers add up? *Functional Ecol.*, 19: 173-176.
- NEE, S., 2006. Birth-death models in macroevolution. *Ann. Rev. Ecol. Evol. Syst.*, 37: 1-17.
- NEE, S., BARRACLOUGH, T.G., HARVEY, P., 1996. Temporal changes in biodiversity: detecting patterns and identifying causes. In K. J. Gaston (Ed.) *Biodiversity: A biology of numbers and difference*: 230-252, Blackwell Science, London.
- NEE, S., MOERS, A.S., HARVEY, P.H., 1992. Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci. USA*, 89: 8322-8326.
- NEWMAN, M.E.J., 2005. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46: 323-351.
- PEARSON, E.S., 1923. Natural selection and the Age and Area Theory of Dr J. C. Willis. *Biometrika*, 15: 89-108.
- PETERS, R.H., 1983. *The ecological implications of body size*: 1-329, Cambridge University Press, New York.
- PETKOVSKI, T.K., 1969. Einige neue und bemerkenswerte Candoninae aus dem Ohridsee und einigen anderen Fundorten in Europa (Crustacea-Ostracoda). *Acta Mus. Maced. Sci. Nat.*, 11: 81-110.
- PIPIK, R., BODERGAT, A.-M., 2005. Espèces du groupe *Candona candida*, *Candona neglecta* et quelques *Candona* à l'aspect morphologique problématique (Candonidae, Ostracoda) du bassin de Turiec (Miocène supérieur, Slovaquie). *Ann. Paléont.*, 91: 279-309.
- PIPIK, R., BODERGAT, A.-M., 2006. Groupe du *Candona clivosa*, nouveau groupe de Candoninae (Crustacea, Ostracoda) et sa diversification dans le Bassin de Turiec (Slovaquie) au Miocène supérieur. *Geobios*, 39: 394-414.
- REED, W. J., HUGHES, B.D., 2002. On the size distribution of live genera. *J. Theor. Biol.*, 217: 125-135.
- SCHLUTER, D., 2000. *The ecology of adaptive radiation*: 1-218, Oxford University Press, New York.
- SIMKIN, M.V., ROYCHOWDHURY, V.P., 2011. Re-inventing Willis. *Physics Reports*, 502: 1-35p.
- SIMON, H.A., 1955. On a class of skew distribution functions. *Biometrika*, 42: 425-440.
- STOCH, F., GALASSI, D.M.P., 2010. Stygobitic crustacean species richness: a question of numbers, a matter of scale. *Hydrobiologia*, 653: 217-234.
- SZCZECZURA, J., 1978. Freshwater ostracodes from the Nemegt Formation (Upper Cretaceous) of Mongolia. Results of the Polish-Mongolian Paleontological Expedition. Part 3. *Palaeont. Pol.* 38: 65-121.
- TAMBOVTSSEV, Y., MARTINDALE, C., 2007. Phoneme frequencies follow a Yule Distribution (The form of the phonemic distribution in world languages). *J. Theor. Linguistics*, 4: 1-11.
- TONELLI, R., CONCAS, G., LOCCI, M., 2010. Three efficient algorithms for implementing the preferential attachment mechanism in Yule-Simon Stochastic Process. *WSEAS Transact. Information Sci. and Applications*, 7: 176-185.
- TRIEBEL, E., 1963. Ostracoden aus dem Sannois und jüngeren Schichten des Mainzer Beckens: 1. Cyprididae. *Senck. Leth.*, 44: 157-207.
- TURNER, G.F., 1999. Explosive speciation of African cichlid fishes. In: A. Magurran and R. M. May (Eds.), *Evolution of biological diversity*: 130-159, Oxford Univ. Press, Oxford.
- VAN FRAASSEN, B.C., 2000. Sola experientia? Feyerabend's refutation of classical empiricism. In: J. Preston, G. Munévar and D. Lamb (Eds.) *The worst enemy of science? Essays in memory of Paul Feyerabend*: 28-36, Oxford Univ. Press, New York.
- VON FOERSTER, H., 1995. Anthology of principles, propositions, theorems, road-signs, definitions, postulates, aphorisms, etc. <http://www.cybsoc.org/heinz.htm>.
- WALTERS, S.M., 1986. The name of the rose: a review of ideas on the European bias in Angiosperm classification. *New Phytologist*, 104: 527-546.
- WILLIS, J.C., 1922. *Age and area, a study in geographical distribution and origin of species*. Edition 2009, BiblioBazar, LLC.; 274p, <http://books.google.at/books>.
- WILLIS J.C., 1926. Age and Area. *Quart. Rev. Biol.*, 1: 553-571.
- WILLIS, J.C., YULE, G.U., 1922. Some statistics of evolution and geographical distribution in plants and animals, and their significance. *Nature*, 109: 177-179.
- WRIGHT, S., 1941. The course of evolution by differentiation or divergent mutation rather than by selection. *Ecology*, 22: 345-347.
- WOUTERS, K., MARTENS, K., 2001. On the *Cyprideis* species flock (Crustacea, Ostracoda) in Lake Tanganyika, with the description of four new species. *Hydrobiologia*, 450: 111-127.
- YULE, G.U., 1925. A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis. *Philos. Trans. R. Soc. London B*, 213: 21-87. (published 1924).