Extinctions and taxonomy in a trophic model of coevolution

Juan Camacho[1] and Ricard V. Solé[2,3]

[1]*Departament de Física (Física Estadística), Edifici C, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain*
[2]*Complex Systems Research Group, Department of Physics, FEN-UPC, Campus Nord B4, 08034 Barcelona, Spain*
[3]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501*

We investigate the statistics of extinction sizes and the taxonomy in a trophic model of evolution recently proposed [Phys. Rev. Lett. **82**, 652 (1999)]. By further exploring the parameters of this model, we find that the distribution of extinction sizes $N(s)$ shows typically a characteristic maximum before developing the power-law behavior $N(s) \approx s^{-\alpha}$ with $\alpha \approx 2$, in agreement with empirical observations. Furthermore, the derivation of the $\alpha = -2$ exponent given by Drossel [Phys. Rev. Lett. **81**, 5011 (1998)] for this model is completed. The extinction sizes in each trophic level are also analyzed; one finds that at the fourth level and up ($l \geqslant 4$) the extinction size statistics is a power law with exponent $\alpha_l \simeq 1.4$, and exponential-like at the second level, also in agreement with some empirical data not previously explained by current models. On the other hand, in contrast to the observed power-law distribution of the number of species in genera, numerical simulations yield an exponential law. A modification of the model is presented that provides an approximate potential behavior for taxonomy, and some consequences for future modeling are outlined.

PACS number(s): 87.10.+e, 05.40.−a, 05.45.−a

I. INTRODUCTION

The problem of large-scale evolutionary patterns has attracted the attention of physicists in recent years [1]. Several approaches to macroevolution and extinction patterns have been suggested, some based on internal dynamics [2–4] and others on externally driven phenomena [5,6] or mixed descriptions assuming nonstationarity [6,7]. Recently, Amaral and Meyer (AM) have proposed a simplified model for macroevolution based on a trophic, multilayer description of Earth's ecology [8]. The model introduces L trophic levels with a number N of niches each. Living species can create new species at some rate in the nearest levels provided that the randomly chosen niche is empty. Species at the lowest level vanish randomly at rate p, and species at higher levels disappear only if all their preyspecies become extinct (i.e., ecological avalanches are at work).

This model presents some interesting properties, such as a power law for the distribution of extinction sizes, $P(s) \sim s^{-\alpha}$, with $\alpha \simeq 2$, in agreement with available data from the fossil record [9]. This value for the exponent can also be obtained analytically [10] (see Sec. II for a comment on this demonstration). Another interesting property of the model is that time correlations of the number of species also fit well with fossil record databases. Specifically, the number of species displays $1/f$ fluctuations in time [9].

Our aim in this paper is to study numerically the predictions of the trophic model for two main patterns in the fossil record, namely, the statistics of extinction size and the fractality of taxonomy. In this context, it has been shown that the number of genera $G(s)$ formed by s species also scales as $G(s) \approx s^{-\tau}$ where $\tau \approx -2$ for most groups [11].

The paper is organized as follows. Section II deals with the distribution of extinction sizes. A point is made about the proper definition of an extinction event, and the demonstration of the -2 exponent given by Drossel [10] is completed; some minor modifications are also introduced in the model in order to make it slightly more realistic. In Sec. III we analyze the distribution of extinction sizes in each trophic level and compare with some empirical data. Section IV explores the distribution of species in genera, and the last section is devoted to conclusions.

II. STATISTICS OF EXTINCTION SIZES

The AM model is defined by a simple set of rules [8]. First, species at the bottom (basal) layer ($l=1$) are extinct with some probability μ, presumably associated with random external events. Each species in $L>l>1$ layers receives inputs from its immediate lower trophic levels. These connections are not weighted and simply indicate whether or not a species in the $l-1$ layer supports a species in the lth layer. We can indicate the connections between species $i \in l-1$ and species $j \in l$ (here $i,j = 1, \ldots, N$) by $W(i,l-1;j,l) \in \{0,1\}$. If the state of a species $j \in l$ at a given time step t is indicated as $S(j,l;t)$, then the rule for extinction (for $l>1$) is

$$S(j,l;t+1) = \Theta\left(\sum_{i=1}^{N} W(i,l-1;j,l) \right), \qquad (1)$$

where $\Theta(z) = 1$ if $z>0$ and zero otherwise. In other words a species cannot survive if no inputs from lower layers are present. In Fig. 1 we see an example of AM dynamics, for a network with $L=5$ levels and $N=50$ species per level. At $t=2485$ (and starting from a random initial condition) a fully developed ecology is formed, where we see that a large number of species is sustained by the bottom layer. But the extinction of $S(1,13)$ triggers a massive extinction event that propagates to the upper layers [Fig. 1(b)].

The model is completed by defining a diversification process. Diversification occurs as follows: For each surviving species at layer l a nearest layer $l' \in \{l-1,l,l+1\}$ is chosen (except for the top and bottom layers, where only two nearest levels are available). Then a random site $m \in l'$ is also se-
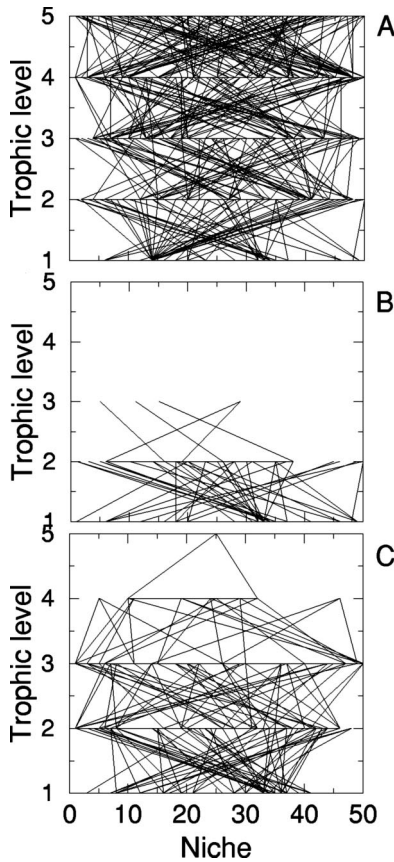
FIG. 1. Example of Amaral-Meyer model of extinction dynamics: a five-level system with $N=50$ niches is shown for $t=2485$ (where we can see a well-established trophic web), at $t=2487$, just after a mass extinction event, and at $t=2500$, recovering.

lected. If empty, it is occupied by a new species, with k prey species chosen at random from the layer below. After large extinction events, upper layers can be empty, but new species are generated and a new, evolving graph of connections develops [Fig. 1(c)]. For very small p's, a power law is obtained. But for larger p's, the distribution of extinction sizes follows a curve that is no longer a simple straight line on a log-log plot (see Fig. 2). It increases up to a maximum and later on decays as a power law with exponent $\alpha \simeq 2$. It is interesting to notice that the fossil record (using a 2-Myr resolution in the histogram) also shows a maximum before developing the power-law tail (inset, Fig. 2) when enough time resolution (here $\delta T=2$ Myr) is considered. The presence of a maximum has an immediate interest in terms of the underlying dynamics responsible for the fossil record. If a pure self-organized critical system were at work, then the power law $N(s) \approx s^{-\alpha}$ *should be recovered at the smallest scales.* This is certainly not the case. It introduces a characteristic extinction size that needs some interpretation.

It is convenient to make a point about the adequate definition of the distribution of extinction sizes. Drossel [10] demonstrated that, when one species in the lowest level dies, the probability for an extinction of size s scales as s^{-2}. One must not confuse this distribution, corresponding to the extinction sizes *per dead species in the lowest level* [let us call it $P_{ds}(s)$], with the distributions plotted in Fig. 2, namely, the distribution of extinction sizes *per time step*, $P(s)$; the second distribution—which is what the fossil record actually
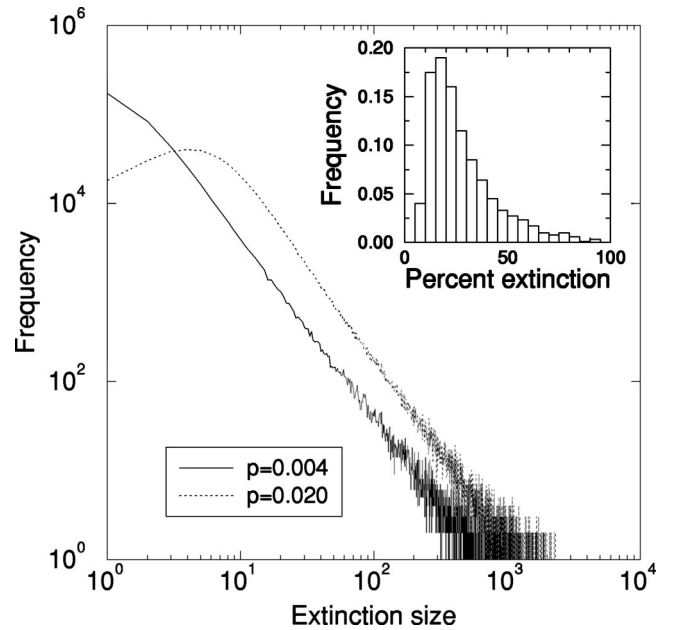

FIG. 2. Distribution of extinction events for the AM model. The values of the parameters are $\mu=0.02$, $p=0.01$ and $\mu=0.004$, $p=0.002$. A maximum is found in the first case, in agreement with the fossil record (see inset).

supplies—adds all the extinctions of the first one along a time step. In this sense, it is worth stressing that the exponent -2 derived theoretically by Drossel is related to, but strictly is not, the exponent of the extinction size distribution per time step, $P(s)$. In order to calculate this second distribution, one has to combine $P_{ds}(s)$, the extinction size distribution for the death of only one species in the lowest level, with the Poisson statistics $P_{Po}(l)$ for the number of extinctions in this level. Even if one neglects the correlations between one extinction and the subsequent ones, the resulting expression, namely,

$$P(s) = \sum_{l=1}^{s} P_{Po}(l) \left\{ \sum_{\Sigma_{j=1}^{l} s_j = s} P_{ds}(s_1) \cdots P_{ds}(s_l) \right\}, \quad (2)$$

is very cumbersome to deal with analytically and, in practice, only numerical simulations can solve it. If one simulates a Poisson process with $P_{ds} \sim s^{-2}$ for the parameters used in our simulations, one finds curves similar to Fig. 2. Therefore, there is no contradiction between the distributions displayed in Fig. 2 [for $P(s)$] and the exponent found by Drossel for $P_{ds}(s)$.

On the other hand, Fig. 2 shows that $P(s)$ decays as a power law with exponent $\alpha \simeq 2$. The value of the exponent is found to be slightly dependent on the death rate p. For $N=1000, \mu=0.02$, and $p=0.01$, one has $\alpha \simeq 2.15$, and for $\mu=0.004$ and $p=0.002$, $\alpha \simeq 1.99$. In both cases, α is very close to 2, and in any case inside the error bounds from the fossil record data.

With these points clarified, one realizes that the demonstration of the $\simeq -2$ exponent developed by $P(s)$ at high enough extinction sizes must still be completed. One may give the following argument. As long as the average number of dead species per unit time in the lowest level is not high,

if one of these deaths gives rise to a large extinction event, it is quite probable that this is the only large extinction that takes place in that time step, since $P_{ds} \sim s^{-2}$ decreases rapidly with s. Therefore, a large extinction of size s during a time step is approximately equivalent to a large extinction of size s per dead species, namely, $P(s) \simeq P_{ds}(s) \sim s^{-2}$.

One can also estimate the location of the maximum in the extinction distribution $P(s)$. We know that when one species in the lowest level dies, the probability for an extinction of size s scales as s^{-2}. This means that 60% of the extinctions triggered by a death in the lowest level are of size 1 [since $\zeta^{-1}(2) \simeq 0.6$]. Then, most of the time a species at the $l=1$ layer disappears, and no further extinction takes place. If N_1 denotes the average number of species in the first level at the stationary state, this implies that most of the time, the number of extinctions in a time step will be $\simeq N_1 p$. It is easy to see that $N_1 \approx N(1-p/\mu)$, so that for $N=1000, \mu=0.02$, and $p=0.01$, $N_1 p \approx 5$; meanwhile, for the case $\mu=0.004$ and $p=0.002$, one has $N_1 p \approx 1$. These predictions for the location of the maxima agree well with numerical simulations, as shown in Fig. 2: the maximum is found at $s \sim 5$ in the first case, but it is absent in the second one.

One major point that makes the AM model unrealistic from the ecological point of view is that species diversify with equal probability into all the available levels, i.e., the parent level, the one immediately above, and the one below. It is more reasonable to consider that the probability of speciation toward lower trophic levels is very small, and that speciation into the same level is much more common than into the others (to generate a new species with similar trophic traits is rather common, to become a higher-level species requires the development of new sophisticated strategies, and the return to a lower level is a very rare event, usually linked with the development of parasitic strategies). To test the AM model's robustness to this biologically sensible criticism, we have considered that speciation takes place only into two levels, the parent level and the one above, the former having a larger probability. One observes that this slight, more realistic modification does not change the generic behavior of the model with respect to the extinction size distribution.

III. EXTINCTION DISTRIBUTION PER TROPHIC LEVEL

One point that can be directly addressed by the model (which has not been previously analyzed in other models to our knowledge) is the statistics of extinctions in each trophic level. This is especially interesting since it can eventually be compared with observations. All previous studies have been focused on comparing the model results with the frequency distribution for the whole fossil record. This is an interesting additonal property that is likely to have a key role in future developments of evolution models: different trophic levels are likely to react in different ways to environmental perturbations and thus display a range of possible statistical patterns. Since information available from the fossil record gives specific data for different trophic groups [13], extinctions per trophic level are highly valuable information.

Let us indicate by $N_l(s)$ the distribution of extinction sizes for the lth level. The pattern produced by the model at different levels is the following. Extinctions in the lowest
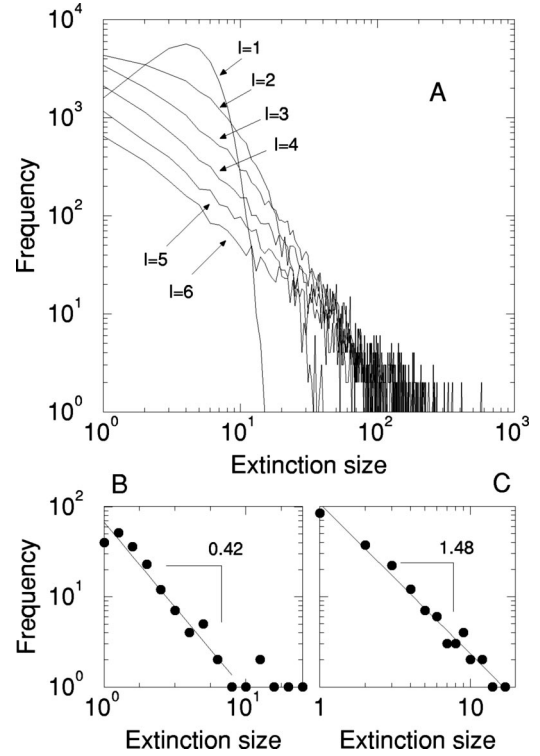

FIG. 3. (a) Distribution of extinction sizes per trophic levels in the AM model. From top to bottom, $l=2$ (approximately exponential), $l=4$, and $l=6$ (potential curve with exponent ≈ 1.4). ($N=1000, k=3, \mu=0.02, p=0.01$.) Frequency of extinction sizes for (b) planktonic foraminifera and (c) ammonites.

level obey a Poisson distribution for $N_l(s)$, consistently with the random nature of the dynamics. At the second level, they satisfy an exponential law $N_2(s) \approx e^{-\epsilon s}$ with $0.13 < \epsilon < 0.25$; and the distributions $N_{l>3}(s)$ clearly exhibit power laws, i.e., $N_{l>3}(s) \approx s^{-\alpha_l}$ with $\alpha_l \approx 1.4$ [Fig. 3(a)].

Numerical simulations show that the value of the exponent is only slightly dependent on the parameters; for $k=10$, the power-law behavior is obtained at lower trophic levels than for $k=3$, and the exponent is again $\beta_l \approx 1.5$. For smaller μ and p values, the exponent decreases slightly.

It is interesting to note that this result is consistent with available data from the fossil record. If we look at well-defined extinct groups, roughly related to trophic levels (it is almost impossible to have a whole group with all species confined to a single level in the trophic chain) we can see that the AM model shows remarkable agreement with them. Data from planktonic foraminifera (which occupy lower levels in marine food chains) are clearly exponentially distributed. This is shown in Fig. 3(b), where we have plotted data from a study by Patterson and Fowler [12], who analyzed the extinction record of planktonic foraminifera from the Jurassic through the Holocene. The observed distribution is in fact close to an exponential and to a Poisson distribution (with better agreement with the first). Data fitting provides an exponent $\epsilon=0.48 \pm 0.04$. Data available from other well-preserved groups, such as ammonites (which have a record spanning 320 Myr) have a power-law behavior with $\beta=1.48 \pm 0.07$ [14], consistently with their higher position in food webs.

IV. TAXONOMY

The fossil record at all taxonomic levels (species, genera, classes, orders, etc.) shows that the frequency distribution of the number of subtaxa (say, species) within taxa (say, genera) follows a power law [11]. These observations are important since the self-similar character of the taxonomy is related to the fact that the tree of life (defined by species that branch into new species) is the result of a given dynamical process. Some models of large-scale evolution have been shown to be able to link the self-similar fluctuations of the extinction record with the self-similarity observed in the taxonomy. This is thus relevant structural information that provides an independent test for theoretical models.

We have performed numerical simulations in order to analyze the frequency distribution predicted by the trophic model. To do so, we have defined a genus as the number of species at the same level having the same parent. The results provide an exponential behavior [Fig. 4(a)], with coefficients in the exponential of order unity: 0.69 for $k=1$, 0.84 for $k=3$, and 0.77 for $k=20$.

This exponential distribution can be obtained analytically for $k=1$. This is a particularly simple case, since the extinction rate is p for all species, and the distribution of lifetimes is described simply by $P_T(t)=pe^{-pt}$. For $k>1$, however, the lifetime distribution is much more complex; in particular, it depends on the trophic level [10].

Since the speciation rate in the parent level is much larger than in the one higher, we will make the approximation in our calculations that all the descendants belong to the same level. Our aim is to evaluate the statistics of the number of descendants per species along its lifespan. Let $P(n)$ be the probability that a species has exactly n descendants when it dies. The probability that a species with lifetime t has n descendants is simply a binomial distribution,

$$P(n,t)=\binom{t}{n}\Pi^n(1-\Pi)^{t-n} \qquad (3)$$

with $t \geq n$, and Π the speciation rate. Let us recall that μ is the rate of speciation attempts, but that speciation actually occurs only if the randomly chosen niche is empty. At the stationary state, speciation rates and death rates are equal, so that $\Pi=p$. Since Π is small (below 0.01), we may approximate $P(n,t)$ by a Poisson distribution. After changing the sum to an integral and some simple calculations, we get

$$P(n)=\sum_{t=n}^{\infty} P_T(t)P(n,t)$$

$$\simeq \int_n^{\infty} pe^{-pt}\frac{(\Pi t)^n}{n!}dt \simeq \frac{p}{p+\Pi}\left(\frac{\Pi}{p+\Pi}\right)^n, \qquad (4)$$

which is an exponential law. This result can also be found by calculating the characteristic function of $P(n)$. Since $\Pi=p$, we get $P(n)\sim e^{-n\ln 2}$. Then the exponent is found to be $\ln 2 \simeq 0.7$, as obtained in the simulations.

Taxonomy in the fossil record follows a power law. Since a power-law distribution implies the absence of a characteristic scale, there cannot exist a characteristic number of offspring per parent species. However, the AM model seems to
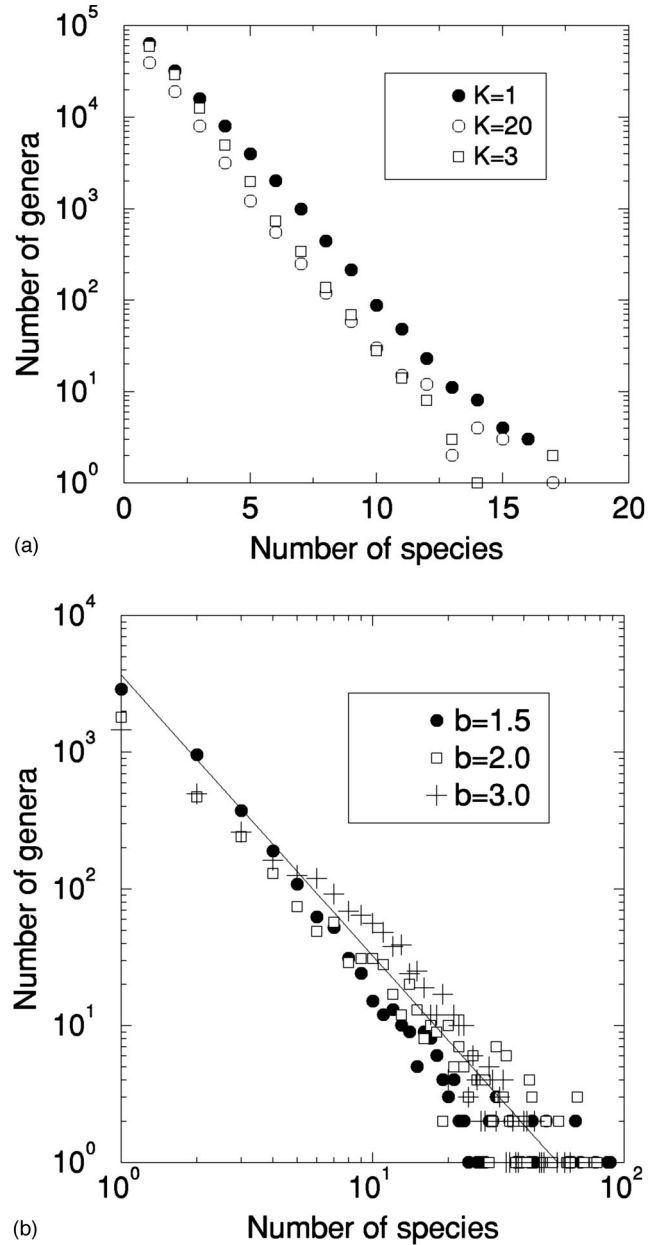

(a)

(b)

FIG. 4. Frequency of species in genera. (a) The trophic model provides an exponential law, in disagreement with the fossil record: $k=1$ (●), $k=3$ (□), and $k=20$ (○); the other parameters are as in Fig. 3. (b) By taking an exponential dependence of the speciation rate on the number of predators, $\mu=\mu_0 b^{np}$, one has approximately power-law curves ($\mu_0=0.004$; three values of b are shown).

provide it. To see this, let us realize that in this model the lifetime of species has a well-defined time scale $\tau \approx k/p$ (in fact, it grows from p^{-1} at the lowest level to k/p at higher levels [10]). Furthermore, since the system reaches a (un-stable) stationary state, the origination rate equals the death rate, namely, $\Pi \simeq \tau^{-1}$. Then the typical number of offspring will be $\tau\Pi \simeq 1$, i.e., a few units, as confirmed by the numerical simulations. Therefore, in order to get a power law in the taxonomy, one has to introduce some *ad hoc* modifications either in the death or in the origination mechanism.

First, we have analyzed the case where the number of preyspecies is not the same for all predators. This changes the lifetime distributions. But even if the number are chosen

at random (between 1 and 100, for instance) or follow a power-law distribution, the offspring frequency distribution is exponential. Secondly, in order to keep somehow the biological resemblance between a species and its progeny, we have assumed that species have the preyspecies of their parents, plus some others, chosen at random up to k preyspecies. The subsequent taxonomy for a wide range of k values does not change: it remains exponential with the same exponents.

Finally, we have considered that the speciation rate increases with the number of predators. This can be justified by the theory that evolutionary rates depend inversely on effective population size [15]. Since, in general, the greater the number of predators, the less their population abundance, then the larger the mutation rates. The results are only partially successful. If one assumes the speciation rate μ to depend exponentially on the predator number np, i.e., $\mu = \mu_0 b^{np}$, one obtains approximately potential curves for $b \simeq 2$ [see Fig. 4(b)] with exponent ≈ -2; the results are better for lower values of μ_0. However, if one takes a linear dependence of μ on np one does not observe a power law. The introduction of a critical number of predators over which a species disappear does not appreciably change this behavior.

V. SUMMARY AND DISCUSSION

We have studied several aspects of a trophic model of macroevolution recently presented by Amaral and Meyer. First, we have seen that it can provide distribution functions for extinction sizes having a maximum before developing a power-law tail with exponent $\alpha = 2$, in agreement with available data from global extinction statistics. Since this expo-

nent has been obtained in other studies also (in both internalist and externalist models) it only partially supports the model accuracy. Analysis of the distribution of extinction sizes in each trophic level yields interesting additional information not previously available from other (nontrophic) models: it shows exponential shapes for the second level, and power laws for higher levels, with exponent ≈ 1.4. Interestingly, these results are consistent with fossil data, such as those studied for Ammonoid families (Ref. [14]).

Finally, we have obtained the frequency distribution of species into a genus predicted by the model; it is exponential, in contrast to the power-law behavior displayed by the fossil record, and in this sense it is a flaw of the model. We have introduced some *ad hoc* modifications in the model in order to recover the fossil data observations. The best results were obtained by considering that the speciation rate depends on the number of predators. If this dependence is exponential, the taxonomy approximately follows a power law, but only for a narrow range of the exponential base. Our study suggests that future models of macroevolution should incorporate the trophic structure present in real ecologies. There is information available for different groups and different spatial and temporal scales that might be of great help for testing the models.

[1] For recent reviews, see R.V. Solé, in *The Statistical Mechanics of Biocomplexity*, edited by D. Reguera, J.M.G. Villar, and J.M. Rubí (Springer, Berlin, 1999), pp. 217–250; see also M. E. J. Newman and R. Palmer, Santa Fe Institute Working Paper No. 99-08-061 (unpublished).

[2] S. Kauffman. *At Home in the Universe* (Oxford U. Press, Oxford, 1994).

[3] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993); K. Sneppen, P. Bak, H. Flyvbjerg, and M.H. Jensen, Proc. Natl. Acad. Sci. USA **92**, 5209 (1995).

[4] R.V. Solé, J. Bascompte, and S.C. Manrubia, Proc. R. Soc. London, Ser. B **263**, 1407 (1996); R.V. Solé and S.C. Manrubia, Phys. Rev. E **55**, 4500 (1996).

[5] R.E. Plotnick and M.L. McKinney, Palaios **8**, 202 (1993).

[6] M.E.J. Newman, Proc. R. Soc. London, Ser. B **263**, 1605 (1996).

[7] P. Sibani, Phys. Rev. Lett. **79**, 1413 (1997).

[8] L.A. Nunes Amaral and M. Meyer, Phys. Rev. Lett. **82**, 652 (1999).

[9] R.V. Solé, S.C. Manrubia, M. Benton, and P. Bak, Nature (London) **388**, 764 (1996); see also M.E.J. Newman and G. Eble, Proc. R. Soc. London, Ser. B **266**, 1267 (1999), for a critical reanalysis. One of the most complete compilations, the Fossil Record 2, can be downloaded from http:// palaeo.gly.bris.ac.uk/frwhole/FR2.html.

[10] B. Drossel, Phys. Rev. Lett. **81**, 5011 (1998).

[11] B. Burlando, J. Theor. Biol. **163**, 161 (1993). In this study, the scaling distribution of whole groups of organisms from different periods of time revealed a consistent power law. Burlando estimated the scaling exponent τ using an appropriate statistical test based on maximum likelihood analysis. He obtained an estimator of τ given by $\bar{\tau} = (N-1)/\Sigma \ln(F_i) + 1$ where $F_i(i = 1,2,\ldots,N)$ is the frequency of genera involving i species.

[12] R.T. Patterson and A.D. Fowler, Geology **24**, 215 (1996).

[13] R.V. Solé and M.E.J. Newman, in *Encyclopaedia of Global Environmental Change*, edited by H. Mooney *et al.* (Wiley, New York, in press). There are a number of well-known groups restricted to specific positions in the food chain that have been studied at different time scales, indicating that both biotic and abiotic responses have played a relevant role. See, for example; D. Jablonski, Science **253**, 754 (1991).

[14] R.V. Solé and J. Bascompte. Proc. R. Soc. London, Ser. B **263**, 161 (1996).

[15] N.H. Barton, in *Speciation and Its Consequences*, edited by D. Otte and J.A. Endler (Sinauer Associates, Inc., Sunderland, MA, 1989), pp. 229–256; J.D. Pelletier, Phys. Rev. Lett. **82**, 1983 (1999).