

# Some properties of the speciation model for food-web structure—Mechanisms for degree distributions and intervality

A.G. Rossberg\*, H. Matsuda, T. Amemiya, K. Itoh

*Yokohama National University, Graduate School of Environment and Information Sciences, Yokohama, Kanagawa 240-8501, Japan*

Received 6 February 2005; received in revised form 26 May 2005; accepted 31 May 2005

Available online 20 July 2005

## Abstract

We present a mathematical analysis of the speciation model for food-web structure, which had in previous work been shown to yield a good description of empirical data of food-web topology. The degree distributions of the network are derived. Properties of the speciation model are compared to those of other models that successfully describe empirical data. It is argued that the speciation model unifies the underlying ideas of previous theories. In particular, it offers a mechanistic explanation for the success of the niche model of Williams and Martinez and the frequent observation of intervality in empirical food webs.

© 2005 Elsevier Ltd. All rights reserved.

**Keywords:** Food-web; Evolution; Network dynamics; Degree distribution; Intervality

## 1. Introduction

The theoretical study of the topology of food webs, the networks formed by the trophic interactions in ecological communities, has led to increasingly precise descriptions of the empirically observed structures. In the early work of Cohen (1978), Briand and Cohen (1987), Sugihara (1983) and others, several simple food-web models had been investigated. The *cascade model* (Cohen et al., 1990) was identified as a description that reproduced the available data particularly well. In the cascade model, a food web consists of a fixed number  $S$  of species, and each species consumes any species which precedes it in a given linear ordering with a fixed probability  $C_0$ . The analysis of this model led to several predictions (Cohen et al., 1990) which inspired a more systematic and accurate collection of food-web data by empiricists (e.g., Hall and Raffaelli, 1991; Havens, 1992; Martinez, 1991; Polis, 1991).

Based on the new data, Williams and Martinez (2000) showed that their *niche model* was a significant improvement. In this model, species are ordered according to their niche value  $n$  that is chosen randomly from the interval  $[0, 1]$ . To determine the diet of a species, an interval of random width  $\leq n$  is drawn with even distribution from within<sup>1</sup>  $[0, 1]$ , restricted by the condition that at least half of the interval is located below the niche value  $n$  of this species. Its diet then consists of all species with a niche value in this interval.

A mathematical analysis by Camacho et al. (2002a) revealed the importance of the specific rule for determining the width of the feeding intervals: by choosing it from an approximately exponential distribution, the resulting food webs show a distribution of generality (the number of a species' resources) which is strongly skewed towards low values, in good accordance with observations (Camacho et al., 2002b; Stouffer et al., 2005).

By construction, the niche model also reproduces a property called *intervality* (Cohen, 1978): Species can be

\*Corresponding author. Tel.: +81 45 339 4369; fax: +81 45 339 4353.

E-mail addresses: [rossberg@ynu.ac.jp](mailto:rossberg@ynu.ac.jp) (A.G. Rossberg), [matsuda2@ynu.ac.jp](mailto:matsuda2@ynu.ac.jp) (H. Matsuda), [amemiyat@ynu.ac.jp](mailto:amemiyat@ynu.ac.jp) (T. Amemiya), [itohkimi@ynu.ac.jp](mailto:itohkimi@ynu.ac.jp) (K. Itoh).

<sup>1</sup>The original description of the model (Williams and Martinez, 2000) is inaccurate at this point.

ordered on a line in such a way that the diet of each consumer is a contiguous set. Intervality is surprisingly often found in small webs (Cohen, 1978). Larger webs exhibit it to some degree (Cattin et al., 2004; Cohen et al., 1990). Cattin et al. (2004) argued that intervality can be a consequence of the fact that similar, evolutionary related species consume similar resources. They proposed the *nested hierarchy model*, a modification of the niche model which incorporates this idea and better accounts for the observed degree of intervality.

Apart from these mostly descriptive models of food-web topology there have also been several attempts to explain the structure of food webs by the interaction of population dynamical and evolutionary mechanisms (e.g., Caldarelli et al., 1998; Drossel et al., 2001; Tokita and Yasutomi, 2003; Yoshida, 2003). Characteristic for most of these models is their high computational complexity, which makes their quantitative statistical validation difficult. Therefore, it can be advantageous to consider first explanatory models that are explicit in terms of either population dynamics (e.g., Montoya and Solé, 2003; Pimm, 1984) or evolutionary mechanisms (e.g., Amaral and Meyer, 1999; Camacho and Solé, 2000; Drossel, 1998) alone.

The recently proposed *speciation model* (Rossberg et al., 2005) is of the purely evolutionary type. It combines mechanisms corresponding to speciations and extinctions with simple assumptions regarding the evolutionary inheritance of trophic links. In spirit, the model is similar to the duplication-divergence model of proteome evolution by Vazquez et al. (2003) or the related model by Pastor-Satorras et al. (2003), even though in the speciation model directed links and the possibility of extinctions complicate the situation.

Furthermore, the speciation model takes the tendency of food webs to respect a “pecking order”, as it is ideally realized in the cascade model, into account. It is currently unclear if the dominating mechanism imposing this ordering of species is the physical advantage that larger predators have over smaller prey, energy conservation and dissipation, or some other constraint. The idea that the pecking order is essentially an ordering by body size has often been discussed (Cattin et al., 2004; Cohen et al., 1993; Memmott et al., 2000; Warren and Lawton, 1987). The speciation model makes this hypothesis explicit by postulating an allometric relationship between body sizes and evolution rates.

The speciation model has been validated by a systematic statistical analysis based on a comparison of 12 model properties—such as the average chain length, the fraction of top predators, the degree of intervality, or the clustering coefficient—with empirical data (Rossberg et al., 2005). These numerical results suggest that the speciation model reproduces observed food-web properties even better than the niche model or the nested hierarchy model. The aim of the current work

is to present some analytic results that allow insights into how important food-web properties derive from the model specifications. After stating the model definition in Section 2, the steady-state distribution of the number of species  $S$  and the expectation value of the directed connectance  $C$  (sometimes referred to as the food-web “complexity”) are derived in Section 3. These quantities are important because they are used as control parameters in other models. Section 3 also contains a characterization of the species pool in terms of evolutionary “clades” which invites a comparison with empirical data. Section 4 is devoted to a characterization of the model in terms of the distributions of generality and vulnerability (the number of a species’ consumers). Based on these results, the speciation model is compared with the cascade model, the niche model, and the nested hierarchy model in Section 5; common properties and differences are pointed out. Two variants of the speciation model, which leave the analytic properties derived below unchanged, are introduced in Section 6. A discussion and interpretation of the results is provided in Section 7.

## 2. Definition of the speciation model

This section restates the definition of the speciation model given elsewhere (Rossberg et al., 2005), since it will be the starting point for the subsequent analysis. For a motivation of the model and a discussion of design decisions we refer to the original work. The speciation model describes an abstract species pool, the set of trophic links between the species, and the evolution of both. The model is described in terms of a stochastic process characterized by the parameters  $r_1, r_+, r_-, R, D, \lambda, C_0$ , and  $\beta$ .

### 2.1. The evolution of the species pool

Each species  $i$  in the pool is associated with a *speed parameter*  $s_i$  in the range  $[0, R]$ . The speed parameter characterizes the evolution rate of a species and is thought to be inversely correlated with the logarithm of the species’ body mass by an allometric law (see Rossberg et al. (2005) for discussion). In any infinitesimal time interval  $[t, t + dt]$  three kinds of events can occur: *adaptations* of foreign species to the habitat (i.e. invasions on an evolutionary time scale), *extinctions*, and *speciations*. The probability for the adaptation of a new species  $k$  with speed parameter in the infinitesimal range  $s_k \in [s, s + ds]$  is  $r_1 \exp(s) ds dt$ . When a new species is adapting to the habitat, it is added to the species pool. The probability that some species  $i$  of the species pool becomes extinct is  $r_- \exp(s_i) dt$ . When a species becomes extinct, it is removed from the species pool. Finally, the probability that some species  $i$  from

the species pool speciates is  $r_+ \exp(s_i) dt$ . When  $i$  speciates, a new species  $j$  with speed parameter  $s_j = s_i + \delta$  is added to the species pool, where  $\delta$  is a zero-mean Gaussian random variable with  $\text{var } \delta = D$ . If  $s_i + \delta$  exceeds the range  $[0, R]$ ,  $s_j = -(s_i + \delta)$  or  $s_j = 2R - (s_i + \delta)$  are used instead (reflecting boundaries). The probabilities for any of these events to occur are independent.

### 2.2. The evolution of the food web

The food web is described by a connectivity (or adjacency) matrix  $(m_{ij})$ , with connectivity values  $m_{ij} = 1$  when  $j$  eats  $i$  and  $m_{ij} = 0$  otherwise. Possible consumers  $l$  of species  $i$  are defined as species with  $s_l < s_i + \lambda R$ , possible resources  $h$  as those with  $s_h > s_i - \lambda R$ . The connectivity  $m_{ij}$  can be 1 only when  $i$  is a possible resource of  $j$ . The connectivity of a new species adapting to the habitat to all possible consumers and resources is set to 1 with probability  $C_0$  and to 0 otherwise. Upon speciation, the connectivity values of the decedent species  $j$  to possible consumers and resources are copied from the corresponding connectivity values of the parent species  $i$  with probability  $1 - \beta$  (i.e., links break with probability  $\beta$ ). The connectivity values to all possible resources and consumers of  $j$  which have not been copied are set to 1 with probability  $C_0$  and to 0 otherwise.

### 2.3. Typical parameters

In our previous study (Rossberg et al., 2005) the predictions of the speciation model were compared to empirical data, and maximum likelihood fits of the model to empirical data sets for fixed  $R = \ln 10^4$ ,  $D = 0.0025$ ,  $r_- = 1$  were computed. For brevity, we refer to these parameter sets as “typical values” hereafter. For the convenience of the reader the fitted values are listed in Table 1 together with some derived expressions relevant for the calculations below.

## 3. Basic statistical properties of the model steady state

The number  $S$  of species in a food web and the number  $L$  of trophic links connecting them belong to the simplest quantities used to characterize food webs. Often  $L$  is expressed in terms of the directed connectance  $C = L/S^2$  or related quantities. In what follows, the steady-state distribution of  $S$  and the expectation value of  $C$  for the speciation model are derived. For these calculations, it is helpful to imagine the species pool as being divided into clades. Following Yoshida (2002, 2003), a *clade* is here defined as the group of all currently existing descendant species of a *founder species* that entered the species pool through an adaptation process,

in close correspondence with the standard phylogenetic notion. When  $D$  is sufficiently small, the speed parameter  $s$  is approximately the same for all species in a clade, and the ranges of  $s$  covered by different clades do not overlap. We can then divide the  $s$ -axis into small intervals  $[s, s + \Delta s]$ , and account for the number of species in each interval separately. The absence of overlap between clades is used only as a trick to simplify accounting. The final results do not depend on this assumption. The condition that the spread of  $s$  within clades is small will be made more precise in the detailed discussion of the clades in Section 3.2 below.

### 3.1. The steady-state distribution of the species number $S$

In order to obtain the steady-state distribution of the total number of species, consider first only a small interval  $[s, s + \Delta s]$  on the speed-parameter axis. The master equation for the probability distribution  $p_n$  of the number  $n$  of species in the interval is given by

$$\frac{dp_n}{dt} = j_{n-1,n} - j_{n,n+1} \tag{1}$$

for  $n \geq 1$  and

$$\frac{dp_0}{dt} = j_{0,1}, \tag{2}$$

with the probability current  $j_{n,n+1}$ , resulting from the balance of processes incrementing and decrementing  $n$ , given by

$$j_{n,n+1} = e^s [(nr_+ + r_1 \Delta s) p_n - (n+1)r_- p_{n+1}]. \tag{3}$$

The possibility of speciations that cross the boundaries of the range  $[s, s + \Delta s]$  is ignored here, because the corresponding corrections would cancel out when summing up the  $n$  values from different intervals below. The reflecting boundary conditions at the endpoints of the full  $s$ -range  $[0, R]$  ensure that Eq. (3) holds also for the intervals adjacent to the endpoints.

For the steady state  $j_{n,n+1} = 0$  one gets

$$p_1 = \frac{r_1}{r_-} p_0 \Delta s \tag{4}$$

and for  $n \geq 1$  the recursive relation

$$p_{n+1} = \frac{nr_+}{(n+1)r_-} p_n + \mathcal{O}(\Delta s), \tag{5}$$

which is solved by

$$p_n = \frac{1}{n} \left( \frac{r_+}{r_-} \right)^n \frac{r_1 p_0}{r_+} \Delta s + \mathcal{O}(\Delta s^2). \tag{6}$$

With the abbreviations  $\rho = r_+/r_-$  and  $\kappa = r_1/r_+$ , the corresponding moment generating function is

$$m(z) = \langle z^n \rangle = p_0 [1 - \kappa \Delta s \ln(1 - \rho z)] + \mathcal{O}(\Delta s^2), \tag{7}$$

with

$$p_0 = 1 + \kappa \Delta s \ln(1 - \rho) + \mathcal{O}(\Delta s^2) \tag{8}$$

Table 1

Maximum-likelihood model parameters for the speciation model obtained for seven empirical food webs and quantities derived thereof

Food web	BB	Sk	Co	Ch	SM	Yth	LR
<i>Model parameters</i>							
$r_+$ ( $= \rho$ )	0.914	0.934	0.961	0.959	0.801	0.949	0.991
$r_1$	0.17	0.21	0.13	0.21	0.92	0.67	0.13
$\lambda$	0.12	0.082	0.006	0.25	0	0.001	0.025
$C_0$	0.37	0.53	0.58	0.064	0.23	0.081	0.16
$\beta$	0.059	0.012	0.014	0.029	0.034	0.040	0.0063
<i>Derived quantities</i>							
Web size (before lumping) $\langle S \rangle$	18.2	29.0	31.4	47.9	42.7	122.0	137.4
$\text{var}S/\langle S \rangle^2$	0.64	0.53	0.81	0.51	0.12	0.16	0.81
Clade size $\langle n \rangle$ : Eq. (14)	4.3	5.2	7.6	7.3	2.5	6.3	23.5
Number of clades $\langle c \rangle$ : Eq. (15)	4.2	5.5	4.2	6.6	17.1	19.5	5.8
Clade lifetime in generations: $-\ln(1 - \rho)$	2.5	2.7	3.2	3.2	1.6	3.0	4.7
Clades in diet: Eq. (33), $A = R$	2.3	3.2	2.8	0.7	4.5	2.8	1.5
Diet breakout: Eq. (55)	0.44	0.16	0.31	0.41	0.12	0.43	0.44

The abbreviations stand for BB: Bridge Brook Lake (Havens, 1992), Sk: Skipwith Pond (Warren, 1989), Co: Coachella Desert (Polis, 1991), Ch: Chesapeake Bay (Baird and Ulanowicz, 1989), SM: St. Martin Island (Goldwasser and Roughgarden, 1993), Yth: Ythan Estuary (Hall and Raffaelli, 1991), LR: Little Rock Lake (Martinez, 1991).

given by the normalization condition  $m(1) = 1$ . From  $m(z)$  one obtains the cumulant generating function

$$k(z) = \ln m(z) = \ln p_0 - \kappa \Delta s \ln(1 - \rho z) + \mathcal{O}(\Delta s^2) \\ = \kappa \Delta s \ln \left( \frac{1 - \rho}{1 - \rho z} \right) + \mathcal{O}(\Delta s^2). \quad (9)$$

Cumulant generating functions of this form and the corresponding distributions are discussed in Appendix A. For example, by Eq. (A.2), the density of species along the speed-parameter line is

$$\lim_{\Delta s \rightarrow 0} \frac{\langle n \rangle}{\Delta s} = \frac{\kappa \rho}{1 - \rho} = \frac{r_1}{r_- - r_+}. \quad (10)$$

The cumulant generating function of the sum of independent random variables is the sum of their cumulant generating functions. Thus, the cumulant generating function for the total number of species  $S$  can be obtained by dividing the range  $[0, R]$  into small intervals of width  $\Delta s$  and summing the contributions. With  $\Delta s \rightarrow 0$  corrections  $\mathcal{O}(\Delta s^2)$  become negligible and the summation goes over into an integration

$$\sum \frac{k(z)}{\Delta s} \Delta s + \mathcal{O}(\Delta s^2) \rightarrow \int_0^R \kappa \ln \left( \frac{1 - \rho}{1 - \rho z} \right) ds \\ = \kappa R \ln \left( \frac{1 - \rho}{1 - \rho z} \right). \quad (11)$$

This is again of the general form Eq. (A.1) discussed in Appendix A. Hence, the steady-state distribution of the species number  $S$  is  $P(\kappa R, \rho; S)$  as defined by Eq. (A.4). Fig. 1 shows a typical distribution and corresponding simulation results. The curves agree well. Only the probability for  $S$  near zero seems to be overestimated by

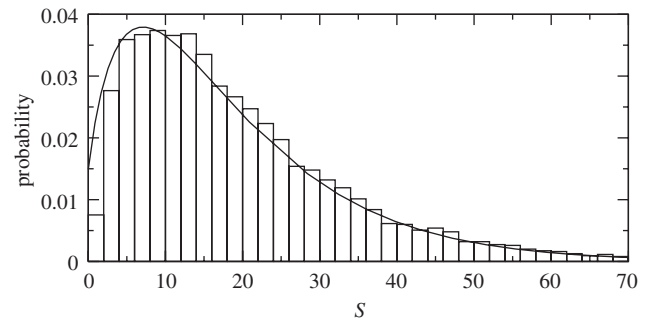


Fig. 1. Typical steady-state distribution of the number of species  $S$ . The solid line is  $P(\kappa R, \rho; S)$  as defined by Eq. (A.4); the histogram was obtained by direct simulations. Parameters correspond to Bridge Brook Lake (Table 1).

the theory. By Eq. (A.2), the mean number of species is

$$\langle S \rangle = \frac{\kappa R \rho}{1 - \rho}, \quad (12)$$

and by Eq. (A.3) the relative variance ( $\text{var}S/\langle S \rangle^2 = 1/\kappa R \rho$ ). Typical relative variances (Table 1) can become of the order unity. Thus, in the model,  $S$  fluctuates strongly on evolutionary time-scales.

### 3.2. Basic properties of clades

The division of  $S$  into clades can be made more explicit. For example, the distribution of the number  $n$  of species in a single clade is given by Eq. (6) conditional to  $n \geq 1$ :

$$p_n = -\frac{\rho^n}{n \ln(1 - \rho)}. \quad (13)$$

Thus, the mean number of species per clade is

$$\langle n \rangle = \sum_n n p_n = -\frac{\rho}{(1-\rho) \ln(1-\rho)}. \quad (14)$$

Further, the expectation value of the number of clades  $c$  in the food web can be estimated as

$$\langle c \rangle = \frac{\langle S \rangle}{\langle n \rangle} = -\kappa R \ln(1-\rho). \quad (15)$$

(An exact calculation yields the same result.) Since appearances and extinctions of clades are statistically independent, the number of clades is Poisson distributed. For typical values of  $\langle n \rangle$  and  $\langle c \rangle$  see Table 1.

To obtain the average lifetime  $\tau_c$  of a clade founded by a species with speed parameter  $s$ , note that the probability that a clade exists in the interval  $[s, s + \Delta s]$  is  $1 - p_0$  with  $p_0$  given by Eq. (8). On the other hand, new clades are founded at a rate  $r_1 \sigma \Delta s$  with  $\sigma = \exp(s)$ . The fraction of time when there is a clade in the interval is thus  $\tau_c r_1 \sigma \Delta s$ . (Note that in the limit  $\Delta s \rightarrow 0$  there is no overlap in the clade lifetimes.) Thus

$$\tau_c = \lim_{\Delta s \rightarrow 0} \frac{1 - p_0}{r_1 \sigma \Delta s} = -\frac{\ln(1-\rho)}{r_+ \sigma}. \quad (16)$$

The time that it takes for the system to reach the steady state can be estimated by the lifetime of the slowest clade, i.e., by Eq. (16) with  $\sigma = \exp(0) = 1$ . This quantity is important for model simulations. For a detailed discussion of the dynamics of the birth/death process relevant here, including the clade lifetime distribution, see the book of Bailey (1964).

The typical number of evolutionary “generations” that a clade exists is  $\tau_c / (\text{generation time}) = \tau_c r_+ \sigma = -\ln(1-\rho)$  (see Table 1 for typical values). Since in each generation the variance of the distribution of  $s$  over a clade increases by  $D$ , the width of a clade on the speed-parameter line is of the order

$$\text{std } s \approx \sqrt{-D \ln(1-\rho)}. \quad (17)$$

The assumption made above that all members of a clade have approximately the same  $s$  is justified when  $s \ll 1$ .

### 3.3. The expected directed connectance

A food-web property that has found much attention in both empirical and theoretical research is the connectance, for example measured in terms of the directed connectance  $C = L/S^2$  (Martinez, 1991) with  $L$  denoting the total number of trophic links. To compute the expectation value of this quantity, note that from all  $S^2$  topologically possible links only some are allometrically possible in the model, namely those from consumers  $i$  to their possible resources  $h$  with  $s_h > s_i - \lambda R$  (see Section 2). A fraction  $(1 - \lambda)^2/2$  of the  $s_h - s_i$  plane is forbidden. By construction, exactly a fraction  $C_0$  of all allometrically possible links is realized on the

average in the model. Thus, as a simple estimate one gets  $S^2[1 - (1 - \lambda)^2/2] = S^2(1 + 2\lambda - \lambda^2)/2$  allometrically possible links and

$$C \approx C_0(1 + 2\lambda - \lambda^2)/2. \quad (18)$$

The exact value differs due to subtle correlations stemming from intra-clade links. As an example, we derive  $C$  for the case that the typical intra-clade spread of  $s$  given by Eq. (17) is much smaller than  $\lambda R$ , so that all intra-clade links are allometrically possible. As in Section 3.1, we divide the  $s$  axis into small intervals of width  $\Delta s$ , and do again as if each clade was located in its own interval. Let the  $p$ th interval range from  $s_p$  to  $s_p + \Delta s = s_{p+1}$  and denote the number of species it contains by  $n_p$ . We first compute the expected number of allometrically possible links conditional to fixed  $S$

$$\langle L_{al} | S \rangle = \sum_{\substack{p,q \\ s_p > s_q - \lambda R}} \langle n_p n_q | S \rangle = \sum_{\substack{p \neq q \\ s_p > s_q - \lambda R}} \langle n_p n_q | S \rangle + \sum_p \langle n_p^2 | S \rangle. \quad (19)$$

Consider the last term first. The distribution  $p_n$  of  $n_p$  is given by Eqs. (6), (8). Since clades appear and disappear independently, the probability that there are  $S - n_p$  species outside the  $p$ th interval is, just as for the total number of species,  $P(\kappa R, \rho; S - n_p)$ , defined by Eq. (36) to lowest order in  $\Delta s$ . The probability for a particular pair  $(n_p, S)$  is therefore  $p_{n_p} P(\kappa R, \rho; S - n_p)$ . This can be used to calculate the probability  $p(n_p | S)$  of  $n_p$  conditional to  $S$  in the usual way, giving

$$\begin{aligned} \langle n_p^2 | S \rangle &= \sum_{n=0}^S n^2 p(n | S) = \sum_{n=0}^S n^2 \frac{p_n P(\kappa R, \rho; S - n)}{P(\kappa R, \rho; S)} \\ &= \frac{S(\kappa R + S)}{R(1 + \kappa R)} \Delta s + \mathcal{O}(\Delta s^2). \end{aligned} \quad (20)$$

The dependence on  $\rho$  drops out. By a similar argument one obtains to lowest order in  $\Delta s$ ,

$$\langle n_p n_q | S \rangle = \sum_{m+n \leq S} n m p(m, n | S) = \frac{\kappa(S-1)S}{R(1 + \kappa R)} \Delta s^2. \quad (21)$$

Inserting both results into Eq. (19) and taking the limit  $\Delta s \rightarrow 0$  yields

$$\langle L_{al} | S \rangle = S \frac{S + \kappa R [1 + \frac{1}{2}(1 + 2\lambda - \lambda^2)(S - 1)]}{1 + \kappa R} \quad (22)$$

$$= S^2 \left[ \frac{1 + \kappa R \frac{1}{2}(1 + 2\lambda - \lambda^2)}{1 + \kappa R} + \mathcal{O}\left(\frac{\kappa R}{S}\right) \right]. \quad (23)$$

Expression (23) is often a good approximation of Eq. (22). The expected directed connectance conditional to  $S$  is  $\langle C | S \rangle = C_0 \langle L_{al} | S \rangle / S^2$ . Dropping the undefined case  $S = 0$ , the expected connectance for freely fluctuat-



ing  $S$  can be evaluated as

$$\langle C \rangle = C_0 [1 - P(\kappa R, \rho; 0)]^{-1} \sum_{S=1}^{\infty} \frac{\langle L_{al} | S \rangle}{S^2} P(\kappa R, \rho; S), \tag{24}$$

either directly numerically or, for a (complicated) closed-form expression, with the help of symbolic algebra software. For the parameters of Bridge Brook Lake (Table 1), for which  $\lambda R / \text{std } s = 14.7$ , Eq. (24) yields  $\langle C \rangle = 0.294$  while the simple estimate (18) gives  $\langle C \rangle = 0.230$ . Simulations yield  $\langle C \rangle = 0.286$ . The cases that  $\lambda R > 0$  is much smaller than the typical intra-clade spread of  $s$  and that  $\lambda = 0$  (no cannibalism) can be handled by replacing  $n^2$  in Eq. (20) by  $n(n+1)/2$  or  $n(n-1)/2$ , respectively. For both cases the approximation  $\langle C | S \rangle = \frac{1}{2} C_0 [1 + \kappa R (1 + 2\lambda - \lambda^2)] / (1 + \kappa R) + \mathcal{O}(S^{-1})$  holds. For the parameters of St. Martin Island ( $\lambda = 0$ ) this yields  $\langle C \rangle = 0.115$ , while numerically  $\langle C \rangle = 0.112$  is obtained.

#### 4. The distributions of generality and vulnerability

In this section, analytic approximations for the distributions of generality  $k$  (the number of resources of a consumer) and vulnerability  $m$  (the number of consumers of a resource) are derived. When defining the direction of trophic links in the standard way from the resource to the consumer, these are the distributions of the in-degree and the out-degree of the food web, respectively. Degree distributions are often thought to belong to the major determinants of the overall network topology. Due to the inherent randomness of food webs and their finite size, instances of degree distributions of empirical or model webs are also random quantities. Nevertheless, they contain information regarding the probability distributions of generality  $P_{gen}(k)$  and vulnerability  $P_{vul}(m)$  in the steady state. Specifically, if  $N(k)$  denotes the number of species with generality  $k$  in a web and the total number of species is  $S$ , then  $\langle N(k) / S \rangle = P_{gen}(k)$  in the steady state. While this is trivial for fixed  $S$ , it is worth noting that this relation is valid also when the value of  $S$  fluctuates randomly and when the generalities of individual species are strongly correlated with each other and with  $S$ , as can be seen by a straightforward calculation. Below it is shown that the conditional probability  $P_{gen}(k|S)$ , i.e. the conditional expectation value  $\langle N(k) / S | S \rangle$ , does in fact strongly depend on  $S$ . For a comparison with single instances of empirical distributions  $N(k) / S$  the conditional distribution  $P_{gen}(k|S)$  is therefore better suited than  $P_{gen}(k)$ . Similar considerations hold for the vulnerabilities. Thus, the conditional distributions are computed below.

Following Camacho et al. (2002a), we consider the distinguished limit of large food-web sizes  $S$  and small

connectances  $C$  while keeping the link density  $Z := L/S = CS$  fixed. (Fixing  $Z$  for asymptotic expansions is not meant to suggest that  $Z$  is actually fixed for large food webs.) For simplicity, we make use of the hypothesis that resources typically evolve faster than their consumers in the extreme form that resources evolve *much* faster than their consumers. This corresponds to assuming a large spread of time scales  $R$  and a small loopiness  $\lambda$ . Errors due to intra-clade trophic links, which violate this hierarchy of timescales, are small when the total number of clades (15) is large, due either to large  $\kappa R$  or to small  $1 - \rho$ . We note that in the case  $\kappa R \gg 1$  the combined effect of these assumptions would reduce the formula for the directed connectance derived above to  $\langle C \rangle = C_0 / 2$ , which shows that the approximations employed here are much coarser than those used in the foregoing sections. Nevertheless, they retain the main effects that determine the general forms of the degree distributions.

##### 4.1. Reduction to the dynamics of the actual resources

When most resource species evolve much faster than their consumers, the distribution of generality for a given consumer can be approximated by the steady-state generality distribution with the consumer assumed fixed while its resources evolve. We first show that, using a simple mean-field-type approximation, the stochastic dynamics of the actual resources of the fixed consumer can be separated from the dynamics of the possible resources which are not actual resources (called *spurned resources* below) in a self-consistent way.

To derive the dynamics of the actual resources, consider a small interval  $[s, s + ds]$  in the range of possible resources. Let  $\sigma = \exp(s)$ . The rate at which actual resource species in the interval speciate in such a way that the descendant species remain actual resources is  $r_+^* \sigma$  with

$$r_+^* = (1 - \beta)r_+ + \beta C_0 r_+. \tag{25}$$

The first term accounts for trophic links that do not break in the speciation, and the second term for trophic links that break but are immediately reconnected. The probability that a resource species becomes extinct in a time interval of length  $dt$  is simply  $r_-^* \sigma dt$  with

$$r_-^* = r_-. \tag{26}$$

Finally, the consumer can acquire a novel resource species either by an adaptation of a new species to the habitat or by a speciation of a spurned resource in such a way that the decedent species becomes an actual resource. For the rate at which the latter event occurs, a mean-field type approximation is employed: The number of spurned resources  $n^o$  in the speed-parameter range  $[s, s + \Delta s]$  is approximated by its expectation value  $\langle n^o \rangle$ . The rate at which a predator acquires novel

resources (that did not speciate from an existing resource species) in this range is then given by  $C_0 r_1^* \sigma \Delta s$  with

$$r_1^* = r_1 + \frac{\beta \langle n^o \rangle r_+}{\Delta s}. \tag{27}$$

The first term represents new adaptations, the second term mutations of spurned species. With this approximation, the expectation value for the number  $n^*$  of actual resources in the range  $[s, s + \Delta s]$  can be calculated as

$$\langle n^* \rangle = \frac{C_0 r_1^*}{r_-^* - r_+^*} \Delta s \tag{28}$$

by methods analogue to those used in Section 3.1. Deviations from this mean-field approximation occur because the expectation value  $\langle n^o \rangle$  is correlated to  $n^*$  by the breaking of actual links, which occurs at a rate  $\mathcal{O}(\beta)$ . Since the contribution of  $\langle n^o \rangle$  to the dynamics of  $n^*$  is also of order  $\mathcal{O}(\beta)$ , the resulting error in the distribution of  $n^*$  is  $\mathcal{O}(\beta^2)$ .

For the dynamics of the number of spurned resource, a set of equations corresponding to Eqs. (25)–(28) can be set up by replacing  $C_0 \rightarrow 1 - C_0$  and interchanging the indices  $*$  and  $o$  (25–28). These equations can be used to eliminate  $\langle n^o \rangle$  from Eq. (27), yielding

$$r_1^* = r_1 + \frac{\beta(1 - C_0)r_+}{r_- - r_+} r_1. \tag{29}$$

#### 4.2. The generality distribution for fluctuating $S$

Analogous to the calculations of Section 3.1, the cumulant generating function for the number of actual resources for a species with speed parameter  $s$  can now be obtained as

$$K_{gen}(s, z) = C_0 \kappa^* A(s) \ln \left( \frac{1 - \rho^*}{1 - \rho^* z} \right), \tag{30}$$

where  $\rho^* = r_+^*/r_-^*$  and  $\kappa^* = r_1^*/r_+^*$  are given by Eqs. (25), (26), (29), and  $A(s) = \min[(1 + \lambda)R - s, R] \approx R - s$  is the size of the speed-parameter range of possible resources. The corresponding distribution function is

$$P_{gen}(s, k) = P(C_0 \kappa^* A(s), \rho^*; k) \tag{31}$$

as defined by Eq. (A.4). In particular, the expected number of a consumer’s resources is

$$\langle k \rangle = \frac{C_0 \kappa^* A(s) \rho^*}{1 - \rho^*} = C_0 A(s) \frac{r_1}{r_- - r_+}. \tag{32}$$

Comparison with Eq. (10) shows that the mean-field approximation preserves the model property that, on the average, a fraction  $C_0$  of all allometrically possible links is realized. Just as for the overall species pool, the diet of a consumer can be divided into several clades, each descending from a single newly acquired resource. For example, the expected number of resource clades for a

consumer is

$$-C_0 \kappa^* A(s) \ln(1 - \rho^*), \tag{33}$$

in analogy to Eq. (15).

Since, on the average, species are homogeneously distributed along  $s$ , the probability distribution  $P_{gen}(k)$  of the generality of a species chosen arbitrarily from a food web is, to a good approximation, the average of  $P_{gen}(s, k)$  over  $s$ . Analytically, this average is more easily calculated in terms of the moment generating function  $M_{gen}(s, z) = \sum_k P_{gen}(s, k) z^k = \exp K_{gen}(s, z)$ . For the simple case  $\lambda = 0$  one obtains

$$M_{gen}(z) = \frac{1}{R} \int_0^R M_{gen}(s, z) ds = \frac{u - 1}{\log u} \quad \text{with} \tag{34}$$

$$u = \left( \frac{1 - \rho^*}{1 - \rho^* z} \right)^{C_0 \kappa^* R}.$$

The generality distribution  $P_{gen}(k)$  itself can be calculated by a Taylor expansion of  $M_{gen}(z)$  in  $z$  or numerically from the Fourier transformation of  $\text{Re}\{M_{gen}(e^{i\phi})\}$ . A comparison with direct numerical simulations shows that the condition that  $R$  is large is important for the numerical validity of Eq. (34). For example, Fig. 2 shows analytic and numerical results for  $R = \ln 10^{20}$  in good agreement.

#### 4.3. The generality distribution for fixed $S$

In order to compute the generality distribution  $P_{gen}(k|S)$  conditional to fixed  $S$ , we start again from the distribution  $P_{gen}(s, k|S)$  for a consumer with speed parameter  $s$ . In order to simplify the calculations  $\beta = 0$  is assumed here. Then  $\kappa^* = \kappa^o = \kappa$  and  $\rho^* = \rho^o = \rho$ .

For a given consumer, the species pool can be divided into three subsets: (i) the actual resources of the consumer, (ii) the allometrically possible but spurned resources, and (iii) the allometrically forbidden resource

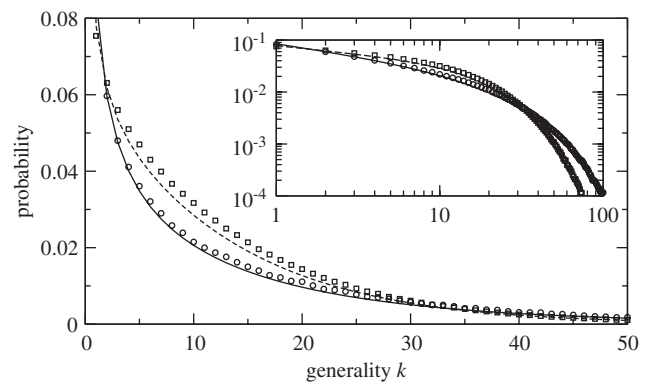


Fig. 2. Steady-state generality distributions for fluctuating species number, calculated from Eq. (34) with  $\beta = 0$  (solid) and  $\beta = 0.05$  (dashed) in comparison with direct numerical simulations (circles, squares). The other parameters were  $R = \ln 10^{20}$ ,  $D = 0.005$ ,  $\rho = 0.95$ ,  $\kappa = 10/R$ ,  $C_0 = 0.1$ ,  $\lambda = 10^{-3}$  (no fitting). The inset shows the same data on a double-logarithmic scales.

(see Section 3.3). For small enough  $D$ , each clade is located in a single subset, and the species distributions in the three subsets become independent. We first calculate the probability distribution for the number of species in the union of sets (ii) and (iii) for freely fluctuating  $S$ . As above, denote the width of the range of allometrically possible resources on the  $s$  axis by  $A$ . The distribution of the species number in set (ii) can be obtained from Eq. (30) by substituting  $C_0 \rightarrow 1 - C_0$  and is therefore given by  $P((1 - C_0)\kappa A, \rho; n)$  as defined in Eq. (A.4). The distribution of the number of species in (iii) can be obtained in the same way as the distribution of the total number of species (Section 3.1), just that the relevant range of  $s$  is now  $R - A$ , and not  $R$ . Hence this distribution is given by  $P(\kappa(R - A), \rho; n)$ . The distribution of the number of species in the union of these two sets is given by the convolution

$$P_{union}(n) = P((1 - C_0)\kappa A, \rho; n) * P(\kappa(R - A), \rho; n) = P(\kappa(R - C_0A), \rho; n). \tag{35}$$

The second equation is easily verified by comparing the corresponding cumulant generating functions (A.1).

The number of species in set (i) is given by  $P_{gen}(k) = P(C_0\kappa A, \rho; k)$  as defined above. Using the known distribution  $P(\kappa R, \rho; S)$  for  $S$ , the conditional distribution of generality can be obtained as

$$P_{gen}(k|S) = \frac{P_{gen}(k) P_{union}(S - k)}{P(\kappa R, \rho; S)} = \frac{\Gamma(C_0\kappa A + k)\Gamma(\kappa R)\Gamma(1 + S)\Gamma(\kappa(R - C_0A) - k + S)}{\Gamma(C_0\kappa A)\Gamma(1 + k)\Gamma(\kappa(R - C_0A))\Gamma(1 - k + S)\Gamma(\kappa R + S)}. \tag{36}$$

Remarkably, just as the conditional expectations, Eqs. (20) and (21), this result is independent of  $\rho$ . The parameter  $S$  is playing a similar role instead (see below). Eq. (36) is now evaluated for large  $S$ . Specifically, we assume (i)  $S \gg \kappa R$ , which is natural when  $S$  is of the order of its expectation value  $\kappa R\rho/(1 - \rho)$  and  $1 - \rho \ll 1$ , (ii)  $S \gg 1$ , (iii) we restrict ourselves to values of  $k \ll S$ , and (iv) in order to take the distinguished limit of fixed link density, we set  $C_0 = Z_0/S$  with fixed  $Z_0$ . Expanding the logarithm of Eq. (36) for large  $S$  (e.g., using Stirling’s formula) then gives

$$\ln P_{gen}(k|S) = \ln \frac{\kappa AZ_0}{kS} + \frac{-(\kappa R - 1)k + \kappa LZ_0[\gamma + \psi_0(\kappa R) + \psi_0(k) - \ln S]}{S} + \dots, \tag{37}$$

where  $\gamma \approx 0.57$  is the Euler constant and  $\psi_0(x) = (d/dx)\ln \Gamma(x)$  the digamma function.

A similar expansion can be obtained for a distribution of the form  $P(C_0\kappa A, \tilde{\rho}; k)$  given by Eq. (A.4), when the parameter  $\tilde{\rho}$  is assumed to behave such that  $S = b/(1 - \tilde{\rho})$  with fixed  $b$  for large  $S$ , which is natural in view of  $\langle S \rangle \sim 1/(1 - \rho)$ . One obtains

$$\ln P(C_0\kappa A, \tilde{\rho}; k) = \ln \frac{\kappa AZ_0}{kS} + \frac{-bk + \kappa LZ_0[\gamma + \ln b + \psi_0(k) - \ln S]}{S} + \dots. \tag{38}$$

A comparison of the two expansions shows that

$$P_{gen}(s, k|S) \approx \mathcal{N}P(C_0\kappa A(s), \tilde{\rho}; k), \tag{39}$$

where

$$\mathcal{N} = \mathcal{N}(s) = \exp\{C_0\kappa A(s)[\gamma_0(\kappa R) - \ln(\kappa R - 1)]\} \tag{40}$$

and

$$\tilde{\rho} = 1 - \frac{\kappa R - 1}{S}. \tag{41}$$

Hence, apart from the new parameters  $\mathcal{N}$  and  $\tilde{\rho}$ , the form of the generality distribution for fixed  $S$  is approximately the same as for fluctuating  $S$ .

The additional normalization factor  $\mathcal{N}$  enters because  $k$  can never exceed  $S$ , while  $P(C_0\kappa A, \tilde{\rho}; k)$  is non-zero for all  $k$ . When the expected number of consumers is much smaller than  $S$ , i.e., for small connectances  $C_0$ , the value of  $\mathcal{N}$  approaches 1. This can be seen by noting that  $\gamma_0(x) - \ln(x - 1) = 1/(2x) + \mathcal{O}(x^{-2})$ , so that we can write  $\mathcal{N} = \exp[\tilde{C}A/(2R)]$  with  $\tilde{C} \approx C_0$ . The dependence on  $S$  is fully contained in the new parameter  $\tilde{\rho}$ . Its relation to  $\rho$  can be understood by substituting  $S$  in Eq. (41) by  $\langle S \rangle = \kappa R\rho/(1 - \rho)$ , which leads to

$$1 - \tilde{\rho} = \frac{\kappa R - 1}{\kappa R\rho} (1 - \rho) \approx (1 - \rho). \tag{42}$$

Of course, the forgoing interpretation of Eq. (39) makes sense only when  $\kappa R > 1$ . Yet, Eq. (39) is numerically valid also when continued analytically to the region  $0 < \kappa R \leq 1$  where  $\tilde{\rho} \geq 1$ .

In Section 4.1 it was shown that the effect of a small, non-zero  $\beta$  can be approximated by a renormalization of the coefficients  $\kappa$  and  $\rho$ . Eq. (39) shows that for  $\beta = 0$  the effect of fixing  $S$  is also essentially a renormalization of  $\rho$ . Even though the generality distribution for fixed  $S$  and non-zero  $\beta$  is difficult to compute analytically, it is reasonable to assume that this too can be approximated by an expression of the form (39) with an appropriate pair of parameters  $\tilde{\kappa}$  and  $\tilde{\rho}$ .

In order to obtain the overall conditional generality distribution we go, again, over to the moment-generating function

$$M_{gen}(s, z|S) := \sum_{k=0}^{\infty} P_{gen}(k, z|S)z^k \approx \mathcal{N} \left( \frac{1 - \tilde{\rho}}{1 - \tilde{\rho}z} \right)^{C_0\tilde{\kappa}A(s)}. \tag{43}$$



The average of this expression over  $s$  for the simple case  $\lambda \rightarrow 0$  is

$$M_{gen}(z|S) \approx \frac{\tilde{u} - 1}{\log \tilde{u}} \quad \text{with} \quad \tilde{u} = \exp\left(\frac{\tilde{C}}{2}\right) \left(\frac{1 - \tilde{\rho}}{1 - \tilde{\rho}z}\right)^{C_0 \tilde{\kappa} R} \quad (44)$$

This result was verified by comparison with a direct numerical simulations of the model. Fig. 3 shows simulation results in comparison with the predictions of Eq. (44) and with the results of numerically averaging Eq. (36) directly over  $\Lambda = 0..R$ . Although the precision of the approximation, Eq. (44), decreases for increasing  $C_0$  and  $k$  in comparison with the prediction using Eq. (36), it is surprisingly good even for large values of  $C_0$  and  $k$ . For large  $C_0$  and small  $k$  the simulations deviate noticeably also from the prediction using Eq. (36), because in this parameter range the effects of intra-clade consumption, that had here been ignored, become relevant. Even for smaller  $R$ ,  $\kappa R = \mathcal{O}(1)$ , and  $\beta > 0$ , where Eq. (44) does not make quantitative predictions, the general form of this expression still seems to be valid. Fig. 4 shows some examples of numerical results in this regime compared with curves obtained by fitting  $\tilde{C}$ ,  $\tilde{\rho}$  and  $\tilde{\kappa}$  in Eq. (44). The fitted curves describe the distributions similarly well as the quantitative predictions above: deviations occur many for very small and very large  $k$ .

#### 4.4. The vulnerability distribution for fixed $S$

The distribution of the vulnerability  $m$  is most easily computed directly conditional to fixed  $S$ : Assume

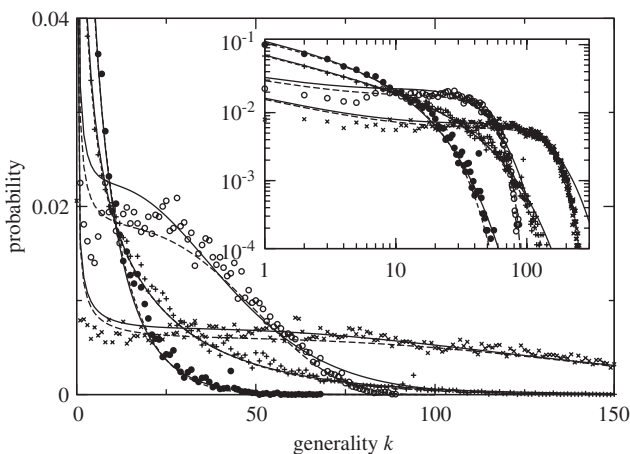


Fig. 3. Steady-state generality distributions conditional to fixed species number  $S$  obtained from simulations with  $C_0 = 0.1$ ,  $S = 100$  ( $\bullet$ ),  $C_0 = 0.1$ ,  $S = 300$  ( $+$ ),  $C_0 = 0.5$ ,  $S = 100$  ( $\circ$ ), and  $C_0 = 0.5$ ,  $S = 300$  ( $\times$ ) in comparison with the corresponding predictions by Eq. (44) (solid) and by directly averaging Eq. (36) over  $\Lambda = 0..R$  (dashed). The other parameters were  $R = \ln 10^{20}$ ,  $D = 0.005$ ,  $\rho = 0.95$ ,  $\kappa = 10/R$ ,  $\lambda = 10^{-3}$ ,  $\beta = 0$  (no fitting). For all examples  $\langle S \rangle = 190$ . The inset shows the same data on a double-logarithmic scale.

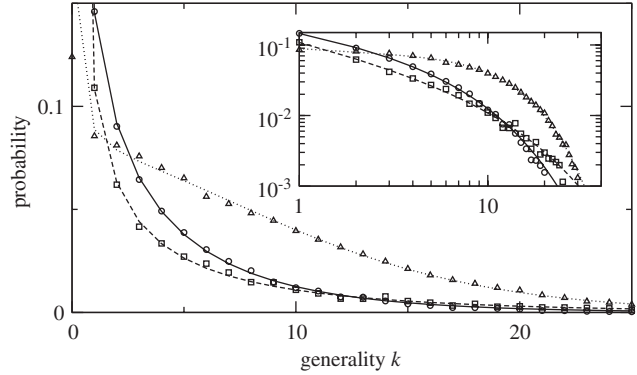


Fig. 4. Simulation results for the generality distributions conditional to  $S = 40$  with  $C_0 = 0.1$ ,  $\beta = 0$  (squares),  $C_0 = 0.1$ ,  $\beta = 0.05$  (circles),  $C_0 = 0.3$ ,  $\beta = 0.05$  (triangles) compared to distributions fitted by adjusting the parameters  $\tilde{C}$ ,  $\tilde{\rho}$  and  $\tilde{\kappa}$  in Eq. (44) (dashed, solid, dotted line). The other parameters were  $R = \ln 10^4$ ,  $D = 0.005$ ,  $\rho = 0.95$ ,  $\kappa = 2/R$ ,  $\lambda = 10^{-3}$ .

species to be indexed in the order of increasing  $s$  starting with 1. For  $\lambda \rightarrow 0$  the number of possible consumers of species  $i$  is then simply  $i$ . When assuming again that resources evolve much faster than their consumers, the consumers of  $i$  are determined by (i) the random connection of consumers with probability  $C_0$  when the resource-clade founder enters the food web and (ii) random re-connections with probability  $C_0$  during speciations of resources. Neither of these processes introduces correlations in the connectivities within the set of possible consumers of  $i$ . Thus, links are statistically independent and the vulnerability of  $i$  is given by a binomial distribution. Averaging over the food web yields

$$P_{vul}(m|S) = \frac{1}{S} \sum_{i=m}^S \binom{i}{m} C_0^m (1 - C_0)^{i-m} \quad (45)$$

This is exactly the expression that Camacho et al. (2002a) obtained in their analysis of the niche model. Following their observation that in the limit of large  $S$  with constant  $Z_0 = C_0 S$  and  $i = \mathcal{O}(S)$  the binomial distribution can be approximated as Poisson and the sum by an integral, one obtains

$$P_{vul}(m|S) = \frac{1}{Z_0} \int_0^{Z_0} \frac{t^m e^{-t}}{m!} dt \quad (46)$$

As is shown in Fig. 5, this result predicts the vulnerability distribution similarly well as Eq. (44) the generality distribution.

Note that the Poisson distribution entering Eq. (46) is the special case  $P(t/B, B; n)$ ,  $B \rightarrow 0$  of the general distribution  $P(A, B; n)$  entering Eq. (39). Thus, integral (46) is also a limiting case of the general form, Eq. (44). In the case of generality distributions, however,  $B$  is typically close to one.

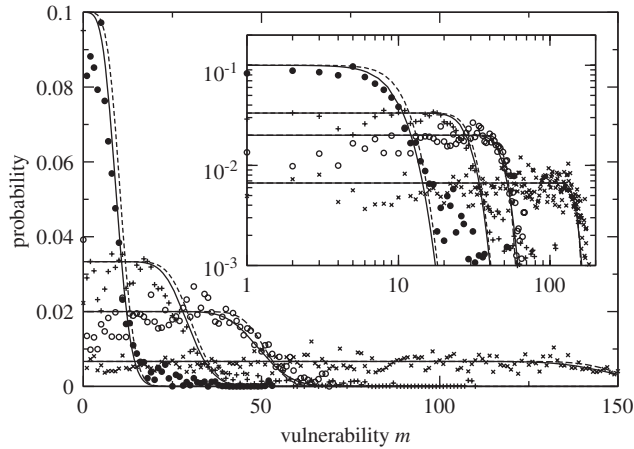


Fig. 5. Steady-state vulnerability distributions conditional to fixed species number  $S$ . Parameters are the same as in Fig. 3. The solid and dashed lines correspond to Eqs. (46) and (45), respectively. The inset shows the same data on a double-logarithmic scale.

## 5. Comparison with other topological food-web models

### 5.1. Comparison with the cascade model

The main idea upon which the cascade model is based, random connections restricted by a trophic hierarchy, is retained in the speciation model, albeit refined in several ways. The cascade model is recovered from the speciation model in the limit of no loops ( $\lambda = 0$ ), and no speciations,<sup>2</sup> i.e.  $r_+ \rightarrow 0$ . Then all species enter the species pool by adaptations and are independently, randomly connected to their resources and consumers, just as it was assumed for the consumers alone in the foregoing section. However, the limit  $r_+ \rightarrow 0$  does not describe empirical data particularly well (Rossberg et al., 2005). Typical parameter sets for the speciation model have  $r_+ \approx r_-$  (Table 1).

### 5.2. Comparison with the niche model

#### 5.2.1. Degree distributions

It was mentioned already that the distribution of vulnerability in the niche model is approximately the same as in the speciation model, in both cases given by Eq. (46). In the case of the niche model  $Z_0 = 2CS = 2Z$  where the targeted connectance  $C$  and the species number  $S$  are parameters of the model. In both cases the distribution is due to random connections with possible consumers.

For the generality distribution the situation is more complex. As the analysis of Camacho et al. (2002a) showed, it is for the niche model essentially determined

by the distribution of the “niche width”, i.e., the size of the interval containing the resources of a species on the niche-parameter scale. Williams and Martinez (2000) chose this width for each species as its niche value  $n$  times a random variable  $x$  with a beta distribution of the form

$$p_x(x) = b(1-x)^{(b-1)} \approx be^{-bx}, \quad (47)$$

where  $b = (1 - 2C)/(2C)$  depends on the targeted directed connectance  $C$ . The approximation by an exponential is valid for  $b \gg 1$ , i.e. for  $C \approx 1/(2b) \ll 1$ . Williams and Martinez (2000) used this particular form for its computational simplicity. No ecological arguments to motivate it seem to be known. Since species are independently and evenly distributed with density  $S$  in the one-dimensional “niche space”, the number of species in the niche interval follows a Poisson distribution with expectation value  $Snx$  when  $x$  is fixed. Averaging over all  $x$  yields the geometric distribution

$$\begin{aligned} P_{gen}^{(niche)}(n, k) &= \int_0^1 \frac{(Snx)^k}{k!} e^{-Snx} be^{-bx} dx \\ &= \frac{1}{1 + nZ_0} \left( \frac{nZ_0}{1 + nZ_0} \right)^k. \end{aligned} \quad (48)$$

The overall generality distribution is obtained by averaging Eq. (48) over  $n$ . The calculation is simplified by the approximation  $k \approx Snx$ , i.e.

$$P_{gen}^{(niche)}(n, k) \approx \frac{1}{Sn} p_x\left(\frac{k}{Sn}\right) = \frac{1}{nZ_0} \exp\left(-\frac{k}{nZ_0}\right), \quad (49)$$

which is valid for  $nZ_0 \gg 1$  [cf. Eq. (48)]. This leads to the result of Camacho et al. (2002a)

$$P_{gen}^{(niche)}(k) = \int_0^1 P_{gen}^{(niche)}(n, k) dn \approx \frac{1}{Z_0} E_1\left(-\frac{k}{Z_0}\right) \quad (50)$$

with  $E_1(x) := \int_x^\infty t^{-1} \exp(-t) dt$  denoting the exponential integral function. Camacho et al. (2002b) concluded that the distribution of the scaled generality  $k/(2Z)$  or, for single instances of food webs more appropriate, its cumulative distribution, should have the universal form

$$P\left(\frac{k}{2Z} \geq x\right) = \int_x^\infty E_1(x') dx' = \exp(-x) - xE_1(x), \quad (51)$$

and verified this impressively by a comparison with empirical data.

In order to see if this observed regularity is reproduced also by the speciation model, cumulative distribution functions for the speciation mode obtained from Eq. (44) were compared with Eq. (51). The value for  $k = 0$  was excluded from the comparison because (i) in many empirical foodwebs the lowest trophic level ( $k = 0$ ) is only poorly resolved and (ii) approximation (50) is undefined at  $k = 0$  and Eq. (44) is not accurate at this point either. The scaling factor  $Z_0^{-1}$  for the

<sup>2</sup>Observe that for  $r_+ \rightarrow 0$  the often encountered combination  $-\kappa \ln(1 - \rho)$  simplifies to  $r_+/r_-$ .

generality and the correction  $\tilde{\mathcal{N}}$  of the normalization constant were therefore determined directly by transforming the cumulative speciation-model distributions to  $\tilde{\mathcal{N}} \sum_{k'=k}^{\infty} P_{gen}(k'/Z_0|S)$  such as to minimize the mean-least-square deviation from Eq. (51) for  $k \geq 1$ . These curves match Eq. (51) surprisingly well over a wide parameter range (Fig. 6a). The empirical data are described well by both distributions (Fig. 6b).

To understand the reason for this apparent scaling law of speciation-model food webs, consider the speciation-model generality distribution (44) conditional to  $k \geq 1$  in the limit of low connectance  $C_0, \tilde{C} \rightarrow 0$  (now at fixed  $S$ ), i.e. the distribution with the moment generating function

$$\lim_{C_0, \tilde{C} \rightarrow 0} \frac{M_{gen}(z|S) - M_{gen}(0|S)}{1 - M_{gen}(0|S)} = \frac{\ln(1 - \tilde{\rho}z)}{\ln(1 - \tilde{\rho})}. \quad (52)$$

This is easily seen to be the distribution of resource-clade sizes [cf. Eq. (13)]

$$\frac{\tilde{\rho}^k}{k \ln(1 - \tilde{\rho})}. \quad (53)$$

In this limit of low connectance most species belong to the lowest trophic level, only a few heterotrophs remain, and the percolation of the network is lost. Therefore, this limit does not correspond to the general situation encountered in the field. But the approximate form of the log-series distribution (53) is retained also for more complex networks. For values of  $\tilde{\rho} \approx 0.8$ , this distribution has a shape quite similar to the exponential integral distribution, Eq. (49). When going over to cumulative distributions, the fit looks even better. Thus, the observed generality distributions can be interpreted mechanistically in terms of the steady-state distributions of evolutionary clade sizes, corrected for fixing  $S$  and trophic link breaking. This also suggests that the “scaling” distribution (53)  $\sim k^{-1} \exp(-k \ln \tilde{\rho})$  or the more accurate result (44) would rather be the adequate functional forms than the exponential integral function (49).

In spite of the similarities of the overall generality and vulnerability distributions, there are marked differences in the detailed predictions of the two models. Consider, for example, the generality distribution for species near the lower end of the trophic cascade, i.e., species with  $\kappa A(s) \ll 1$  in the speciation model and  $n \ll 1$  in the niche model, that have at least one resource species ( $k \geq 1$ ). For the speciation model Eqs. (39) and (A.4) lead again to the clade-size distribution (53), while for the niche model, Eq. (48), predicts

$$(1 - nZ_0)(nZ_0)^{k-1}. \quad (54)$$

Thus, for the niche model it is very probable that such a species has exactly one resource, whereas for the speciation model larger generalities can also be ex-

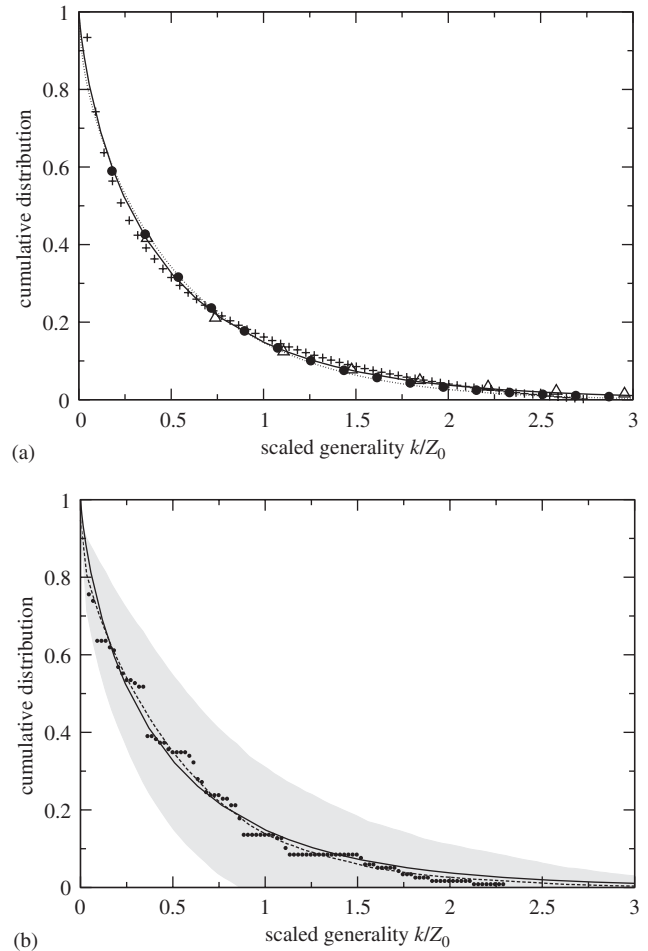


Fig. 6. Comparison of niche-model and speciation-model predictions for the cumulative generality distribution. (a) Approximation (44) for the speciation model with  $C_0 \tilde{\kappa} R = 0.2, \tilde{\rho} = 0.75$  (triangles),  $C_0 \tilde{\kappa} R = 1.5, \tilde{\rho} = 0.75$  (circles),  $C_0 \tilde{\kappa} R = 0.2, \tilde{\rho} = 0.98$  (plus),  $C_0 \tilde{\kappa} R = 1.5, \tilde{\rho} = 0.98$  (dotted line) in comparison with approximation (51) for the niche model (solid line). (b) The empirical distribution for Little Rock Lake Martinez (1991) (dots), the speciation model prediction from numerical simulations (dashed, shaded area is the 1- $\sigma$  range of fluctuations before scaling), and again approximation (51) for the niche model (solid line). All distributions have the point  $k = 0$  removed and are scaled and normalized to minimize mean-square deviations from Eq. (51).

pected. An empirical test should be capable of distinguishing these two predictions.

### 5.2.2. Intervality

A major distinction of the niche model from the cascade model is the intervality it enforces upon the diets of consumers. While the degree of intervality obtained with the cascade model is typically too small compared with empirical data (Cohen et al., 1990), it is too large for the niche model (Cattin et al., 2004). Under certain conditions the speciation model can also produce a high degree of intervality. Consider some arbitrary ordering of clades, for example by the speed parameter of the

founder species, and an ordering of the species within each clade given by a traversal of the evolutionary tree.<sup>3</sup>For this ordering (which differs from an ordering by  $s$ ) diets will form contiguous sets when (i) the average number of resources clades (33) is low, i.e., when most consumers have either one or no resource clade and (ii) the probability that resources break out of resource clade during the clade's lifetime is low. Then the set of a consumer's resources is usually simply the non-extinct part of an evolutionary subtree. The probability of resource break-out is small when  $\beta \times$  (resource clade size)  $\times$  (clade lifetime in generations) is small, which, by arguments analogous to those used in Section 3.2, is the case when

$$\beta\rho^*/(1 - \rho^*) \ll 1. \quad (55)$$

For typical model parameters we find that these two conditions are satisfied to some extent but not too well (Table 1), in accordance with expectations. Correspondingly, the degree of intervality  $D_{diet}$  (Cattin et al., 2004) of empirical data is reproduced well by the model (Rossberg et al., 2005).

We conclude with Cattin et al. (2004) that the larger-than-random intervality observed in food webs may not so much result from a low dimensionality of the niche space, as has been proposed (Cohen, 1978), but rather reflects the importance of the phylogenetic history for the food-web structure.

### 5.3. Comparison with the nested hierarchy model

Just as for the niche model, the generality distribution for the nested hierarchy model is imposed “by hand” by specifying distribution (47) and setting  $k \approx Snx$ . But the structure of the set of resources is determined by a more complex algorithm that has been designed in such a way that consumers and resources form groups ( $\approx$  clades), and consumers and resources from the same groups share resources and consumers, respectively. The algorithm is intended to mimic a structure that would result from a phylogenetic evolution of the web, without explicitly modeling this evolution. The speciation model achieves a similar effect by explicitly modeling the evolutionary dynamics.

<sup>3</sup>For example, the order given by the recursive algorithm  $list(A)$  defined as

1. if  $A$  has not become extinct  
  print  $A$ ;
2. for all direct descendants  $B$  of  $A$  in order of appearance  
   $list(B)$ ;

starting with  $list$ (clade founder).

## 6. Variants of the speciation model

Modeling complex ecological systems often requires difficult decisions with regards to which kinds of effects ought to be incorporated into a model and which can be ignored. Here, two variants of the speciation model are shortly discussed that include aspects of the real system that had been left out in the original model. For both variants, the analytic results derived in the previous sections remain valid without change.

### 6.1. A variant with asymmetric link persistence

In the analysis above it was assumed that consumer–resource links are statistically independent of the phylogenetic history of the consumers. If this assumption is valid, one may as well modify the model such as to choose all resources of a descendant species at random after its speciation, without affecting the analytic results obtained above. More generally, one might incorporate an asymmetry in the persistence (or reconnection probability) of links between consumers and resources in the following way:

In the original form of the model, the connectivity of the descendant species was (randomly) re-assigned for a fraction  $\beta$  of all possible trophic links. In the asymmetric variant of the model, the connectivity from the descendant species to its consumers is re-assigned for a fraction  $\beta_c$  of all possible consumers, and the connectivity to resources is re-assigned for a fraction  $\beta_r$  of all possible resources, with  $\beta_c \neq \beta_r$  in general.

In fact, there is no ecological reason to expect  $\beta_c = \beta_r$ . A large difference between the values of  $\beta_c$  and  $\beta_r$  such as considered above ( $\beta_c = \beta \ll \beta_r = 1$ ) could be understood from the assumption that in the competition between related species their sets of resources are much more important than their sets of consumers: In order to avoid competitive exclusion, related species need drastically different sets of resources ( $\beta_c = 1$ ), while there is only little evolutionary pressure for a descendant species to have a different set of consumers than its predecessor ( $\beta_c \ll 1$ ).

However, one might also argue that by the direct resource–consumer interaction alone. Then one could expect it to be advantageous for a descendant species to evade its predecessors consumers (large  $\beta_c$ ), while maintaining its resources (small  $\beta_r$ ). This would lead to the reverse relation between  $\beta_c$  and  $\beta_r$ . An empirical test to establish which of these two mechanisms is more relevant might be possible.

### 6.2. A variant with quantitative link strength

Topological food-web models are often criticized for ignoring the fact that the link strength in food webs, instead of being either 1 or 0, is in reality a continuous



quantity (Berlow et al., 2004). There is a simple way to incorporate continuously varying link strengths in the speciation model without affecting its statistical properties.

Instead of assigning to each possible trophic link a connectivity of either 0 and 1, quantify the strength of each possible link by a real number between 0 and 1. Where the connectivity was copied during speciations in the original model, the links strength is copied now. Where the connectivity was set to 1 with probability  $C_0$  and to 0 otherwise, set the link strength to an appropriately distributed random number between 0 and 1 now. For a characterization of the resulting food webs in terms of topological food-web statistics, count each link with strength larger than some threshold as present, and all other links as absent. That is, the thresholding of the link strength is just delayed to the time of the characterization. While this modification is straightforward for the speciation model, modifications of other topological models to postpone the thresholding of link strength might be possible, if at all, only at the price of increasing the model complexity.

Of course, an evolution where the link strength either does not change at all or is reset to a completely new random value is quite artificial. More natural it would be to vary the link strength by a small random amount at each evolutionary step. In such a model, link breaking and reconnecting events relative to some threshold ( $1 - C_0$ ) would be correlated. They would be concentrated at certain pairs of consumer and resources clades with link strength near the threshold. Further studies are required to understand what effect this would have on the overall network structure.

## 7. Discussion and outlook

Besides improving the general understanding of the properties of the speciation model and their dependence on model parameters, a purpose of this work was also to show that the speciation model integrates the underlying ideas from previous, simpler models (see Section 5). The speciation model retains the trophic ordering of the cascade model. In fact, it contains the cascade model as a special case. By the interplay of speciations, extinctions, and adaptations of new species to the habitat, the speciation model reproduces three key features of the niche model and the nested hierarchy model at the same time: (1) the empirical distributions of generality, which in the niche model and similarly in the nested hierarchy model are obtained only by a special, ecologically unmotivated choice of the niche-width distributions; (2) intervality, to the degree that is actually observed (Cattin et al., 2004; Rossberg et al., 2005); (3) the organization of resources into groups of related species

that share consumers and *vice versa*. This unifying character of the speciation model is probably the main reason for its high accuracy in reproducing empirical data (Rossberg et al., 2005).

The observed broad, log-series-like generality distributions have been traced back to, among others, a condition  $1 - \rho \ll 1$ . This means that the rate constant for speciations  $r_+$  is numerically close to the rate constant for extinctions  $r_-$ . For any phylogenetically closed system, a steady state always requires that extinction rates and speciation rates are equal, independent of the statistical details of the branching pattern. For the half-open system considered here,  $1 - \rho \ll 1$  implies that the contributions from foreign adaptations to the species pool are small compared to the contribution from speciations. In fact,  $1 - \rho$  directly equals the fraction of species in the food web that have entered by foreign adaptations. However, in order to obtain broad, left-skewed generality distributions, the independence of the speciation and extinction probability of a species from the actual size of its clade is also important. If, instead, large clades would notably favor extinctions and small clades speciations, clade size distributions would be dominated by a “typical” clade size, which would, in the model, also lead to a narrower generality distribution. In an analysis of paleontological time series Raup (1991) applied a model for the size of genera identical to the model used here for the dynamics of clade sizes (Eqs. (1)–(3)). While, on the average, this model (with  $\rho = 0.996$ ) reproduced the data well, the scatter in the paleontological data was larger than in the model. Raup could explain this observation by assuming that the overall evolution rate varies over time. Since such a variation can also be described by a (random) nonlinear transformation of the time axis, it does not affect statistics that refer only to a particular moment in time, such as food-web structures. Thus our assumption of a simple birth/death process is supported by paleontological observations.

As a direct consequence of this birth/death process, a characterization of food webs in terms of “clades” has been derived. Table 1 lists expectation values for characteristic quantities corresponding to some empirical food webs. It might be interesting to compare these results with the taxonomic structure of the actual empirical webs or the model dynamics with paleontological records.

In Section 4 it was shown that a correlation between the evolution rates and the trophic height leads to the observed asymmetry between generality and vulnerability distributions. However, in the present model this requires evolution rates spanning an unrealistically large range of about 20 orders of magnitude. We are currently evaluating a variant of the speciation model that achieves a similar effect without any differences in evolution rates by making not directly the trophic links



hereditary but the properties of species determining link strengths. An asymmetry of the heredity between species-as-consumers and species-as-resources leads effectively to an asymmetry of the link persistence as described in Section 6.1 above. Numerical results with the new model are promising, but analytically we understand it only in so far as it can be approximated by the speciation model, so that the analysis presented here remains valid. Details regarding the new model will be reported elsewhere.

Our findings indicate that a food web's population dynamical stability and persistence are not as important determinants of its structure as is sometimes assumed. From a technical point of view, this is good news. It appears possible to obtain natural food-web structures without time-consuming population dynamical simulations. These food webs could then be investigated also with respect to the question how their structure affects population dynamical stability.

In the course of this work, analytic approximations for several empirically testable predictions of the speciation model could be obtained. These include the average clade size  $\langle n \rangle$ , the number of clades  $\langle c \rangle$  in the web, the age of clades in generations (speciation times)  $-\ln(1 - \rho)$ , the average number of resource clades (Eq. (33)), and the generality distribution of consumers at low trophic levels (Eq. (53)). A careful comparison of the models discussed here and other food-web models with existing empirical data and new results from ongoing efforts in the field will reveal discrepancies and, hopefully, suggest new ideas to bringing us another step closer to understanding this fascinating aspect of life on earth.

## Acknowledgements

The authors acknowledge generous support by The 21st Century COE Program "Bio-Eco Environmental Risk Management" of the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## Appendix A. A family of distribution functions encountered in the analysis of the speciation model

The analysis of the steady state of a simple model of evolutionary dynamics (Section 3.1) naturally leads to probability distributions  $p_n$  for species number  $n$  with a cumulant generating function

$$\ln \sum_n p_n z^n = K_{A,B}(z) = A \ln \left( \frac{1-B}{1-Bz} \right), \quad (\text{A.1})$$

where  $0 < A$ ,  $0 < B < 1$ .

From this, the mean

$$\langle n \rangle = \left. \frac{dK_{A,B}(e^u)}{du} \right|_{u=0} = \frac{AB}{1-B}, \quad (\text{A.2})$$

and variance

$$\text{var } n = \left. \frac{d^2 K_{A,B}(e^u)}{du^2} \right|_{u=0} = \frac{AB}{(1-B)^2} \quad (\text{A.3})$$

can be calculated directly. The ratio  $(\text{var } n)/\langle n \rangle$  is  $(1-B)^{-1}$  times larger than for Poisson distributions.

The distribution function itself is given by

$$p_n = P(A, B; n) := \frac{(1-B)^A B^n \Gamma(A+n)}{\Gamma(A) \Gamma(1+n)}. \quad (\text{A.4})$$

This implies that the ratio of consecutive probabilities is

$$\frac{p_{n+1}}{p_n} = \frac{B(A+n)}{1+n}. \quad (\text{A.5})$$

In particular, the most probable value is  $n = 0$  whenever  $AB < 1$ . Since  $B < 1$ , this is always the case when  $A \leq 1$ . For  $A = 1$  one gets exactly a geometric distribution

$$p_n = (1-B) B^n, \quad (\text{A.6})$$

and for small  $A$  Eq. (A.4) simplifies to the log-series distribution

$$p_n = \begin{cases} 1 + A \ln(1-B) + \mathcal{O}(A^2) & \text{for } n = 0, \\ \frac{AB^n}{n} + \mathcal{O}(A^2) & \text{otherwise.} \end{cases} \quad (\text{A.7})$$

For small  $B$  a Poisson distribution is obtained: With fixed  $AB$ ,

$$p_n = \frac{(AB)^n}{n!} e^{-AB} + \mathcal{O}(B) \quad (\text{A.8})$$

uniformly in  $n$ . Finally, when  $AB \gg 1$  the distribution  $p_n$  can be approximated by a Gaussian with mean and variance given by Eqs. (A.2) and (A.3).

## References

- Amaral, L.A.N., Meyer, M., 1999. Environmental changes, coextinction, and patterns in the fossil record. *Phys. Rev. Lett.* 82 (3), 652–655.
- Bailey, N.T.J., 1964. *The Elements of Stochastic Processes*. Wiley, New York.
- Baird, D., Ulanowicz, R.E., 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecol. Monogr.* 59, 329–364.
- Berlow, E.L., Neutel, A.-M., Cohen, J.E., de Ruiter, P.C., Ebenman, B., Emmerson, M., Fox, J.W., Jansen, V.A.A., Jones, J.I., Kokkoris, G.D., Logofet, D.O., McKane, A.J., Montoya, J.M., Petchey, O., 2004. Interaction strengths in food webs: issues and opportunities. *J. Anim. Ecol.* 73, 585–598.
- Briand, F., Cohen, J.E., 1987. Environmental correlates of food chain length. *Science* 238, 958–960.
- Caldarelli, G., Higgs, P.G., McKane, A.J., 1998. Modelling coevolution in multispecies communities. *J. Theor. Biol.* 193, 345.
- Camacho, J., Solé, R.V., 2000. Extinctions and taxonomy in a trophic model of coevolution. *Phys. Rev. E* 62 (1), 1119.

- Camacho, J., Guimerà, R., Amaral, L.A.N., 2002a. Analytical solution of a model for complex food webs. *Phys. Rev. E* 65, 030901R.
- Camacho, J., Guimerà, R., Amaral, L.A.N., 2002b. Robust patterns in food web structure. *Phys. Rev. Lett.* 88 (22), 228102.
- Cattin, M.-F., Bersier, L.-F., Banasek-Richter, C., Baltensperger, R., Gabriel, J.-P., 2004. Phylogenetic constraints and adaptation explain food-web structure. *Nature* 427, 835–839.
- Cohen, J.E., 1978. *Food Webs and Niche Space*. Princeton University Press, Princeton, NJ.
- Cohen, J.E., Briand, F., Newman, C.M., 1990. *Community Food Webs: Data and Theory*, Biomathematics, vol. 20. Springer, Berlin.
- Cohen, J.E., Pimm, S.L., Yodzis, P., Saldana, J., 1993. Body sizes of animal predators and animal prey in food webs. *J. Anim. Ecol.* 62, 67–78.
- Drossel, B., 1998. Extinction events and species lifetimes in a simple ecological model. *Phys. Rev. Lett.* 81, 5011–5014.
- Drossel, B., Higgs, P.G., McKane, A.J., 2001. The influence of predator–prey population dynamics on the long-term evolution of food web structure. *J. Theor. Biol.* 208, 91–107.
- Goldwasser, L., Roughgarden, J., 1993. Construction of a large Caribbean food web. *Ecology* 74, 1216–1233.
- Hall, S.J., Raffaelli, D., 1991. Food-web patterns: lessons from a species-rich web. *J. Anim. Ecol.* 60, 823–842.
- Havens, K., 1992. Scale and structure in natural food webs. *Science* 257, 1107–1109.
- Martinez, N.D., 1991. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.* 61, 367–392.
- Memmott, J., Martinez, N.D., Cohen, J.E., 2000. Predators, parasitoids and pathogens: species richness, trophic generality and body size in a natural food web. *J. Anim. Ecol.* 69, 1–15.
- Montoya, J.M., Solé, R.V., 2003. Topological properties of food webs: from real data to community assembly models. *Oikos* 102 (3), 614–622.
- Pastor-Satorras, R., Smith, E., Solé, R.V., 2003. Evolving protein interaction networks through gene duplication. *J. Theor. Biol.* 222, 199–210.
- Pimm, S.L., 1984. The complexity and stability of ecosystems. *Nature (London)* 307 (321), 326.
- Polis, G.A., 1991. Complex desert food webs: an empirical critique of food web theory. *Am. Nat.* 138, 123–155.
- Raup, D.M., 1991. A kill curve for Phanerozoic marine species. *Paleobiology* 17 (1), 37–48.
- Rossberg, A.G., Matsuda, H., Amemiya, T., Itoh, K., 2005. An explanatory model for food-web structure and evolution. *Ecol. Complexity*, in press. URL <http://arXiv.org/q-bio.PE/0410030>
- Stouffer, D.B., Camacho, J., Guimerà, R., Ng, C.A., Amaral, L.A.N., 2005. Quantitative patterns in the structure of model and empirical food webs. *Ecology* 86 (5), 1301–1311.
- Sugihara, G., 1983. *Niche hierarchy: structure, organization, and assembly in natural communities*. Ph.D. Thesis, Princeton University.
- Tokita, K., Yasutomi, A., 2003. Emergence of a complex and stable network in a model ecosystem with extinction and mutation. *Theor. Popul. Biol.* 63, 131–146.
- Vazquez, A., Flammini, A., Maritan, A., Vespignani, A., 2003. Modeling of protein interaction networks. *ComplexUs* 1, 38.
- Warren, P.H., 1989. Spatial and temporal variation in the structure of a freshwater food web. *Oikos* 55, 299–311.
- Warren, R.J., Lawton, J.H., 1987. Invertebrate predator-prey body size relationships: an explanation of upper triangularity in food webs and patterns in food web structure. *Oecologia* 74, 231–235.
- Williams, R.J., Martinez, N.D., 2000. Simple rules yield complex food webs. *Nature* 404, 180–183.
- Yoshida, K., 2002. Long survival of “living fossils” with low taxonomic diversities in an evolving food web. *Paleobiology* 28 (4), 464–473.
- Yoshida, K., 2003. Evolutionary dynamics of species diversity in an interaction web system. *Ecol. Model.* 163, 131–143.