J. theor. Biol. (2001) **212,** 355–366
doi:10.1006/jtbi.2001.2382, available online at http://www.idealibrary.com on IDEAL®

Species Orthogonalization

PETR KRÁL*

Department of Chemical Physics, Weizmann Institute of Science, 76100 *Rehovot, Israel*

(*Received on* 11 *September* 2000, *Accepted in revised form on* 21 *June* 2001)

We discuss the general formation of complementary behaviors, functions and forms in biological species competing for resources. We call *orthogonalization* the related processes on macro and micro-level of a self-organized formation of correlations in the species properties. Orthogonalization processes could be, for example, easily observed in sympatric speciation, as we show in numerical studies carried out with a new population equation. As a practical result, we find that the number of species is proportional to the effective richness of resources and depends on their history.

© 2001 Academic Press

Introduction

Individuals in any biological species differ a little in their behaviors, forms and other parameters, which can help to relax their otherwise large competition. If the differences become too large, individuals can decrease their interest for interbreeding, and species can consequently split up. During speciation and species coexistence, correlations are formed between parameters of the species, which adjust distributions for different species in such a way that resources are effectively shared. In this process, complex patterns of complementary sets of parameters are formed, where species can be seen to approach discrete *orthogonal states*. These "orthogonalization processes" (OP) can be modeled in multidimensional mathematical spaces, with coordinates given by the species parameters.

It appears natural to define species by a mutual reproducibility of their individuals (Mayr, 1942, 1963). Since OP are strong in clusters of evolutionary and functionally close species, they could give a clear substance to the species definitions. Globally, we can define *species orthogonalization* as a process in which species approach a state, where they share resources and habitats in the most effective way, fully stabilized in time. Since species not only passively approach this ideal state, but they prefer to overturn the "status quo", by permanently improving the quality of their gene equipment, full orthogonalization cannot be realized. Instead, evolution gives a relatively stable coexistence (quasi-orthogonality) of biological species, time from time disturbed by speciation events. Close to these events OP are the most intense, but they are strong until the new species "settle down". Species interactions, which enable this process, last even after the species get in the resulting orthogonal states.

The need for orthogonalization might have caused the disappearance of many "intermediate" species, since these would hamper orthogonalization in the short- and long-term run, the last because *both* the resources volume and speciation rate are relatively stable. A more specific trace

* Present address: ITAMP, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, U.S.A.

0022–5193/01/190355 + 12 $35.00/0

© 2001 Academic Press

of OP are *character shifts* observed in species sharing habitats or in sympatric speciation (Brown & Wilson, 1956; Smith, 1966; May & Mac Arthur, 1972; Slatkin, 1980; Schluter & Mac Phail, 1992; Doebeli, 1996a; Drossel & Mac Kane, 1999), which reveal an efficient division of resources from the point of structure, day time or form of consumption. Competition for (orthogonal) discrete niches (Hutchinson, 1968) is confirmed, for example, by simple speciation/extinction models (Valentine & Walker, 1988). Therefore, OP might even control the numbers of species living in given areas (Rosenzweig, 1995; Plotkin *et al*., 2000).

To understand OP, we discuss first some evolution aspects of species on the micro and macro-level, and flow of information between the levels. Then, we focus on modeling of a sympatric speciation of species (Doebeli, 1996b; Bagnoli & Bezzi, 1997; Kondrashov & Kondrashov, 1999; Dieckman & Doebeli, 1999; Drossel & Mac Kane, 2000), and find a new speciation equation. We present it in numerical examples, where OP can be observed.

The Concept of Species Orthogonalization

We can look on species orthogonalization from different levels of the OP activity. In a self-organized molecular system (Eigen, 1971), different types of molecules play different roles, expressed by their microscopic structures. Evolution processes leading to the separation of these roles can be called *orthogonalization on the micro-level* (OMIL). Since correlations are also built between clusters of cooperating but different types of molecules, natural orthogonalizing units are *molecular quasispecies* (Eigen & Schuster, 1977; Eigen *et al*., 1989). Therefore, eigensolutions of the Eigen's equations implicitly incorporate quasispecies orthogonalization on the micro-level.

The molecular system can be a part of a macro-system, competing or cooperating with other macro-systems of the same or different types. Different types of macro-systems (individuals in different species or different organs inside individuals) analogously develop complementary functions and signaling activities in different times or space regions. Processes leading to such a differentiation of biological species can be called *orthogonalization on the macro-level* (OMAL). We could again find that OMAL is rather realized between self-organized clusters of species, analogously to the quasispecies on the micro-level, which can help to improve species definitions. Forces leading to OMAL have predominantly macroscopic origin, but their influence must be "inheritably" fixed on a micro-level, which can be realized there as a part of "externally driven" OMIL.

REALIZATION OF ORTHOGONALIZATION PROCESSES

Realization of OP on all structural levels of biological species is rather fascinating. Starting from the micro-level, proteins on the tertiary level, for example, resemble macro-tools, where prototypical sections can be identified with letters in a hypothetical alphabet (Lesk, 1951; Creighton, 1992; Holm & Sander, 1998), showing the activity of OMIL, i.e. a functional orthogonalization of certain protein sections. Recently, for example, orthogonalization as a measure of similarity between different molecules was used in a modeling of prebiotic species formation (Segré *et al*., 2000). Proteins also reflect the environment temperature, pressure or acidity by the compactness of their folding (Lumsden *et al*., 1997), which shows that some degrees of freedom can be tuned in proteins from the macro-level, without qualitatively affecting their *functions*.

Comparative studies show (Creighton, 1992) that proteins, common among different species, have nearly zero mutations at some regions, reflecting functional importance of these sections. Other sections are nearly *neutral* with respect to protein's functions, since compositions of amino acids present here are reasonably varied between evolutionary distant species (Wang, 1996). Such a structural "self-averaging" often applies to elongated sections stretching out of the proteins. The fact that fluctuations in these sections between individuals in a single species are still *statistically small* could reveal their use for ongoing OMAL processes. These processes can be engaged in a structural tuning of the neutral sections, with the goal to influence speed (catalytic strength), timing and other aspects of protein's functions, and thus vary the species *morphology*

and general activity. Some genes, present in the gene-pool in many similar copies (Creighton, 1992), are also tuned by OMAL to perform the same functions in different evolution stages of the organism, which carries them. Therefore, proteins and other microscopic units probably first internally develop by OMIL, and later are being externally tuned by OMAL, according to the needs on the macro-level.

OMAL is realized through an information (vertical) flow between the species macro and micro-levels, where each level expresses its needs and possibilities in a specific way. Stimulations between the levels are rich: the macro-world provides species individuals with various resources, acts on them through a set of *macroscopic variables* Φ (like radiation, temperature, humidity, etc.), and checks their ability to fit a given time-dependent environment. The individuals react according to their inherited and gained informations (Irwin & Price, 1999), coded in their gene-pools and memories, evolving as a response to the living conditions.

Description of Orthogonalization Processes

Mathematically, an individual can be represented by a point in a vector space, where coordinates ϕ_i are related to codons, genes, or other microscopic units (Baake *et al*., 1997). Giving the fluctuations in their composition, a species with many individuals occupies a finite volume $\Delta\phi$ of this space. An extreme example are viral quasi-species (Eigen, 1993), where $\Delta\phi$ is very large and rapidly grows in time by mutations. The species individuals can also be described in a vector space with coordinates given by their macroscopic characteristics $\mathbf{x}$ (Prügel-Bennett, 1997). Some sets of microscopic parameters ϕ (genotypes) can lead to morphological solutions with close macroscopic parameters $\mathbf{x}$ (phenotypes). Ambiguity of this reflection can be traced by a genotype–phenotype mapping (Fontana & Schuster, 1998).

The two types of vector spaces are formed by a direct product of their subspaces $\mathscr{H} = \mathscr{H}_1 \times \mathscr{H}_2 \times \cdots \times \mathscr{H}_n$ for the involved microscopic or macroscopic parameters. In dependence of the studied level, an individual can be thus identified with the vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ or $\phi = (\phi_1, \phi_2, \ldots, \phi_n)$. A species with many individuals can be described by a (character) distribution $f(\mathbf{x})$ or $\tilde{f}(\phi)$, which reflect various correlations in the populations. The distributions can be smoothed and normalized $f_0(\mathbf{x}) = f(\mathbf{x})/N$, where N is the number of involved individuals.

EVALUATION OF ORTHOGONALIZATION PROCESSES

To quantitatively appreciate OP, we can formally introduce a measure of species orthogonality. We say that two species are *orthogonal with a weight* μ, if a scalar product (Prugovečki, 1981) of their distributions $\langle f_0 | g_0 \rangle$ fulfills

$$\langle f_0 | g_0 \rangle \equiv \int \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \times \sqrt{f_0(x_1, x_2, \ldots, x_n)\, g_0(x_1, x_2, \ldots, x_n)} = 1 - \mu. \tag{1}$$

Any species is orthogonal to itself with a weight $\mu = 0$. Similarly, $\mu = 1$ for any two species, if their distributions for at least one component of the vector $\mathbf{x}$ totally avoid each other. Definition (1) just touches the aspect of a "visual dissimilarity" of the character distributions in species. The value $\mu \approx 1$, for example, does not assure that the two involved species are related in any way, and if they are, it does not prove that OP are *completed*. The last should be independently checked from the time-independence of the distributions f, g, as outlined in the Introduction.

To get more refined tools for capturing OP, we should rather focus on some "dynamical aspects" of OP, like the fact that species living in close contacts often largely *interact*, even if they are practically orthogonal $\langle f | g \rangle \approx 0$. Interactions, which enable orthogonalization, are realized through competition, described here by broadened distributions f_C, g_C [see eqn (8)]. For interacting species, these functions thus have a non-zero overlap $\langle f_C | g_C \rangle \neq 0$, even if $\langle f | g \rangle \approx 0$. When the interactions proceed vertically on the food-web (Amaral & Meyer, 1999), OP can be traced analogously.

It is also important to know in which way OP form *correlations* between the parameters $x_1, x_2, \ldots, x_n$ of interacting species. This can be

tested from the species distributions $F(x_k, x_l, \dots)$, $G(x_k, x_l, \dots)$, resulting by projection of $f(\mathbf{x})$, $g(\mathbf{x})$ on the selected $x_k, x_l, \dots$ parameters

$$F, G(x_k, x_l, \dots) = \int \mathrm{d}x_1 \, \mathrm{d}x_2 \dots \mathrm{d}x_{k-1} \, \mathrm{d}x_{k+1} \dots \times \mathrm{d}x_{l-1} \, \mathrm{d}x_{l+1} \dots \mathrm{d}x_n \, f, g(\mathbf{x}). \quad (2)$$

Similarly as f_C, g_C, the projected distributions F, G could also overlap

$$\langle F|G\rangle \equiv \int \mathrm{d}x_k \, \mathrm{d}x_l \dots \sqrt{F(x_k, x_l, \dots) G(x_k, x_l, \dots)} \neq 0, \quad (3)$$

even if the species are quasi-orthogonal $\langle f|g\rangle \approx 0$. For interacting species, this can be largely due to correlations formed by OP between the individual parameters $x_{1,2,\dots,n}$, in contrast to the overlap of f_C, g_C due to broadening (interaction).

These correlations can be traced, if we project f, g on a *functionally relevant cluster* of parameters. An example of two interacting bird species might be the triplet: x_b—beak length, x_l—leg length and x_w—wing size. This triplet could provide a reasonably small overlap $\langle F|G\rangle \approx 0$, while projections on subsets of (x_b, x_l, x_w) might already result in large overlaps $\langle F|G\rangle \neq 0$. This can be especially relevant for the projected and competition broadened distributions F_C, G_C. We can thus define an *effective orthogonalization variable* x, which would give the least overlap when we project on it. This variable then speaks about the character of correlations built by OP. The simplest possibility of defining it (see Fig. 1) is to project along the vector $\mathbf{x} = (\mathbf{C}_F - \mathbf{C}_G)\, x$, oriented in the direction between the centers of gravity $\mathbf{C}_F = (x_b^F, x_l^F, x_w^F)$ and $\mathbf{C}_G = (x_b^G, x_l^G, x_w^G)$ of the projected species distributions $F(x_b, x_l, x_w)$, $G(x_b, x_l, x_w)$, and integrate these distributions out in the remaining two orthogonal directions. We will apply the effective one-dimensional variable x in our numerical studies.

In Fig. 1, we schematically present the distributions $f(x_1, x_2)$, $g(x_1, x_2)$ for two species. They are practically orthogonal $\langle f(x_1, x_2)|g(x_1, x_2)\rangle \approx 0$, while the projected distributions $F(x_1)$, $G(x_1)$ or $F(x_2)$, $G(x_2)$ are not $\langle F(x_{1,2})|G(x_{1,2})\rangle \neq 0$. When we project $f(x_1, x_2)$, $g(x_1, x_2)$ on the effective orthogonalization variable x, defined by the direction between the distribution "centers", the resulting projected distributions $F(x)$, $G(x)$ also become nearly orthogonal $\langle F(x)|G(x)\rangle \approx 0$. This reveals the OP-built correlations between parameters x_1, x_2 in the two species.

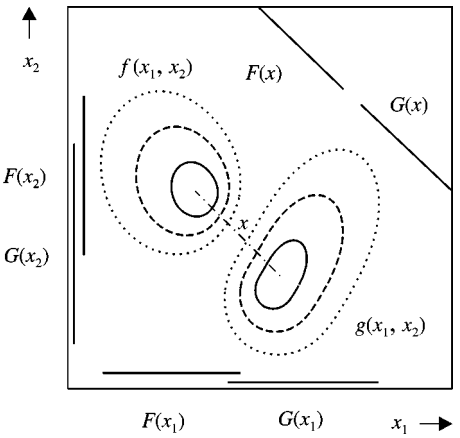


FIG. 1. Scheme of the distributions $f(x_1, x_2)$, $g(x_1, x_2)$ for two species, quasi-orthogonal ($\langle f|g\rangle \approx 0$) in the parameters x_1, x_2. The projected distributions $F(x_1)$, $G(x_1)$ or $F(x_2)$, $G(x_2)$ overlap $\langle F(x_{1(2)})|G(x_{1(2)})\rangle \neq 0$, while the distributions $F(x)$, $G(x)$ resulting from the projection along the effective orthogonalization variable x are still nearly orthogonal $\langle F(x)|G(x)\rangle \approx 0$.

Orthogonalization in Sympatric Speciation

Investigation of OP in species is a demanding task, due to the formation of the complex correlations between species parameters and the different involved levels. OMAL processes can often be observed in evolutionary or functionally close species. *Sympatric speciation* is very interesting from the point of OP, with its many mutually interconnected aspects, like the presence of competition or the role of assortativity in sexual interbreeding (Doebeli, 1996b; Kondrashov & Kondrashov, 1999; Dieckman & Doebeli, 1999). The activity of OMAL could help us to understand some problems of sympatry, like its possible existence on a "flat landscape" solely due to competition (Rosenzweig, 1978).

SPECIATION EQUATION

Here, we model sympatric speciation on the macro-level, where we investigate the presence of

OMAL. For simplicity, we do not explicitly model here the parallel fixation on the micro-level. Only sexually reproducing organisms are considered, but the results could also be extended to clonal species. We describe the total character distribution $f(\mathbf{x})$ by the new equation

$$\frac{\partial f(\mathbf{x})}{\partial t} = \mathscr{R}(\Phi, \mathbf{x}) \int\int d\mathbf{x}_{\mathrm{I}}\, d\mathbf{x}_{\mathrm{II}}\ \mathscr{S}(\mathbf{x}_{\mathrm{I}}, \mathbf{x}_{\mathrm{II}}) \times (f(\mathbf{x}-\mathbf{x}_{\mathrm{I}})\, f(\mathbf{x}+\mathbf{x}_{\mathrm{II}}))^{\alpha} - \frac{f(\mathbf{x})}{\tau(\mathbf{x})} \tag{4}$$

and assume that $f(\mathbf{x})$ splits into several subpopulations $f_i(\mathbf{x})$ for *individual species* with a limited interbreeding. The front term on the r.h.s., representing growth of the species, is controlled by the effective resources $\mathscr{R}$, with abiotic and biotic components, and dependence on factors from the set Φ, like temperature or radiation. It is also influenced by the sexual function $\mathscr{S}$ with a second-order mating (two partners) in a power α. The last term in eqn (4) represents dying of individuals with a time $\tau(\mathbf{x})$, due to natural reasons, predator–prey coupling or catastrophes. It can also reflect possible extinctions, since $f(\mathbf{x})$ would keep a constant norm for $\tau(\mathbf{x}) \approx \infty$.

RESOURCES FUNCTION

The resources function $\mathscr{R}$ plays the role of a "fitness landscape", considered in other speciation studies (Peliti, 2000). We use it in the form

$$\mathscr{R}(\Phi, \mathbf{x}) = \mathscr{R}_0(\Phi, \mathbf{x}) \times \exp\left\{-\left(\int d\bar{\mathbf{x}}\ \mathscr{R}^C(\Phi, \mathbf{x}-\bar{\mathbf{x}})\, f(\bar{\mathbf{x}})\right)^{\beta}\right\}, \tag{5}$$

where competition can be largely varied. Here, $\mathscr{R}_0(\Phi, \mathbf{x})$ represents all the resources available for an individual with parameter $\mathbf{x}$, in the absence of other individuals, and the external parameters Φ. We simply assume that it is given by the Gaussian function $\mathscr{R}_0(\mathbf{x}) = R_0\ \mathrm{e}^{-|\mathbf{x}-\mathbf{x}_R|^2/(\sigma_R)^2}$. Exploitation of the resources is controlled by the exponential function in eqn (5) with the distribution $f(\mathbf{x})$, which is competition-broadened by the function $\mathscr{R}^C$. This broadening expresses the fact that individuals with parameter $\mathbf{x}'$ can eat the food already consumed by individuals with parameter $\mathbf{x}$. If we assume that $\mathscr{R}^C$ is also Gaussian, its width σ_C determines the effective "distance" of the consumption/competition, and additional nonlinearity of this process can be tuned by the parameter β.

Roughening of the landscape $\mathscr{R}_0(\Phi, \mathbf{x})$ reflects the discrete character of species, which form the resources, and the structured ability of other species to consume them. In principle, both can be related to a "limited scaling" in the species parameters. For example, individuals of some species cannot get just smaller without qualitatively changing their properties, like the ability to hide by digging a hole, because of the increasing danger from predators, as they become smaller. In this way, complex correlations in species parameters are built by OMAL, which are roughening the landscape $\mathscr{R}_0(\Phi, \mathbf{x})$. We do not explicitly consider here the food-web structure, where landscape roughening could propagate, and also neglect predator–prey dynamics.

SEXUAL TERM

In the growth term from eqn (4), we assume that the progeny population $f(\mathbf{x})$ receives the properties $\mathbf{x}$ from the parents with distributions $f(\mathbf{x}-\mathbf{x}_{\mathrm{I}}), f(\mathbf{x}+\mathbf{x}_{\mathrm{II}})$. In the *mean heritage* approximation, only parents with $\mathbf{x}_{\mathrm{I}} = \mathbf{x}_{\mathrm{II}}$ contribute to this solution. The dependence of the birth rate on the distance of parents parameters $\mathbf{X} = \mathbf{x}_{\mathrm{I}} + \mathbf{x}_{\mathrm{II}}$ is given by the sexual correlation function $\mathscr{S}(\mathbf{X}, \xi)$. Declination of the progeny properties from this mean solution is described by the difference $\xi = \mathbf{x}_{\mathrm{I}} - \mathbf{x}_{\mathrm{II}}$, which originates in mutations and other effects, like gene mixing and activation. We can again assume that $\mathscr{S}$ has a Gaussian form

$$\mathscr{S}(\mathbf{X}, \xi) = S^0\, \frac{\mathrm{e}^{-|\mathbf{X}|^2/\sigma_S^2 - |\xi|^2/\sigma_M^2}}{2\pi\sigma_S\sigma_M}, \tag{6}$$

where the width of sexual interests σ_S and the declination σ_M can be partly tuned by OMAL. The factor α in eqn (4) reflects the "power of interbreeding". Speciation is rather sensitive to the departure of the equation from the quasi-linearity $\alpha \approx 0.5$, as we discuss below. The nonlinearity could be also incorporated into a more general version of the decay term $-f(\mathbf{x})/\tau(\mathbf{x})$ in

eqn (4), where it would reflect a decreased defense ability of small size populations.

Numerical Studies

In numerical studies, we use projected distributions with just one scalar property x for the *effective orthogonalization variable* (see Fig. 1), in which the species overlap the least. We assume, for simplicity, that the projected distribution $F(x)$ (called here also $f(x)$) follows the same eqn (4) as the full distribution $f(x_1, x_2, \ldots, x_n)$. We define the formed species i as separate subpopulations $f_i(x)$ in the character distribution $f(x)$, similarly as in Doebeli (1996b) and Drossel & Mac Kane (2000). The species i, j are considered to be orthogonal if their final steady-state distributions $f_i(x)$, $f_j(x)$ do not overlap $\langle f_i | f_j \rangle \approx 0$. The subpopulations of the consumption/competition function $f_C(x)$ can still overlap $\langle f_{Ci} | f_{Cj} \rangle \neq 0$, which leads to the fact that the species separation in the variable x is mostly controlled by the finite consumption/competition width σ_C.

We transform in eqn (4) the internal variables $(x_{\mathrm{I}}, x_{\mathrm{II}})$ or equivalently (X, ξ) to the new variables $(z, \bar{z})$, where $X = 2\bar{z}$, $\xi = 2(x - z)$. Then the equation reads

$$\frac{\partial f(x)}{\partial t} = \mathscr{R}(x) \iint 8 \, \mathrm{d}z \, \mathrm{d}\bar{z} \, \frac{\mathrm{e}^{-4\bar{z}^2/\sigma_S^2 - 4(x-z)^2/\sigma_M^2}}{10\pi\sigma_S\sigma_M} \times \left(f(z - \bar{z}) \, f(z + \bar{z}) \right)^{\alpha} - \frac{f(x)}{\tau}, \tag{7}$$

where, the resources function is

$$\mathscr{R}(x) = R_0 \, \mathrm{e}^{-x^2/\sigma_R^2 - a_C (f_C(x))^{\beta}},$$

$$f_C(x) = \int \mathrm{d}\bar{x} \, \frac{\mathrm{e}^{-(x-\bar{x})^2/\sigma_C}}{\sqrt{2\pi}\sigma_C} f(\bar{x}). \tag{8}$$

In eqns (7) and (8) we use the values $S^0 = 0.2$ and $a_C = 4$ to easily present numerical results for different functions.

PRACTICAL EXAMPLES

We use eqns (7) and (8) to study the conditions under which sympatric speciation can be obtained on broad Gaussian resources, and also pay attention to the role of their local inhomogeneities. We also specify the role of OMAL in the obtained solutions.

In Fig. 2, we present the steady-state (globally stable) numerical solutions of eqns (7) and (8) for species formed by sympatric speciation. The dashed and dash-dotted curves correspond to the free and unused resources without and with the competition exponential from eqn (8), respectively. The solid and dotted curves represent the *total* population $f(x)$ and the consumption/competition function $f_C(x)$, respectively. The situations with one to four species are obtained for the Gaussian resources of the width $\sigma_R = 0.2$ and strengths $R_0 = 0.5$, 1, 2, 4. The consumption/competition width is $\sigma_C = 0.1$, the sexual and mutation widths are chosen the same $\sigma_S = \sigma_M = 0.015$, the power parameters are $\alpha = \beta = 0.6$ and the lifetime is $\tau = 100$.

The broad Gaussian resources can be seen as "quasi-flat", since σ_R is larger than other parameters in the model. When the richness of resources R_0 is rather small, only σ_C is close in size to σ_R. As R_0 increases, an *effective width* of resources Σ_R (on which species can survive) becomes larger than σ_R, so the resources can feed more and more species. If σ_R alone (not R_0) is enlarged to get a nearly flat landscape, the number of species grows, but their separation is always given by σ_C. If σ_C does not vary much, the resulting self-organized *speciation quasi-periodicity* as a function of x determines a universal *speciation volume*. In the present broad Gaussian landscape, with an exponential form of competition [see eqn (8)], speciation can be obtained only if the power of interbreeding is slightly nonlinear $\alpha > 0.5$. This adds to the speciation conditions on a quasi-flat landscape.

The unused resources $(- \cdot - \cdot)$ in Fig. 2 are *flattened* by the total distribution $f_C(x)$, which largely copies the shape of the resources $(- - -)$, even though the total distribution $f(x)$ has sharply separated subpopulations. This shows an efficient sharing of resources by the orthogonalized consumption needs of the species (Roughgarden, 1976). We have used eqn (3) to calculate the relative overlaps $r_{Cij} = \langle f_{Ci} | f_{Cj} \rangle / \langle f_C | f_C \rangle$ of the individual contributions $f_{Ci}(x)$ to $f_C(x)$. These $f_{Ci}(x)$ can be obtained from eqn (8), where we substitute the individual distributions $f_i(x)$.

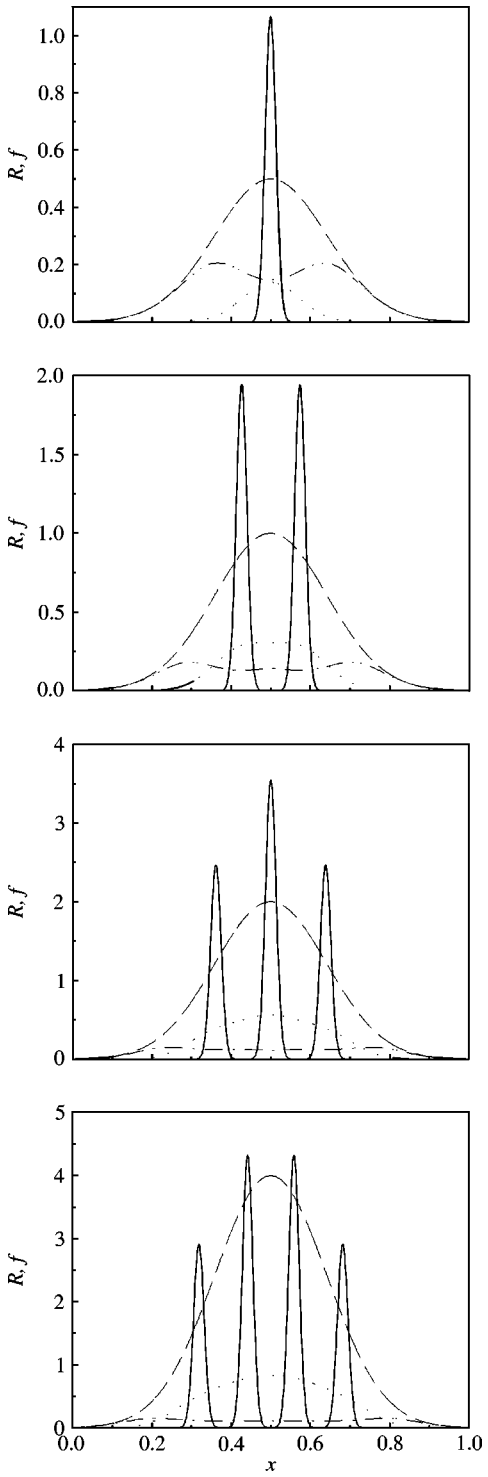


FIG. 2. Species population f (competitive population f_C) in a sympatric speciation is shown by full (dotted) lines as a function of the effective parameter x. The dashed (dash-dotted) lines represent the free (competitive) landscape of resources. The plots with one to four species correspond to strength of resources $R_0 = 0.5, 1, 2, 4$ and other parameters in the text.

The situations with $n = 2$–4 species in Fig. 2 give: $n = 2$, $r_{12} = 0.296$; $n = 3$, $r_{12} = 0.219$, $r_{13} = 0.045$ and $n = 4$, $r_{12} = 0.17$, $r_{13} = 0.062$, $r_{14} = 0.008, r_{23} = 0.213$. The values r_{Cij} show the steady-state strength of coupling of the orthogonal species ($\langle f_i | f_j \rangle \approx 0$). The related pressures can be *relaxed by inhomogeneities* in the broad Gaussian resources.

Note also that the final distributions $f_i(x)$ in Fig. 2 do not depend on the size of the initial Gaussian population used in iterations of eqns (7) and (8), unless this is very small. Then, for the present parameters, initial populations which are by 2–3 orders smaller than the final population tend to follow the single-peaked landscape. This means that the final solution gives $2n - 1$ species where $2n$ species would normally appear, so the initial condition is "frozen in" the final state of a reasonable stability. For even smaller initial populations, species die out, due to finite lifetime τ.

Stepwise speciation resembling our results has been observed, for example, in stickleback species in small lakes of the coastal British Columbia (Schluter & Mac Phail, 1992). Speciation driven by the increased strength of resources could also be the reason for diversity gradients of marine life, as induced by the solar radiation in coastal regions (Roy *et al.*, 1998). Similarly, richness gradients of other species observed towards the equator (Rapoport, 1975) could be the consequence of *enlarged radiation, temperature and food zones*.

In Fig. 3, we show the size f_M and position x_M of the subpopulation maxima as a function of the resources strength R_0 and other parameters as in Fig. 2. In the top plot for x_M, we can see that after a new species emerges, the positions x_M for the peaks slightly move. Since the Gaussian resources rapidly fall down at the tails, the effective width Σ_R increases very slowly with R_0, and so does the number of species n for $n > 3$. For example, the fifth species appears only at $R_0 \approx 13$. In the bottom plot, we show the dependence of the maxima f_M on R_0, and add the number of peaks in each curve. As R_0 grows and the number of species increases from odd to even, the main maxima jump in values for this shape of resources, while in transitions from even to odd the changes are reasonably continuous. The

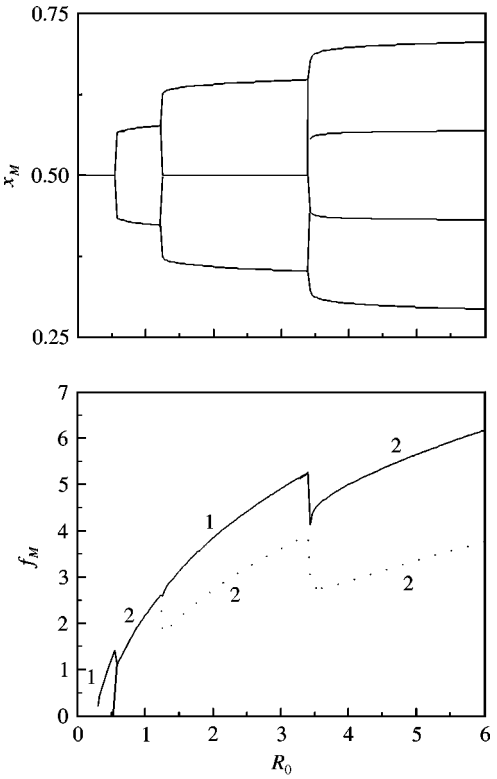


FIG. 3. (Top plot) The maxima x_M of the populations, for the species with parameters in Fig. 2, as a function of the resources strength R_0. (Bottom plot) The same dependence for the maxima of the species populations f_M, with the number of peaks denoted.

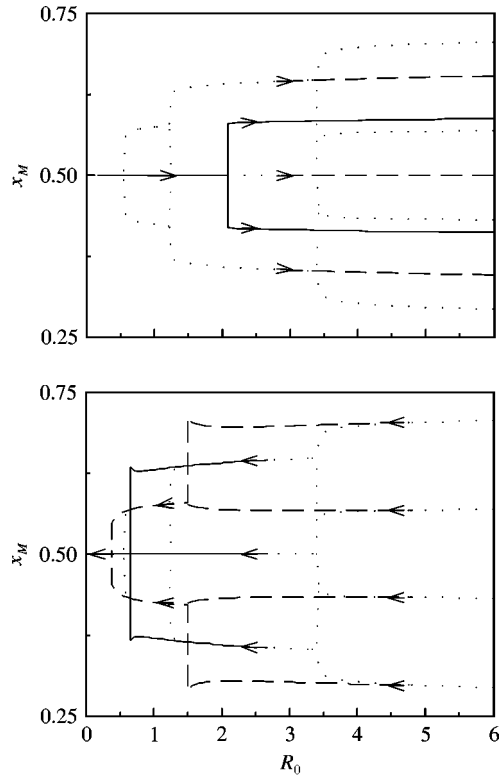


FIG. 4. Positions x_M of the species as the resources richness R_0 is slowly varied. The upper (lower) plot shows situations for increasing (decreasing) R_0. Thin dotted lines correspond to the globally stable solutions from Fig. 3. In the upper plot, thick solid (dashed) lines present runs starting from regions with a single (triple) species. In the lower plot, the solid (dashed) lines present runs starting from regions with a triple (quadruple) species. In all cases, speciation freezing, dependent on history, can be observed.

separation of species parameters (character shift) shows that OMAL is accompanied by a *pressure* between the coexisting species (Slatkin, 1980), induced by the overlap of their $f_{Ci}(x)$ distributions. The activity is especially large close to speciation events, where the species parameters move. Experimentally observed pressure-induced variation of the character shift, as the species move between an island and a continent, is discussed for example by Schluter (1988).

Population branching could also be studied in a time region (Kisdi, 1999). In our work, globally stable solutions in Fig. 3 are calculated by using at each point R_0 a Gaussian initial population comparable in size with the final value. We will also study a time evolution, but limit to the investigation of speciation for an *adiabatic* time evolution of the resources richness R_0; we slightly change R_0 in each time run, and start from the populations of the previous R_0. The results for the species positions x_M are shown in Fig. 4, where thin dotted lines correspond to the solutions from Fig. 3. In the upper plot, we show by thick solid (dashed) lines, runs with increasing R_0 starting from the situation in the single-species (triple-species) region. In the former case, the first speciation occurs late in the double-species region, while three species would appear only deeply in the region where four species should be; a start from the double-species region follows this solution as well. If we start from the triple-species region, the number of species remains unchanged for the used parameters.

If the resources become poorer, the situation is rather different, as we show in the bottom plot. A start from the four-species region provides evolution skipping over solutions with three species, but it "smoothly transfers in advance" to a solution with two species, and the solution with one species appears delayed. Starting from the

triple-species region, the evolution skips over solutions with two species, and later jumps to a solution with one species. Thus, if R_0 increases, all numbers of species appear, but they are delayed, while if R_0 decreases, sequences with odd or even number of species seem to be realized. Since the frozen non-equilibrium solutions are locally stable with respect to perturbations, the system experiences *hysteresis* in the number of species if R_0 slowly oscillates with a large enough amplitude. It is likely that the species parameters would change by OMAL, to partially follow the slow evolution of resources and bypass the mounting pressures from a fixed number of species. The crucial aspect is the ratio of time-scales of these effects. In nature, such a behavior can also be largely stabilized by year seasons and other random effects, but if we take into account the evolution on micro-level, one could essentially face only *irreversible behavior*.

In Fig. 5, we show speciation in the previous landscapes, modified by additional disturbances. The disturbances can *trigger speciation* at smaller competition pressures, induced by the f_{Ci} distributions; speciation could even take place for $\alpha = 0.5$, which would not be possible on the broad Gaussian landscape. In the upper plot, we take the situation with one species in Fig. 2, and subtract from the resources $R_0\, e^{-x^2/\sigma_R^2}$ the exponential $R_1\, e^{-x^2/\sigma_F^2}$ with $R_1 = 0.1$, $\sigma_F = 0.07$. The resulting landscape flattening can split the single species in two subpopulations, since the resources become more efficiently explored by two species. We have also found such splitting/joining jumps for the case of two or three species. For $R_0 = 1.5$ the number of species goes from three to two, while for $R_0 = 2.5$ it goes from three to four, copying thereby in both cases the flatter shape of resources.

From a theoretical point of view, we could also consider that x represents one of the parameters Φ, like temperature, radiation or humidity. Then, the resources show how much food for the species is available at this parameter x, and other changes are analogous. Therefore, we should in principle be able to describe certain situations with *allopatric speciation*. For example, splitting of a Drosophila population, similar to the top part of Fig. 5, was observed on a microsite in a valley with two differently irradiated slopes,

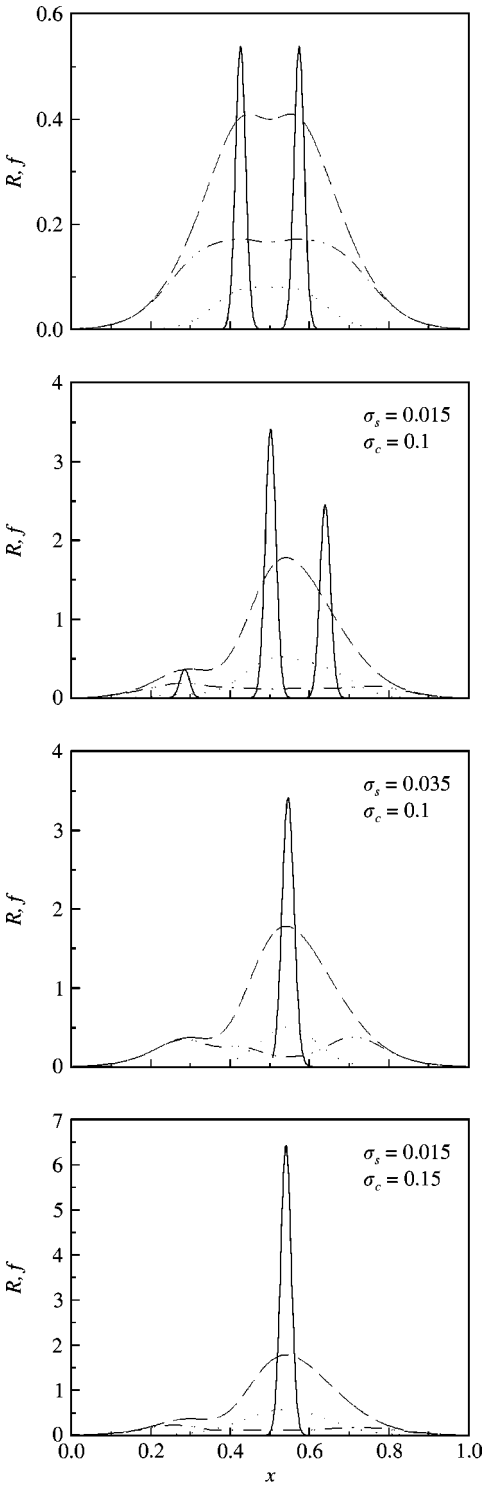


FIG. 5. The situations from Fig. 2 presented for a perturbed Gaussian landscape. The top plot shows how the single species from Fig. 2 breaks into two, when the resources are flattened. Next, one of the side species, from the third plot in Fig. 2, shifts and decreases in size to fit the new landscape. In the third plot, for the same parameters, but the sexuality width $\sigma_S = 0.035$, the side peaks disappear, but the resources are not well explored. This is improved on the last plot, where, instead of σ_S, the consumption width is broadened to the value $\sigma_C = 0.15$.

giving a radiation/temperature space gradient corresponding to our parameter x (Harry *et al.*, 1999). A sharply different character of the slope sides could give flatter resources in the middle radiation/temperatures with side maxima as in Fig. 5.

In the second plot of Fig. 5, we present the population for the situation with three species in Fig. 2, where we subtract, similarly as before, the shifted exponential $R_1\ \mathrm{e}^{-(x+x_1)^2/\sigma_F^2}$ with $R_1 = 1$, $x_1 = 0.1$ and $\sigma_F = 0.07$. Due to this asymmetry of the resources, the first side species largely shifts and decreases, in order to efficiently explore the new landscape, but the total number of species still remains the same. This inhomogeneity-shifted species competes a little with the other species. In the next plot, we use the same resources, but we broaden the sexual width to $\sigma_S = 0.035$. This decreased assortativity fully prevents speciation, and the resources are poorly used in the tails. Much better exploration is obtained in the last plot, where, instead, the consumption width is increased by 50% to $\sigma_C = 0.15$, so the single species population is also larger. The width σ_M is also crucial, since it can broaden and smooth the population. For example, the single-peaked distribution in the third plot of Fig. 5 can also be obtained if σ_M is increased to the value 0.035.

GENERAL TUNING OF PARAMETERS

Other parameters in eqns (7) and (8) could also interplay in a sensitive way. For example, the power α can fluctuate around the value $\alpha \approx 0.5$ to promote/suppress speciation. The power β can reflect the method of food consumption as well as possible deterrence actions aimed in protecting it (screening). For $\beta < 0.5$ ($\beta > 0.5$) side populations become suppressed (enhanced) with respect to the middle populations. The absolute strength of consumption/competition a_C can be tuned to control the size of the population. Finally, the time τ maintains the speciation pressure and maintains the population stability (Smith & Brown, 1986). The term with τ in eqn (4) diminishes individuals present in regions between the landscape maxima. It also gives sharper and a more regular speciation. Without competition and with large τ, the population would simply follow the landscape, while with competition, it would not speciate sharply until τ is decreased.

Recent studies point to the increased role of food-web resources in species survival (Amaral & Meyer, 1999). Proportionality between the number of species and the resources richness, obtained above, can also be supported in different ways. For example, in poor resources, generalist species might be triggered by broadening σ_C and σ_S. Clusters of species can be extinct, and surviving species can broaden their distributions, so that the correlations formed by OMAL decrease. Since mutations are less suppressed, speciation could eventually be also promoted in these conditions. If the resources get richer, individualists could do better. Species can narrow their consumption needs σ_C, but they need to live efficiently in other ways. Consequently, they also narrow their sexual interests σ_S and trigger further speciation. Since the populations largely copy resources, their richness does not make life easier for the numerous individuals. This drives the lasting self-organized evolution. OMAL can tune all these parameters in a limited amount to balance the coexistence of species with an efficiency exploration of resources. The time-scale of this activity goes over many generations, when these changes get fixed on the micro-level.

FIXATION OF OMAL

We noted that neutral sections of proteins, with a relatively stable composition in each individual species, might be used for the transcription of OMAL. This can support selectionist views on the problem of *molecular evolution clock* (Ohta, 1996), which underlines the quasi-deterministic origin of observed constant mutation rates in neutral sections of proteins. Recent observations show, that when bacteria are placed in a new environment, some of their genes start to mutate faster (Sniegowski *et al.*, 1997; Schmid & Tautz, 1997). The growth of RNA *in vitro* also shows that under strong selections the molecules go to the same limit form (Strunk & Ederhof, 1997). These experiments point to a possibly deterministic role of the environment on the proteins, where OMAL takes an active role.

We can imagine, that tuning of protein catalytic strength can be regulated by other proteins,

enforced by the complex molecular environment, evolving under the pressure of globally changing conditions on the macro-level. If these possibilities are exploited, and a larger change is needed, a jump in neutral positions of the tuned proteins could be induced by OMAL, which is analogous to avalanche ticking in other systems (Wu *et al*., 1993). This ticking could largely go *in parallel* with the formation of new species, shown for example in Fig. 2.

These ideas could be tested on bacteria, with short reproduction periods. A working scenario could be obtained by modifying some neutral protein region, and keeping the remaining gene-pool frozen by a periodical exchange. If OMAL is relevant, the protein would eventually evolve in many bacteria generations to the same or a similar form. The problem is whether OMAL can be reached in artificial laboratory conditions. Such experiments could also reveal as to which extent one can speak about proteins in *non-equilibrium* with the cell environment.

Summary

We have introduced the concept of species orthogonalization on the micro and macro-level. OMAL is related with the need for the formation of complementary functions and habits of different biological species living in close contacts. OMIL separates functions of molecular species, and assists OMAL on the micro-level. We have shortly discussed the ways of evaluating orthogonalization processes.

As a practical example, we have investigated OMAL in sympatric speciation, but limited our study so far to the macro-level. To this goal, we have found a new population equation, and presented it in several numerical examples. Its solutions, for the projected distribution with one effective variable x, show that the number of species grows in richer resources, divided between the species. This could help to explain the increase of species diversity close to the equator, where sun radiation provides richer resources. We have also observed speciation dependent on history, which might be useful in the modeling of ecology, and discussed possible ways of OMAL fixation on the micro-level. We expect that orthogonalization mechanisms are active in other systems with competitive structures.

I would like to acknowledge T. Pavlíček for stimulating and encouraging discussions. I would also like to kindly thank M. Shapiro for a partial financial support.

REFERENCES

AMARAL, L. A. N. & MEYER, M. (1999). Environmental changes, coexistence, and patterns in the fossil record. *Phys. Rev. Lett*. **82,** 652–655.

BAAKE, E., BAAKE, M. & WAGNER, M. (1997). Ising quantum chain is equivalent to a model of biological evolution. *Phys. Rev. Lett*. **78,** 559–563.

BAGNOLI, F. & BEZZI, M. (1997). Speciation as pattern formation by competition in a smooth fitness landscape. *Phys. Rev. Lett*. **79,** 3302–3305.

BROWN W. L. & WILSON E. O., (1956). Character displacement. *Syst. Zool*. **5,** 49–64.

CREIGHTON, T. E. (1992). *Proteins. Structure and Molecular Properties*. New York: Freeman.

DIECKMAN, U. & DOEBELI, M. (1999). On the origin of species by sympatric speciation. *Nature* **400,** 354–357.

DOEBELI, M. (1996a). An explicit genetic model for ecological character displacement. *Ecology* **77,** 510–520.

DOEBELI, M. (1996b). A quantitative genetic model for sympatric speciation. *J. Evol. Biol*. **9,** 893–909.

DROSSEL, B. & MAC KANE, A. (1999). Ecological character displacement in quantitative genetic models. *J. theor. Biol*. **196,** 363–376.

DROSSEL, B. & MAC KANE, A. (2000). Competitive speciation in quantitative genetic models. *J. theor. Biol*. **204,** 467–478.

EIGEN, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwisseschaften* **58,** 465–523.

EIGEN, M. (1993). Viral quasi-species. *Sci. Am*. **269,** 42–49.

EIGEN, M. & SCHUSTER, P. (1977). The hypercycle. A principle of natural self-organisation, Oart A: emergenece of the hypercycle. *Naturwisseschaften* **64,** 541–565.

EIGEN, M., MCCASKILL, J. & SCHUSTER, P. (1989). The molecular quasi-species. *Adv. Chem. Phys*. **75,** 149–263.

FONTANA, W. & SCHUSTER, P. (1998). Shaping space: the possible and the attainable in RNA genotype–phenotype mapping. *J. theor. Biol*. **194,** 491–515.

HARRY, M., RASHKOVETSKY, E., PAVLICEK, T., BAKER, S., DERZHAVETS, E. M., CAPY, P. (1999). Fine-scale biodiversity of Drosophilidae in "Evolution Canyon" at the Lower Nahal Oren Microsite, Israel. *Biologia* **54,** 685–705.

HOLM, L. & SANDER, C. (1998). Dictionary of recurrent domains in protein structures. *Proteins* **33,** 88–96.

HUTCHINSON, G. E. (1968). When are species necessary? In: *Population Biology and Evolution* (R. C. Lewontin, ed.), pp. 177–186. Syracuse, New York: Syracuse University Press.

IRWIN, D. E. & PRICE, T. (1999). Sexual imprinting, learning and speciation. *Heredity* **82,** 347–354.

KISDI, E. (1999). Evolutionary branching under asymmetric competition. *J. theor. Biol*. **197,** 149–162.

KONDRASHOV, A. S. & KONDRASHOV, F. A. (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400,** 351–354.

LESK, A. M. (1951). In: *Protein architecture* (Rickwood, D. & Hames, B. D., eds). Oxford: Oxford University Press, 1991.

LUMSDEN, CH. J., BRANDTS, W. A. & TRAINOR, L. E. H. (1997). *Physical Theory in Biology*. Singapore: World Scientific; L. Luo p. 380.

MAY, R. M. & MAC ARTHUR, R. H. (1972). Niche overlap as a function of environmental variability. *Proc. Natl Acad. Sci. U.S.A.* **69,** 1103–1109.

MAYR, E. (1942). *Systematics and the Origin of Species*. New York: Columbia University Press.

MAYR, E. (1963). *Animal Species and Evolution*. Cambridge, MA: Belknapp Press.

OHTA, T. (1996). The neutralist-selectionist debate. *BioEssays*, **18,** 673–683.

PELITI, L. (2000). Fitness Landscapes and Evolution (cond-mat/9505003).

PLOTKIN, J. B., POTTS M. D., YU D. W., BUNYAVEJCHEWIN S., CONDIT R., FOSTER R., HUBBELL S., (2000). Predicting species diversity in tropical forests. *Proc. Natl Acad. Sci. U.S.A.* **97,** 10 850–10 854.

PRÜGEL-BENNETT, A. (1997). Modeling evolving populations. *J. theor. Biol.* **185,** 81–95; **82,** 1983–1986.

PRUGOVEČKI, E. (1981). *Quantum Mechanics in Hilbert Space*. (New York: Academic Press).

RAPOPORT, E. H. (1975). *Areografia: estrategias de las especieses*. Mexico City: Fondo de Cultura Economica.

ROUGHGARDEN, J. (1976). Resources partitioning among competing species—a coevolutionary approach. *Theor. Popul. Biol.* **9,** 388–424.

ROSENZWEIG, M. L. (1978). Competitive speciation. *Biol. J. Lin. Soc.* **10,** 275–289.

ROSENZWEIG, M. L. (1995). *Species Diversity in Space and Time*. NY: Cambridge University Press.

ROY, K., JABLONSKI, D., VALENTINE, J. W. & ROSENBERG, G. (1998). Marine latitudinal diversity gradients: tests of causal hypotheses. *Proc. Natl Acad. Sci. U.S.A.* **95,** 3699–3702.

SCHLUTER, D. (1988). Charactery displacement and the adaptive divergence of finches on islands and continents. *Am. Nat.* **131,** 799–824.

SCHLUTER, D. & MAC PHAIL, J. P. (1992). Ecological charactery displacement in sticklebads. *Am. Nat.* **140,** 85–108.

SCHMID, K. J. & TAUTZ, D. (1997). A screen for fast evolving genes from Drosophila. *Proc. Natl Acad. Sci. U.S.A.* **94,** 9746–9750.

SEGRÉ, D., BEN-ELI, D. & LANCET, D. (2000). Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proc. Natl Acad. Sci. U.S.A.* **97,** 4112–4117.

SLATKIN, M., (1980). Ecological character displacement. *Ecology* **61,** 163–177.

SMITH, J. M. (1966). Sympatric speciation. *Am. Nat.* **100,** 637–650.

SMITH, J. M. & BROWN, R. L. W. (1986). Competition and body size. *Theor. Popul. Biol.* **30,** 166–179.

SNIEGOWSKI, P. D., GERRISH, P. J. & LENSKI, R. E. (1997). Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387,** 703–705.

STRUNK, G. & EDERHOF, T. (1997). Machines for automated evolution experiments in vitro based on the serial-transfer concept. *Biophys. Chem.* **66,** 193–202.

VALENTINE, J. W. & WALKER, T. D. (1988). Diversity trends within a model taxonomic hierarchy. *Physica* **D 22,** 31–42.

WANG, ZHI-XIN. (1996). How many fold types of protein are there in nature? *Proteins* **26,** 186–191.

WU, X-L., MÅLØY K. J., HANSEN A., AMMI M., & BIDEAU D. (1993). Why hour glasses tick. *Phys. Rev. Lett.* **71,** 1363–1366.