

Emergence of Mutualism: Application of a Differential Model to the Coelenterates- Algae Associations

G rard Weisbuch
Guillemette Duchateau

SFI WORKING PAPER: 1992-06-030

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

 NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu



SANTA FE INSTITUTE

Emergence of Mutualism

Application of a Differential Model to the Cœlenterates-algae Associations.

Gérard Weisbuch and Guillemette Duchateau

Ecole Normale Supérieure

Laboratoire de Physique statistique
24 rue Lhomond
75005 Paris
Email: weisbuch@merlin.ens.fr

URA 686 CNRS - Groupe de Bioinformatique
46 rue d'Ulm
75005 Paris
Email: duchat@wotan.ens.fr

Abstract

A population dynamics approach based on a system of differential equations allows us to establish conditions for the emergence of mutualism for cases such as coelenterates-algae symbionts. A central assumption of the model is that a host organism is able to discriminate, via some molecular recognition mechanisms, among different invading organisms and preferentially reject *egoists* rather than *bona fide symbionts*. Large differential rejection rates allow the emergence of mutualism. Different attractors of the population dynamics correspond to the emergence of mutualism, predominance of "egoist" species, or coexistence of many species.

1 Introduction

1.1 The paradox of mutualism

Mutualistic systems are known to occur in nature, e.g., lichen made of algae and fungus, corals-zooxanthellae and hydra-zoochlorellae. Exchanges among the partners are beneficial for the species involved, which live together in close association. The benefits exchanged can be food, energy, protection-habitat and transport (pollination) {Boucher, James and Keeler, 1982}. According to {Begon, Harper and Trunsend, 1986} most of the world's biomass is composed of mutualists: organisms in forests, meadows and corals are involved in symbiotic associations.

Mutualism assumes a large variety of types that can be classified according to the strength of the interaction. At one end, there is *commensalism* where there is a benefit for only one species among the species involved. At the other end, there is *obligate mutualism* where each species involved cannot even survive without the other. Between these extremes, lies a range of mutualistic interactions that show different dependencies {De Angelis, Post and Travis, 1986}. One can separate the *symbiotic mutualism* - where physiological links exist - from the *nonsymbiotic mutualism* - where physiological links do not exist.

The emergence and stability of mutualism constitutes a paradox in terms of individual selection of the fittest. The paradox is the following: since giving food to the other symbiont should be costly for the donor, we expect the donor to be disadvantaged in terms of fitness with respect to a more selfish species which would give nothing.

The purpose of this paper is to describe a mathematical model which accepts the premises of this assertion but refutes its conclusion. An important part of the argument is that selection occurs not only at the level of individual organisms, but also at the level of their mutualistic associations.

1.2 Previous theoretical studies

The stability of mutualistic ecosystems has been studied by a number of authors using cost-benefits analysis, standard non-linear differential systems or "Artificial Life" numerical simulations.

An early work is that of {Roughgarden, 1975} who did a cost benefit analysis of the exchanges among hosts and guests to describe damselfish-anemones mutualism. {Wilson, 1983} in his study of beetles-phoretic mite association insists on the idea of group rather than individual selection. He predicted that when populations are clustered in groups with a varying proportion of mutualists and selfish types, those group with more mutualists should do better and be selected.

The differential system approach is summarized in {De Angelis, Post and Travis, 1986}. It is based on either Volterra-Lotka formalism to characterize global stability or a more general form of the *per capita* rate of increase function to monitor local stability. Classical non-linear analysis criteria are used to obtain inequalities among equation coefficients that ensure stability. The absence

of stability among a set of host, parasite and true symbiont that exist as individual species leads these authors to suggest that mutualism can only exist when the relationship between the host and the guest involves a one-to-one relationship.

Artificial life simulations of {Ikegami and Kanebo, 1990} do not succeed in establishing mutualism permanently since they postulate the existence of parasitic mutants quite harmful to the host. Their model exhibit transient periods during which mutualistic species predominate, alternating with periods when egoist species exploit them. Artificial life models have also been used to study stability and breakdown of symbiotic associations {Taylor, Muscatine and Jefferson, 1989}.

A conclusion of many authors is that long term interactions among host and guest should be necessary for the establishment of mutualism. These conclusions also appear in studies about the increase of the stability of hypercycles by compartmentation {Szathmàry and Demeter, 1987} and in the iterated prisoner dilemma model of {Lindgren, 1992} where cooperation among players can only be established when players play long enough. The present paper works along these lines. We propose here a differential equation model that takes into account a dynamics of aggregation of organisms which is representative of the processes existing among coelenterates-algae associations and we study the range of parameter values for which this aggregation dynamics gives rise to mutualism.

1.3 The hyperbolic model

We use a model of population dynamics introduced in {Weisbuch, 1984}. With respect to the Volterra-Lotka approach, this model does not exhibit divergences of populations in the case of positive interactions among organisms and is analytically soluble in the low mutation rate limit.

The following differential system allows us to study population dynamics in the presence of mutations. Each population varies in time due to three terms:

- The first term describes population growth according to a fitness coefficient α and available resources f which are shared by all populations;

- The second term is simply the rate of death.

- The third term is a mutation term which decreases the population because of all possible point mutations or increases it because of mutations from other existing species j one mutation away.

$$\frac{dP_i}{dt} = \frac{\alpha_i \cdot f \cdot P_i}{P_t} - d \cdot P_i + m \cdot (-n \cdot P_i + \sum_j P_j) ; \text{ every } i$$

and (1.1)

$$P_t = \sum_i P_i$$

where P_i and P_j are the populations of i and j , P_t is the total population, d and m are the death and mutation rates and n is the number of genes of each organism.

The attractor of the system is analytically obtained by a perturbation technique {Morse and Feshback, 1953} when the mutation term is small with respect to the death term. The fittest population predominates in the total population. By equating its derivative to 0 and neglecting mutations from other species one obtains:

$$P_t = \frac{\alpha_m \cdot f}{d} = P_m \quad (1.2)$$

where α_m is the fitness coefficient of the fittest population and P_m is the fittest (or *dominant*) population (equating P_m to P_t is a guess which is supported by further analysis of higher order terms in m).

The populations of the nearest mutants are obtained by equating to zero their time derivative and taking into account the mutation term which comes from the fittest species:

$$\left(\frac{\alpha_i \cdot f}{P} - d \right) \cdot P_i + m \cdot P_m = 0$$

$$P_i = \frac{m \cdot P_m}{d \left(1 - \frac{\alpha_i}{\alpha_m} \right)} \quad (1.3)$$

In the limit of small m , the ratio of the first mutants to the fittest population scales as m/d . The same analysis can be carried to next mutants which decay in population by a factor m/d for each further mutation from the fittest. The perturbation technique described here will allow us to interpret the results obtained later.

2 Building-up of the model

2.1 Phylogeny and interactions

We shall now apply a modified version of the above formalism to a system made of five

populations, whose interactions and phylogeny are represented on Figure 1.

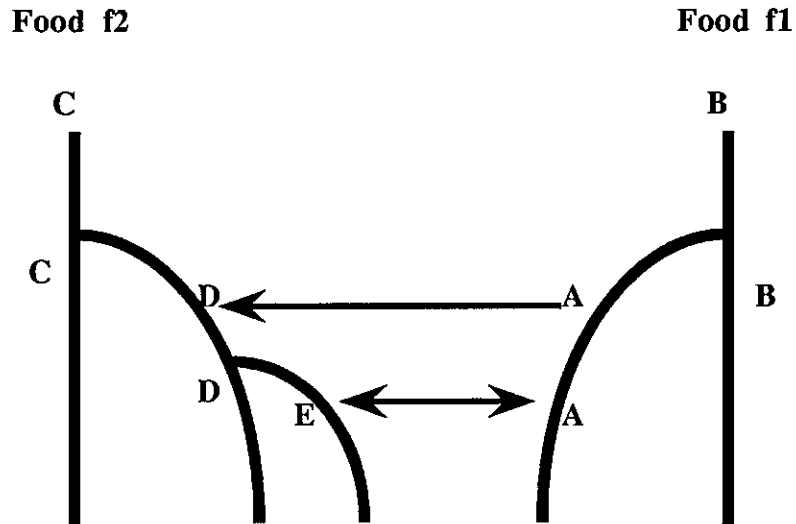


Figure 1 : Phylogeny and interactions of the 5-species system.
The populations of the organisms belonging to the two trees use different resources 1 and 2, available in quantity f_1 and f_2 . Horizontal arrows figure exchanges among organisms.

-Populations C and B are unrelated and are not involved in any interaction with other populations. They can be considered as primeval organisms. Their respective fitnesses are γ and β .

- Population A differs from population B by one mutation. Population A can be considered as a *host* for populations D and E. It produces some nutrients that can be used by D or E.

- D is a *selfish* organism and is further called *egoist guest* of A. It uses nutrients produced by A, but does not give anything in exchange to A.

- E is a *bona fide symbiont* of A. It uses nutrients produced by A and in exchange provides A with nutrients.

Organisms B and A feed on resource 1 available in quantity f_1 and organisms C, D and E on resource 2 available in quantity f_2 . From the hyperbolic model described in section 1.3, we *a priori* expect to find two dominant species, one for each resource tree.

2.2 The elementary processes

The ecosystem is made of free organisms A, B, C, D and E and couples AE and AD, when A is infested respectively by E or D. To simplify the model, we have first supposed that the host A offers only one site where E or D are able to bind. This simplification limits the set of possible associations to couples AD and AE. Changes in population sizes are due to elementary processes represented in Figure 2 and listed further.

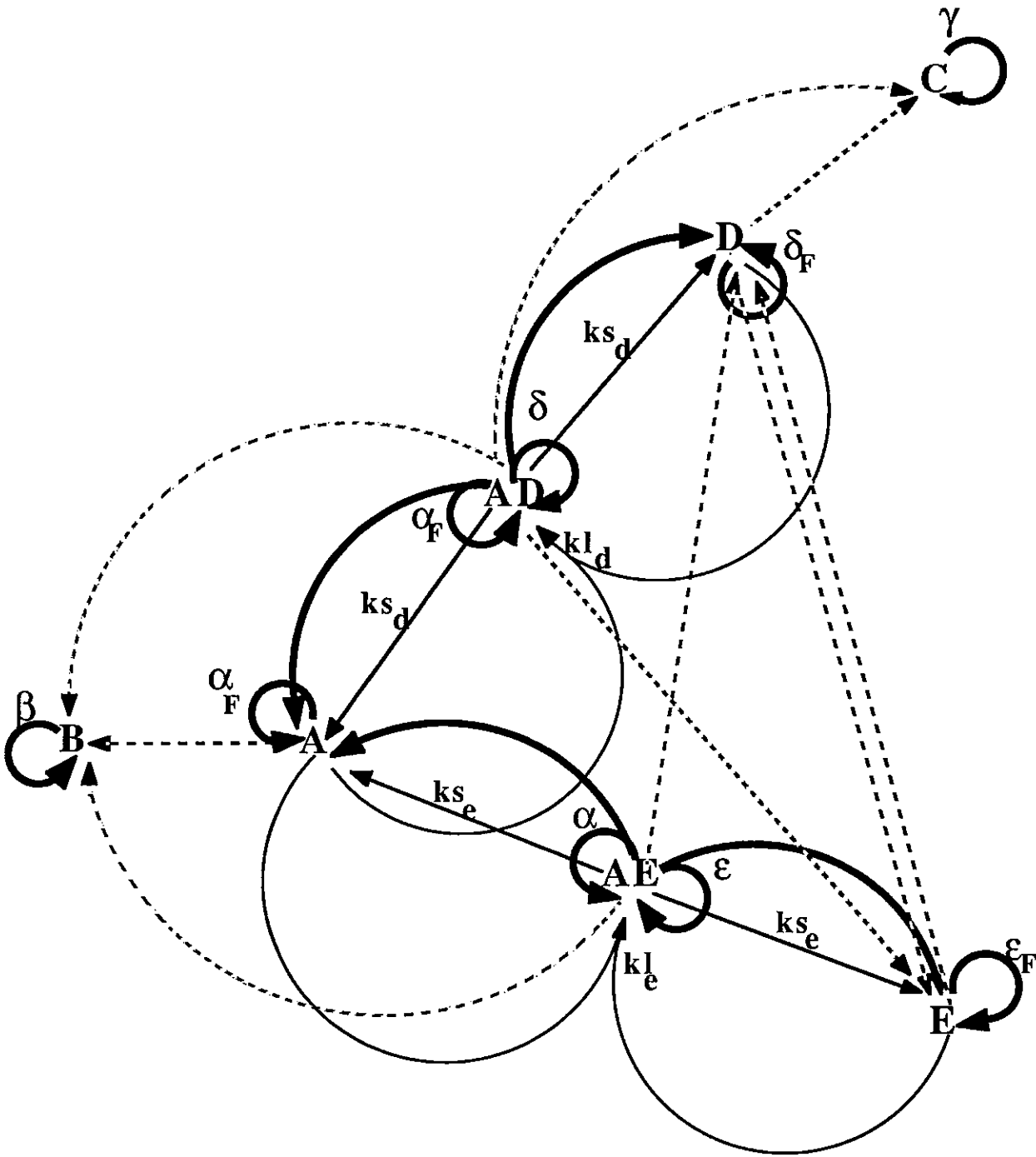
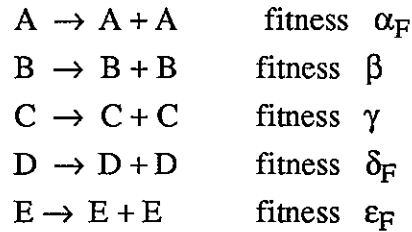


Figure 2: Set of the elementary processes that modify the populations.

Thin continuous lines represent fast association-dissociation processes between hosts and guests organisms with kinetic constants kl_e , ks_e , kl_d and ks_d . Bold lines represent reproduction and death processes (with rate d). The greek letters represent the fitness coefficients associated with the reproduction processes of the organisms, which are either free or associated with another organism belonging to a different species. Dotted lines represent mutation processes with rate m .

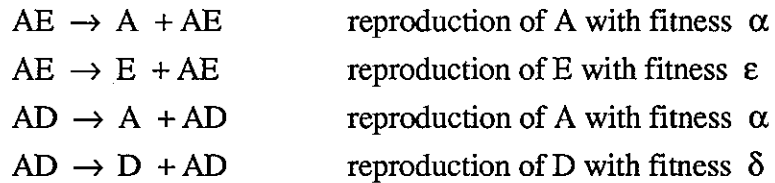
2.2.a Reproduction

Reproduction of free A, B, C, D and E obey simple rules such as:



The frequency of these processes is described as in equation 1 and involves fitness coefficients specific of the concerned species.

In the case of reproduction of one of the symbionts of AD or AE, the other one is liberated. For each one of AE or AD two processes can then occur:



The frequency of these processes is described as in equation 1 but it involves fitness coefficients which are different and eventually higher than those of the free species.

The fitness parameters verify the following relations:

$$\alpha > \beta > \alpha_F \quad (3.1)$$

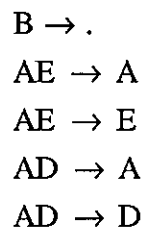
The fitness of B (β) is higher than the fitness of A (α_F) because A provides D and E with nutrients. But when E is bound to A, the fitness of A is larger than that of B thanks to the cooperation of E.

$$\varepsilon_F < \delta_F = \gamma < \varepsilon \leq \delta \quad (3.2)$$

The fitness of the E species (ε or ε_F) is inferior to the fitness of D (δ or δ_F) because E produces with some cost a nutrient for A. $\delta_F = \gamma$ since D does not provide nutrient to anyone. Fitnesses of D and E when associated with A are larger than fitnesses of the free organisms.

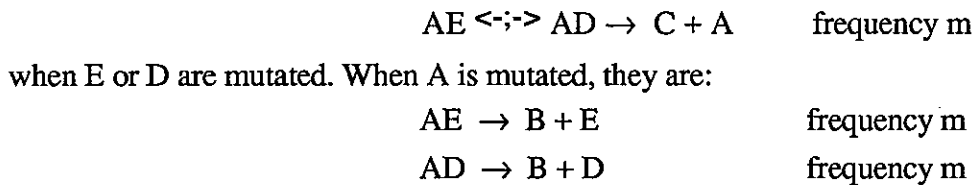
2.2.b Death

Death processes for free species A, B, C, D and E occur with frequency d. But in the case of the death of one of the symbiont of AD or AE, the other one is liberated, with frequency d:



2.2.c Mutations

Mutations involving the free species A, B, C, D and E occur with frequency $m.n$, where n is the number of genes and m the mutation frequency per gene. Only some of these mutations give viable organisms according to the nodes of the phylogenic tree (Figure1). For AD and AE the mutation processes are:



2.2.d Association and dissociation

Association and dissociation processes involve the symbionts:



where k_{l_e} and k_{l_d} are the association (l for linking) constants and k_{s_e} and k_{s_d} are the dissociation (s for separation) constants. Choosing $k_{s_d} > k_{s_e}$ allows E to spend more time inside A than D. This results in a larger time average fitness of E. The basic hypothesis is that the host normally rejects "guests" at a certain rate, but it can somehow appreciate the degree of cooperation of E and reject it less frequently than D. This selective rejection of D might be due to some molecular recognition mechanism, which are known to exist in polypes and sponges, or simply because the increase in the level of nutrient produced by E decreases the rejection rate of E by A.

The only "advantage" of E with respect to D in this model is : $k_{s_e} < k_{s_d}$. We then want to check when, i.e. for what set of parameters, this condition is sufficient to overcome the advantage in fitness of D over E and bring the emergence of mutualism.

3 The Single site model (SSM)

3.1 The differential system

The set of differential equations describing the corresponding population dynamics is:

$$\frac{dA}{dt} = \frac{(\alpha_F.(A+AD) + \alpha.AE).f1}{A+AD+AE+B} - (d+ m.n).A + d_d.AD + d_e.AE + m.(B+AD) + (ks_e.AE) + (ks_d.AD) - kl_e.(A.E) - kl_d.(A.D) \quad (3.3)$$

$$\frac{dAD}{dt} = - 2.(d + m.n).AD + (m.AE) - (ks_d.AD) + (kl_d.(A.D)) \quad (3.4)$$

$$\frac{dAE}{dt} = - 2.(d + m.n).AE + (m.AD) - (ks_e.AE) + (kl_e.(A.E)) \quad (3.5)$$

$$\frac{dB}{dt} = \frac{\beta.B.f1}{A+AD+AE+B} - (d+ m.n).B + m.(A+AE+AD) \quad (3.6)$$

$$\frac{dC}{dt} = \frac{\gamma.C.f2}{C+D+E+AD+AE} - (d+ m.n).C + m.(D+AD) \quad (3.7)$$

$$\frac{dD}{dt} = \frac{(\delta_F.D + \delta.AD).f2}{C+D+E+AD+AE} - (d+ m.n).D + (d.AD) + m.(C+E+AD) + (ks_d.AD) - kl_d.(A.D) \quad (3.8)$$

$$\frac{dE}{dt} = \frac{(\epsilon_F.E + \epsilon.AE).f2}{C+D+E+AD+AE} - (d+ m.n).E + d.AE + m.(D+AE) + ks_e.AE - kl_e.(A.E) \quad (3.9)$$

These equations simply sum the contributions of the processes listed in section 2.2 and set the time variation of each population.

3.2 Simulation results

The initial conditions are such that populations are at equilibrium for the populations B, C, A, D and AD, mutant E being absent:

$$A=0.01 \quad AE=0 \quad AD=0 \quad B=80000 \quad C=20000 \quad D=60000$$

At time $t=0$, E is introduced in the system, with population $E=0.01$

These initial conditions, determined by numerical simulations of the differential system in the absence of E, are those corresponding to the eventual emergence of mutualism, when a true symbiont E is introduced among an ecosystem at equilibrium containing only "egoists" B,C and D,

plus the host species A.

The following set of parameters were used in the simulations, unless otherwise specified.

Parameters	Values
d (frequency of death)	10^{-2}
m (mutation rate)	10^{-5}
n (number of genes)	3
f1 (food shared between A and B)	80
f2 (food shared between C, D and E)	100
<i>Fitness coefficients</i>	
α_F (free organisms A)	8
β (organisms B)	10
γ (organisms C)	8
δ_F (free organisms D)	8
ϵ_F (free organisms E)	7
α (organisms A in AE)	12
δ (organisms D in AD)	12
ϵ (organisms E in AE)	between 8 and 12
<i>Association and dissociation constants</i>	
kl_D (organisms D)	1
kl_E (organisms E)	1
ks_D (organisms D)	between 0.1 and 12
ks_E (organisms E)	0.1

Table 1 : Values of the used parameters in the numerical simulations.

Varying parameters and initial conditions three types of attractors are obtained. Figures 3, 4 and 5 display typical time evolutions of the populations obtained by numerical integration of system (3.3 to 3.9). All attractors are point attractors. The most sensitive parameters, in terms of dynamical behavior, are the difference in fitnesses $\delta - \epsilon$, between species D and E, and the rejection rate of D, ks_D . The resource parameters f1 and f2 simply change the scale of populations. Three time scales, fast, intermediate and slow, are fixed by respectively ks_D , d and m. d/m determines the ratio between dominant and less fit species populations.

For smaller ks_D , i.e., when D spends in A an amount of time comparable to E, AE does not

prevail with respect to AD. D, the "egoist" population benefits from the nutrients it gets from A, and develops faster than E. A does not benefit from D and is not able to grow faster than B. Finally the primeval population B remains at the higher level. A small population of host A is mainly infested by "egoists" D, which gives a small but sufficient advantage to D to overcome C (Figure 3).

In Figure 4, emergence of mutualism is observed for large ks_d , i.e. when D is rapidly expelled from A. A gets support from E and is able to overcome B. The symbiotic organism AE becomes predominant over the primeval populations C and B. The primeval organisms populations are only maintained by the mutations (cf. Figures 1 and 2).

For some values of the fitness parameters a coexistence region is observed for intermediate values of ks_d (Figure 5). In this region, one observes coexistence of the "egoist" and the true symbiont in A, AD and AE. The fitness of A in these couples is comparable to that of B, which coexists with them. A high population of free D is also maintained.

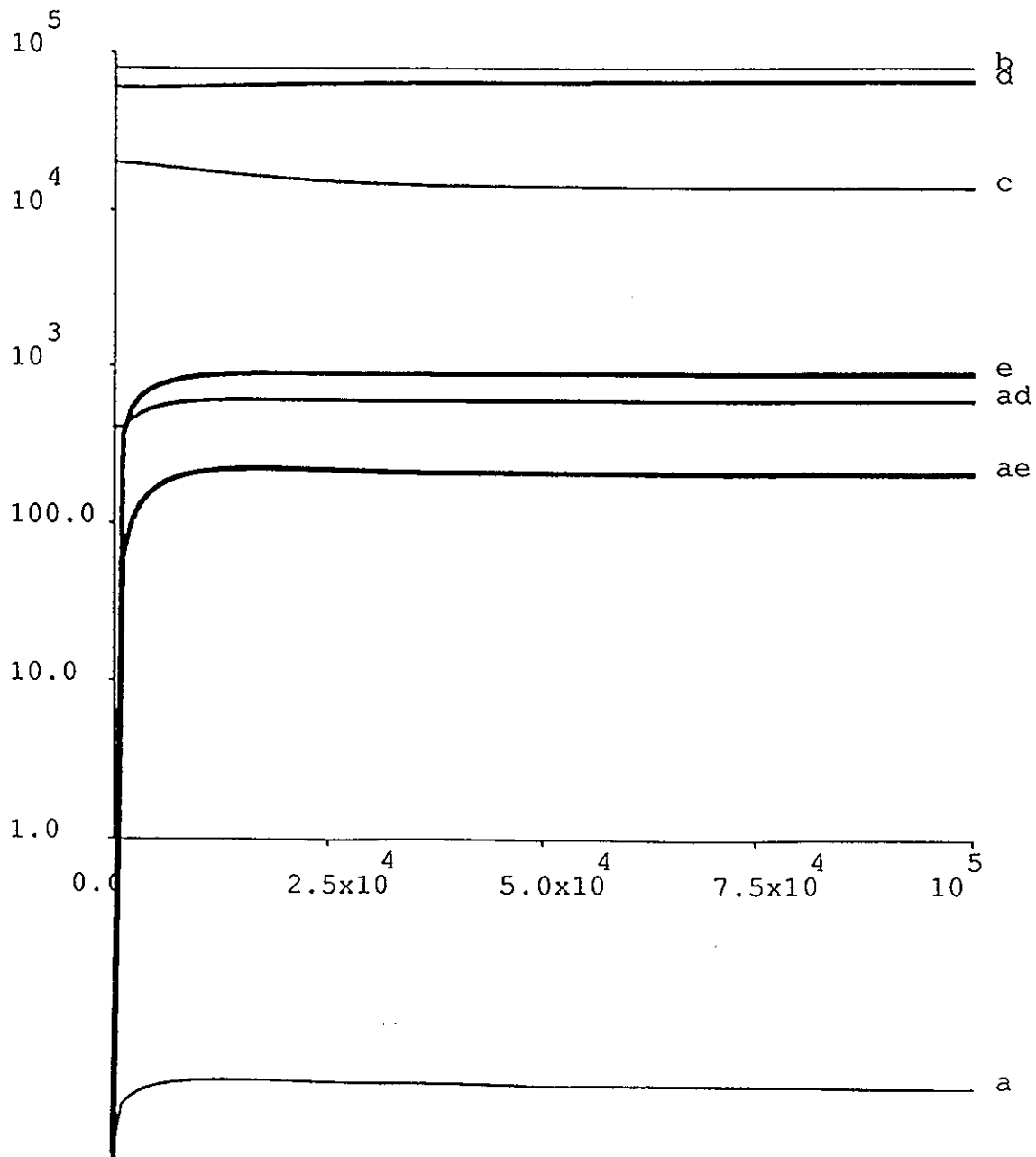


Figure 3 : Prevalence of the primeval species ("egoism attractor") ($ks_d=3$).

Log plot of the populations versus time obtained by computer simulations using GRIND software. Parameters and initial conditions are the same as in table 1, the rejection rate of D is set at a low value $ks_d=3$ and $\epsilon=10$. Mutations from D feed species E. But the insufficient difference in rejection rates between D and E prevents species A and E from growing fast enough to overcome B and D. Although rather little A is available to D, it is enough for D to gain a small selective advantage over C (both otherwise have the same fitness when D is free). D then becomes predominant over C and E.

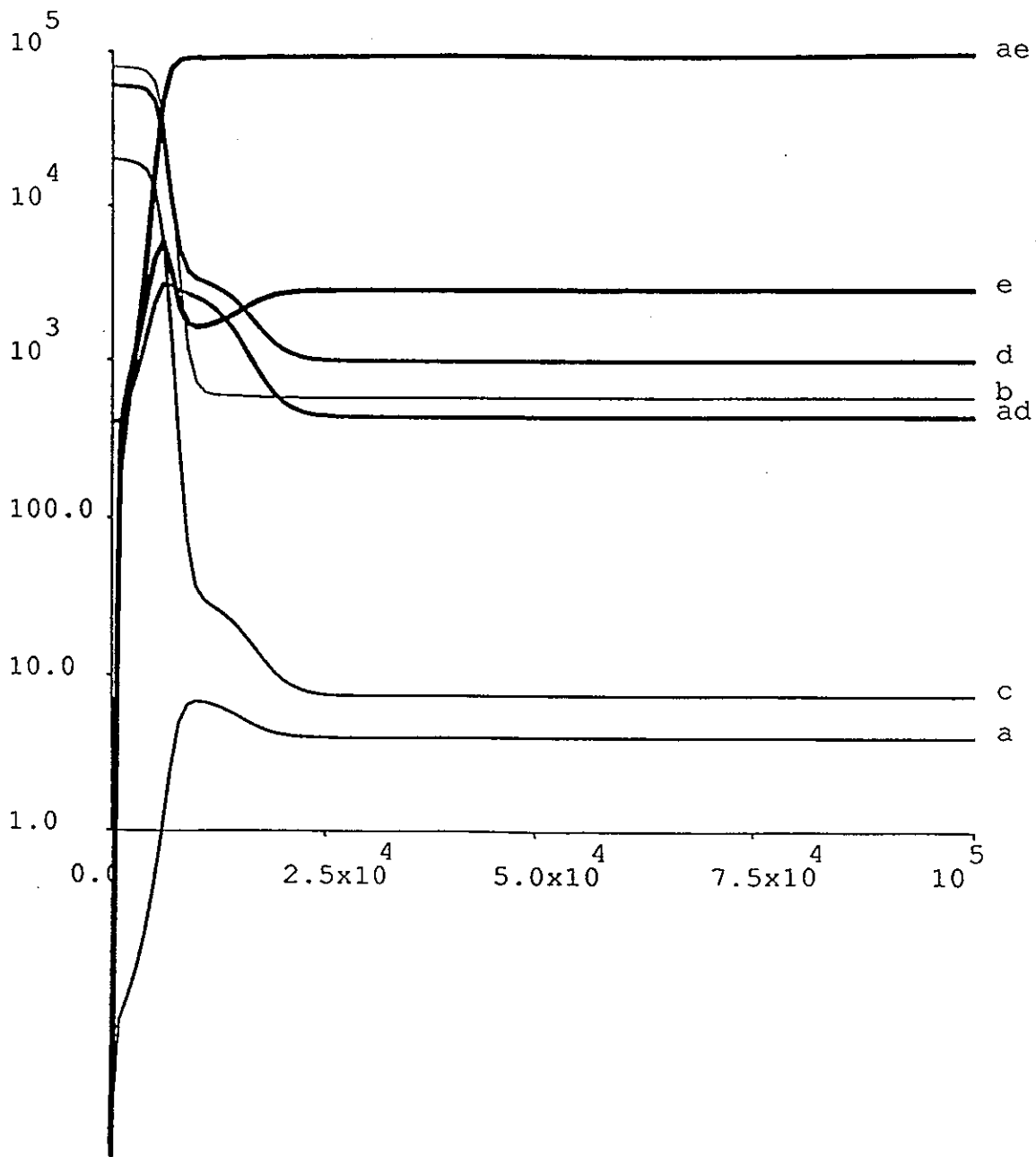


Figure 4 : Emergence of mutualism ($ks_d=9$)

Parameters and initial conditions are the same as in Figure 2, except $ks_d=9$. E can then predominantly invade A, which benefits from its presence. The dominant species is AE, which dominates above all other species in both trees.

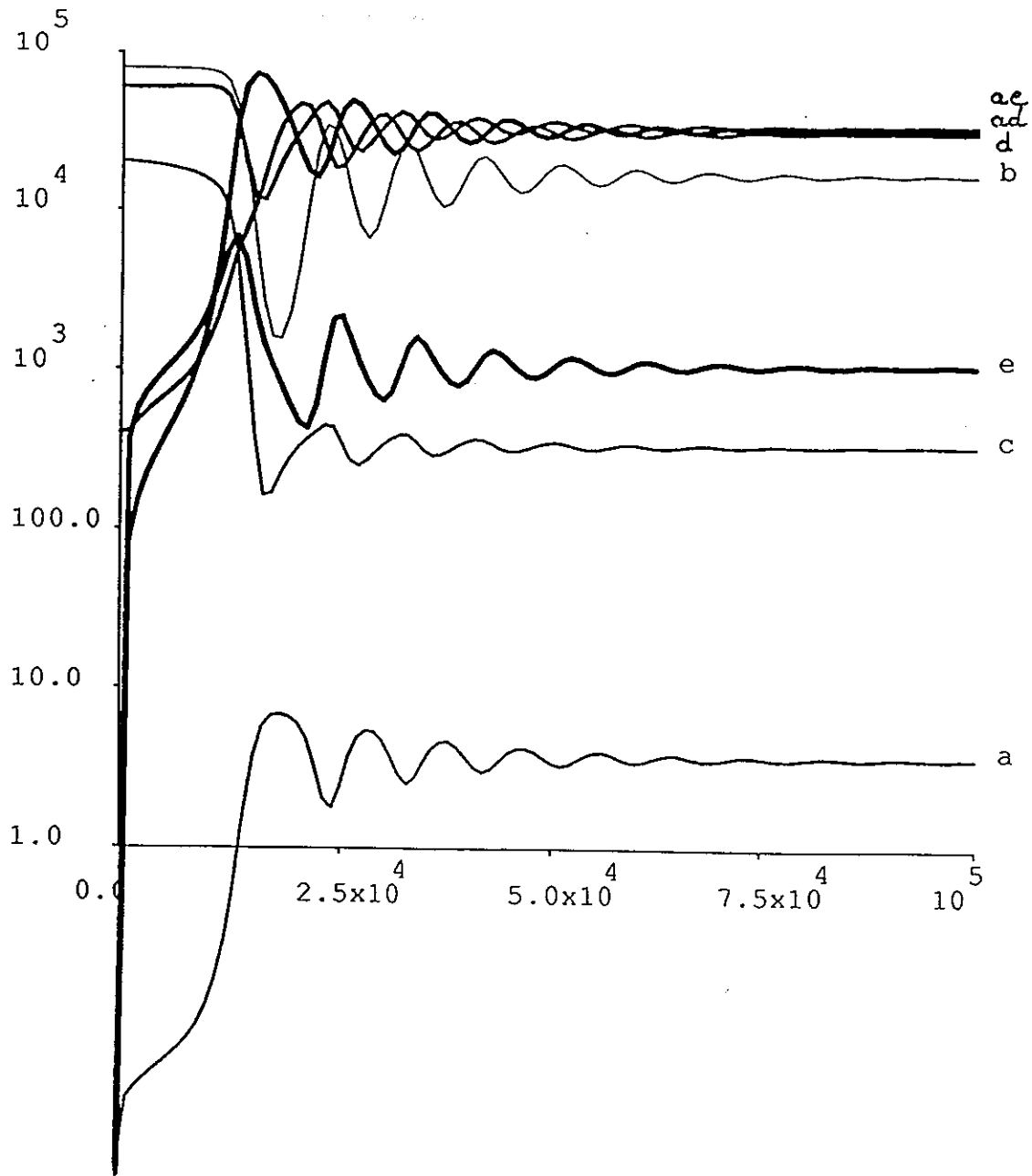


Figure 5 : Coexistence of mutualistic AE, egoist species AD and B ($k_{sd}=7$). The oscillations are an indication that this configuration is only weakly attractive.

The three regimes can be observed on Figure 6 which shows the equilibrium levels of the populations when ks_d increases. The transition value of ks_d between the three dynamics depends upon the difference in fitness between D and E. The "phase diagram" in the plane $(\delta - \epsilon, ks_d)$, Figure 7, shows the different dynamical regimes. These diagrams are complicated because the regime that is reached depends not only from the parameters, but also from the initial conditions. The discontinuities in populations are due to the non-linearity of the equilibrium equations, obtained by setting to zero the time derivatives. Several solutions exist for the same set of parameters, one or two of which are attractors. Dependence from the initial conditions and hysteresis are then observed.

One consequence of this hysteresis is that the conditions for the stability of mutualism when facing the invasion of a new "egoist" is less stringent than the condition for the emergence of mutualism facing the same "egoist" already established. The emergence of mutualism necessitates a transition from a different attractor and is achieved for larger ks_d (or a smaller $\delta - \epsilon$) than the defense against an invading "egoist" which implies a dynamics starting from the mutualistic attractor. In the second case the transition is achieved for a lower ks_d (or a larger $\delta - \epsilon$).

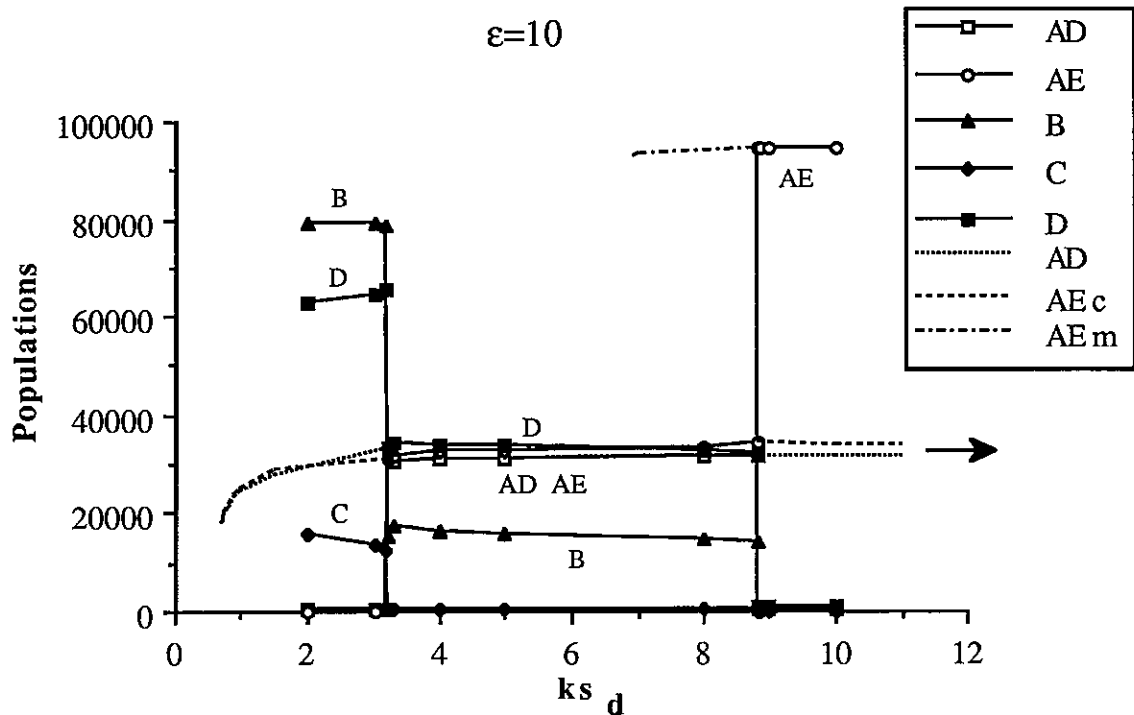


Figure 6: Equilibrium populations measured at time $t=100000$ as a function of ks_d for $\epsilon=10$.

The three regimes, predominance of the ancestors, coexistence and emergence of mutualism are separated by sharp transitions. Populations of A and E, very small, are not represented on the diagram. Continuous lines were obtained from initial conditions in the absence of E species (emergence conditions). The dotted lines are obtained with initial populations of the attractors: AD_c and AE_c corresponding to the coexistence attractor and AE_m to the mutualistic attractor. The arrow indicates that the coexistence attractor exists up to larger values of ks_d .

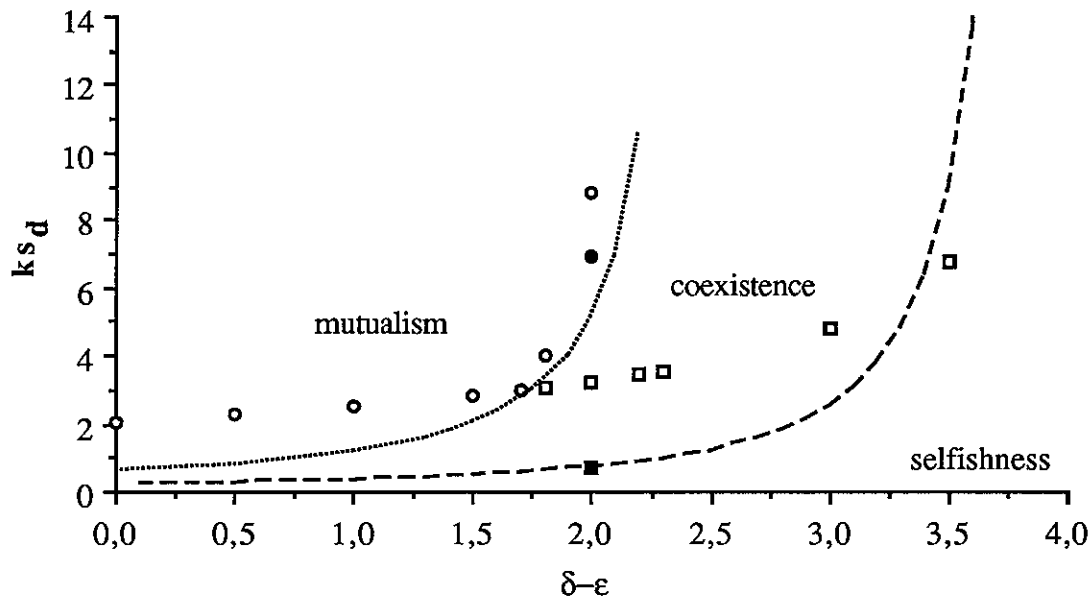


Figure 7 : "Phase diagram" for the emergence of mutualism in the parameter plane ($\delta-\epsilon$, ks_d).

When the difference $\delta-\epsilon$ increases, larger rejection rates of D are necessary for the emergence of mutualism. The two dotted lines were obtained by the slow manifold analysis (see section 4): they limit the domain of existence of the attractors. The open dots, obtained by numerical simulations, mark the transitions between the attractors that are reached from initial conditions in the absence of E (circles signal mutualism/coexistence transition, squares coexistence/non-mutualism). To ensure consistency with theory, they should lie above the transition lines, which is indeed observed, except for one of them. The black dots correspond to the lower limits of the existence domains of the attractors determined by numerical simulations (round for mutualism, square for coexistence); they should lie on the dotted lines.

4 Slow manifold analysis

The slow manifold analysis, detailed in the appendix, allows us to interpret the simulation results and to predict the transitions among the different dynamical regimes. The interaction processes among the different populations have very different time scales: the exchange interaction between free and linked species are very fast with respect to population growth and death terms which are of the same order. Mutations are even slower. One can then suppose that the populations of free and linked organisms follow the ratio at equilibrium given by:

$$(ks_d+2.d).AD = kl_d.A.D \quad (4.1)$$

$$(ks_e+2.d).AE = kl_e.A.E \quad (4.2)$$

After a fast decay towards association equilibria, the populations are describing a slow manifold whose equations are obtained by equating to zero combinations of the time derivatives in which the exchange terms are cancelled (fast exchange). Mutation terms are neglected since they are small with respect to proliferation and death rates:

$$\frac{dA}{dt} + \frac{dAD}{dt} + \frac{dAE}{dt} = \frac{(\alpha_F.(A+AD) + \alpha.AE).f1}{A+AD+AE+B} - d'.(A+AD+AE) = 0 \quad (4.3)$$

$$\frac{dD}{dt} + \frac{dAD}{dt} = \frac{(\delta_F.D + \delta.AD).f2}{C+D+E+AD+AE} - d'.(D+AD) = 0 \quad (4.4)$$

$$\frac{dE}{dt} + \frac{dAE}{dt} = \frac{(\epsilon_F.E + \epsilon.AE).f2}{C+D+E+AD+AE} - d'.(E+AE) = 0 \quad (4.5)$$

$$\frac{dB}{dt} = \frac{\beta.B.f1}{A+AD+AE+B} - d'.B = 0 \quad (4.6)$$

$$\frac{dC}{dt} = \frac{\gamma.C.f2}{C+D+E+AD+AE} - d'.C = 0 \quad (4.7)$$

where $d' = (d + m.n)$. The set of elementary processes that give rise to these equations is reduced to the bold line processes of Figure 2. The first three equations represent the time derivatives of populations of species A, D and E. Here the term species refers to all organisms, whether free or bound. The total population of species A for instance is $A+AD+AE$. In each case the reproduction term involves an "equivalent fitness" for each species. This equivalent fitness is an average over those forms, free or linked, taken by species A, D and E. For A, e.g., it is given by:

$$\frac{(\alpha_F \cdot (A+AD) + \alpha \cdot AE) \cdot f1}{A+AD+AE} \quad (4.8)$$

We might then expect that those species with larger effective fitness, or larger fitnesses as far as B and C are concerned, will become predominant. Since the effective fitnesses depend on the ratio of free to bound organisms, they vary with ks_d , which gives rise to the observed transitions.

We have seven equations with seven unknown, and we should, in principle, be able to compute all variables which we expect a priori to be of the same order of magnitude. In fact this is never observed; if one tries to solve directly the system, some variables have to be set to zero to avoid contradictions. This means that only some of the variables have strictly positive values; they correspond to the dominant populations with larger effective fitness (see Figure 6) and they can be directly computed by solving the slow manifold equations without mutation terms. The other variables are in fact smaller by a factor m/d ; they can be computed from the dominant populations by taking into account the mutation terms that we have previously neglected. The choice of the dominant populations depends on the considered regime. To actually solve the equations, one selects a set of predominant populations, being guided by the simulation results. The self consistency of the choice is checked by solving the equilibrium equations.

There is a limited domain of parameters where a particular solution exists, which gives a necessary condition to reach this attractor. But, unless this solution is unique, knowing this domain does not predict which possible attractor is reached for every initial condition.

In the coexistence domain, for instance, the populations are obtained by only neglecting A and C (A is still kept in the first two equations describing the association/dissociation equilibrium). The boundaries of the coexistence domain occur for those values of ks_d when the solution does not exist anymore. The lower boundary in ks_d is due to the cancellation of the discriminant of a second degree equation, and the upper boundary happens when B goes to 0.

The transition between regimes is also observed when equality among fitnesses of competing species is observed. For instance, the transition between the predominance of AE and the mixed population regime, (when coming from predominance of AE), is obtained for:

$$\frac{\delta_F \cdot D + \delta \cdot AD}{D+AD} = \frac{\epsilon_F \cdot E + \epsilon \cdot AE}{E+AE} \quad (4.9)$$

One of the results of this algebra is that mutualism can be established by compensating the difference in fitness between E and D inside A by a large enough rejection rate for D, provided that the food available to species D and E is in sufficient amount so that E can always saturate A. This condition is written:

$$\epsilon > \frac{f1 \cdot \alpha}{f2} \quad (4.10)$$

Otherwise, if not enough E per A are available, the growth of A cannot be faster than that of B, however large is ks_d .

5 The multi-sites model (MSM)

Binary associations are not the rule and in most cases the host offers a number of sites, p , to the invading species. This is the case, for instance, for corals with algae inside polyps. The straightforward extension of the above model would be to write a system of ordinary differential equations which would include all of species such as AD_rE_s , where r and s vary from 0 to p and indicate the number of D and E organisms inside A. A simpler approach, taking into account the different time scales for association and growth, is to suppose that association and dissociation processes are quasi-instantaneous: after any reproduction, mutation or death event, the induced fluctuations in the rate of occupancy of the host by the "egoists" D and E decay very fast to their equilibrium value. We then do not have to track these fluctuations, which enormously simplifies the model. Apart from the free species, we only have to deal with variables A, AD and AE. AD (resp. AE) is the population of organisms of type D (resp. E) inside A. A is the population of organisms A which are offering p sites per individual to invading organisms D and E. Among these p sites, AE/A sites are occupied by E organisms and AD/A are occupied by D organisms.

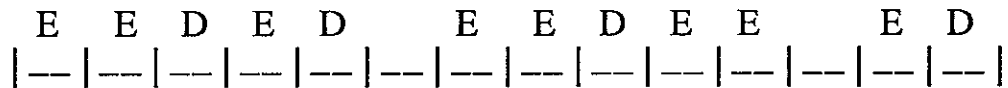


Figure 8 : An organism A with p sites occupied or not by E's or D's.

In other words we are neglecting fluctuations of the occupancy rates of D and E inside the different A organisms. This approximation, equivalent to a mean field approach, is also based on the fast association kinetics. The differential system is then written:

$$\frac{dA}{dt} = \frac{(\alpha_F \cdot (pA - AE) + \alpha \cdot AE) \cdot f1}{p \cdot (A+B)} - d' \cdot A + m \cdot B \quad (5.1)$$

$$\frac{dAD}{dt} = \frac{\delta \cdot AD \cdot f2}{C+D+E+AD+AE} - d' \cdot AD + m \cdot AE - ks_d \cdot AD + kl_d \cdot (p \cdot A - AD - AE) \cdot D \quad (5.2)$$

$$\frac{dAE}{dt} = \frac{\epsilon \cdot AE \cdot f2}{C+D+E+AD+AE} - d' \cdot AE + m \cdot AD - ks_e \cdot AE + kl_e \cdot (p \cdot A - AD - AE) \cdot E \quad (5.2)$$

$$\frac{dB}{dt} = \frac{\beta \cdot B \cdot f1}{A+B} - d' \cdot B + m \cdot A \quad (5.3)$$

$$\frac{dC}{dt} = \frac{\gamma \cdot C \cdot f2}{C+D+E+AD+AE} - d' \cdot C + m \cdot (D+AD) \quad (5.4)$$

$$\frac{dD}{dt} = \frac{\delta_F \cdot D \cdot f2}{C+D+E+AD+AE} - d' \cdot D + m \cdot (C+E+n \cdot AD) + ks_d \cdot AD - kl_d \cdot (p \cdot A - AD - AE) \cdot D \quad (5.5)$$

$$\frac{dE}{dt} = \frac{\epsilon_F \cdot E \cdot f2}{C+D+E+AD+AE} - d' \cdot E + m \cdot (D+n \cdot AE) + ks_e \cdot AE - kl_e \cdot (p \cdot A - AD - AE) \cdot E \quad (5.6)$$

The growth term for A is a linear function of the fraction of sites occupied by E. Mutations in A release infesting organisms, resulting in population changes in AD, AE, D and E. They appear as proportionnal to these populations rather than to A because of the occupancy rates. For instance, when an A organism mutates with frequency m, it frees AE/A E organisms. The corresponding source term in the E differential equation is then:

$$m \cdot A \cdot \frac{AE}{A} = m \cdot AE \quad (5.7)$$

When simulated with p=10 or 100, this model (MSM) gives time evolution of populations which are very similar to those obtained with the single site model (SSM). Populations, are the same in both models for the same set of parameters, provided that the following changes are made:

single site model	multi-site model
f1	p.f1
B	p.B
A	p.A - AD - AE

Except for their short time linking dynamics, both SSM and MSM are pretty much the same, once one realizes that free sites of the A species in the MSM play the same role as free A in the SSM. When the slow manifold analysis is carried out for the multi-site model with the above changes (see the appendix), most equations are identical. This implies that the attractors and their

domain of existence are the same, and that transitions among attractors should occur for similar values of the parameters. We have checked by numerical simulations that the attractors are indeed the same. Even the transition among dynamical regimes are obtained at neighboring couples of parameters in the (k_{sd}, ϵ) plane.

6 Discussion

6.1 Selective protection by the host

Mutualistic species sometimes involve protection in exchange for food, for instance in the case of fish and anemon, or polyp and algae. This protection against predators might be described in our model by changing the death rate inside the host with respect to the death rate outside. By regrouping the first two terms of the differential equation:

$$\frac{dP_i}{dt} = \left(\frac{\alpha \cdot N}{P_t} - d\right) \cdot P_i + \dots$$

One sees that a change in the death rate is equivalent to the same relative change in the fitness coefficient. We have verified by simulations that the decrease of the death rate of E inside A favors the emergence of mutualism.

6.2 Possible scenarios for the emergence of mutualism

6.2.a Differential rates of rejection of selfish organisms

The above computer simulation and mathematical derivations establish the viability of a possible scenario of the emergence of mutualism. The fundamental assumption - existence of a differential rejection rate of the invaders by the host - is based on molecular mechanisms that have been documented in very simple organisms (spongiae, tunicates, coelenterates). A verification of this assumption is the observation of that non symbiotic algae (genus *Chlorella*) injected into hydra are digested, while the true symbiont algae are sequestered in vacuole and migrate towards specialized positions close to digestive cells where they multiply {Begon, Harper and Trunsend, 1986}. We certainly do not claim that the differential rejection rate mechanism we are proposing here is the only way mutualism might establish itself, nor that there is only one possible mechanism. But in the case of coelenterates for instance, we consider it as the most serious candidate.

6.2.b Fast association and dissociation processes

The other important hypothesis of the model concerns the association and dissociation processes, in the multi-sites model: we suppose they are fast compared with the life time of the organisms. This hypothesis applies to coelenterates: symbiotic algae are kept in eggs, and algae and coelenterates are thus never separated. The relevance of the mode of transmission of the symbionts

is stressed in {Maynard Smith, 1991} and {McLaughlin and Zahl, 1966}

6.3 The coexistence region

A somehow unexpected prediction of the model is the existence of the coexistence regime. Coexistence is actually a possibility in those many systems where we cannot figure out the benefits of the association for each individual organism involved. For instance in lichen the benefit for the algae is not obvious {Begon, Harper and Trunsend, 1986}. The coexistence of a variety of intestinal or epidermal flora in insects or mammals could be considered as a possible manifestation of the coexistence regime. On the other hand, since the coexistence situation is less favorable to the host than complete mutualism, we might imagine that further mutations would select host organisms with larger rejection rates of "egoists". One might also view the coexistence of selfish species and mutualists as a source of potential variability which can permit the emergence of new mutualistic species when the biotic and abiotic environments change. The real existence in biological systems of the coexistence regime is certainly an open question worth investigating.

Acknowledgments

Computer simulation used GRIND software {De Boer, 1983}. The Laboratoire de Physique Statistique is associated with the CNRS (URA 1306). This work was started in the Santa Fe Institute which we thank for its hospitality. We thank NATO for partial support (CRG 900998).

We thank Nancy Koppel, Jonathan Roughgarden, Richard Belew and Alan Perelson for helpful discussions.

References

- Begon, M., J. L. Harper and C. R. Trunsend (1986). *Ecology*. Blackwell Scientific Publication.
- Boucher, D. H., S. James and K. H. Keeler (1982). Ecology of Mutualism. *Ann. Rev. Ecol. Syst.* **13**: 315-347.
- De Angelis, D. L., W. M. Post and C. C. Travis (1986). Mutualistic and Competitive Systems. in *Positive Feedback in Natural Systems*. Ed. S. A. Levin. Berlin, Springer Verlag.
- De Boer, R. J. (1983). GRIND: Great Integrator for Differential Equations. Bioinformatics Group, University of Utrecht, The Netherlands.
- Ikegami, T. and K. Kanebo (1990). Computer Symbiosis - Emergence of Symbiotic Behavior Through Evolution. *Physica D 42* : 235-243.
- Lindgren, K. (1992). Evolutionary Phenomena in Simple Dynamics. in *Artificial Life II*. Ed. C. G. Langton, C. Taylor, J. D. Farmer and S. Rasmussen. Addison-Wesley.
- Maynard Smith, J. (1991). A Darwinian View of Symbiosis. in *Symbiosis as a Source of Evolutionary Innovation*. Ed. Margulis. MIT Press.
- McLaughlin, J. A. and P. A. Zahl (1966). Endozoic Algae. in *Symbiosis*. Ed. S. Mark Henry. Academic Press.
- Morse, P. and H. Feshback (1953). *Methods of Theoretical Physics*. McGrawHill.
- Roughgarden, J. (1975). Evolution of Marine Symbiosis - A Simple Cost-Benefit Model. *Ecology*. **56**: 1201-1208.
- Szathmàry, E. and L. Demeter (1987). Group Selection of Early Replicators and the Origin of Life. *Journal of Theoretical Biology*. **128**: 463-486.
- Taylor, C. E., L. Muscatine and D. R. Jefferson (1989). Maintenance and Breakdown of the Hydra-chlorella Symbiosis: a Computer Model. *Proc. R. Soc. Lond.* **238**: 277-289.
- Weisbuch, G. (1984). Un modèle de l'évolution des espèces à trois niveaux, basé sur les propriétés globales des réseaux booléens. *C. R. Acad. Sc. Paris*. **298**(III (14)): 375-378 and in *Complex Systems Dynamics*. Santa Fe Lecture Notes II. Addison-Wesley. pp 161-169.
- Wilson, D. O. (1983). The Effect of Population Structure on the Evolution of Mutualism. *American Naturalist*. **121**: 851-870.

Appendix: The slow manifold analysis

a The single site model (SSM)

The equilibrium equations on the slow manifold are:

$$(ks_d + 2.d).AD = kl_d.A.D \quad A-1$$

$$(ks_e + 2.d).AE = kl_e.A.E \quad A-2$$

$$\frac{(\alpha_F.(A+AD) + \alpha.AE).f1}{A+AD+AE+B} = d'.(A+AD+AE) \quad A-3$$

$$\frac{(\delta_F.D + \delta.AD).f2}{C+D+E+AD+AE} = d'.(D+AD) \quad A-4$$

$$\frac{(\epsilon_F.E + \epsilon.AE).f2}{C+D+E+AD+AE} = d'.(E+AE) \quad A-5$$

$$\frac{\beta.B.f1}{A+AD+AE+B} = d'.B \quad A-6$$

$$\frac{\gamma.C.f2}{C+D+E+AD+AE} = d'.C \quad A-7$$

Dominant species

The five last equations describe the balance between proliferation and death of species A, D, E, B and C. Not all populations can satisfy these equations at the same time. Equations 4 and 7, for instance, when combined, give:

$$\gamma = \frac{\delta_F.D + \delta.AD}{D+AD} \quad A-8$$

which implies:

$$\delta_F < \gamma < \delta \quad A-9$$

in contradiction with expression (3.9), $\delta_F = \gamma$.

In fact, the species with larger equivalent fitness, D, predominates with a large population of order $\delta.f1/d$. The mutation term has to be re-introduced in equation A-7 in order to compute C which is smaller than D+AD by a factor of order m/d . The different dynamical regimes which are observed by computer simulations correspond to different branches of the slow manifold where

different sets of populations are dominant.

The coexistence regime

Let us start with the coexistence regime, where all species except C are dominant. The following change of variable is made:

$$x = \frac{AE}{E+AE}, \quad y = \frac{AD}{D+AD}, \quad z = \frac{E+AE}{D+AD+E+AE}, \quad K = \frac{(ks_d+2.d)}{(ks_e+2.d)} \quad A-10$$

The first two equations being combined in one, the system may be now simplified into:

$$K \frac{y}{1-y} = \frac{x}{1-x} \quad A-11$$

$$\alpha_F.y.(1-z) + \alpha.x.z = \beta.(x.z + y.(1-z)) \quad A-12$$

$$\delta_F.(1-y) + \delta.y = \frac{d.(C+D+E+AD+AE)}{f_2} \quad A-13$$

$$\epsilon_F.(1-x) + \epsilon.x = \frac{d.(C+D+E+AD+AE)}{f_2} \quad A-14$$

$$A + AD + AE + B = \frac{\beta.f_1}{d} \quad A-15$$

Combining equations A-13 and A-14 allows to express y as a function of x, and putting the resulting y in equation A-11 gives a second degree equation in x. Cancelling the discriminant of this equation gives the lower boundary in ks_d for the existence of the coexistence attractor (the corresponding transition line is the lower dotted line on figure 6). An upper boundary in ks_d is obtained by using equation A-12 to compute the actual magnitudes of the populations; it is reached for rather large values of ks_d when B goes to 0. In between, equations A-11 to A-15 allow us to compute the populations of the dominant species and C is computed from the original equation 3.5 with mutations.

The mutualistic regime

In the other regimes, the set of dominant species is even smaller. We could of course systematically try all possible combinations of species to find out whether they satisfy equations A-11 to A-15, but the easiest way is to make use of simulation results to select them.

In the mutualistic regime, the dominant population is AE. It is computed from equation A-12 by neglecting all other populations:

$$AE = \frac{\alpha.f_1}{d} \quad A-16$$

Using this result in equation A-15 gives a second degree equation for x, the ratio of bound E to total E (as defined above in A-10):

$$(\varepsilon - \varepsilon_F). x^2 + \varepsilon_F . x - \frac{\alpha.f_1}{f_2} = 0 \quad A-17$$

The condition $x < 1$ gives the following inequality between the fitnesses and the available foods:

$$\varepsilon.f_2 > \alpha.f_1 \quad A-18$$

When both sides are divided by d , they represent the maximum population size of species E and A. Inequality A-18 is then interpreted as the possibility for E to eventually bind to all available A. Otherwise, however large ks_d is with respect to ks_e , the effective fitness of A is not larger than β , and mutualism cannot be established.

The mutualistic regime is observed as long as the effective fitness of species E is larger than that of D. A transition to coexistence occurs when both fitnesses become equal. The transition line of figure 6 (the upper dotted line), which limits the domain of existence of mutualism, is obtained when the values of the populations computed from equations A-11 to A-15 are replaced in the equality:

$$\frac{\delta_F.D + \delta.AD}{D+AD} = \frac{\varepsilon_F.E + \varepsilon.AE}{E+AE} \quad A-19$$

b The multi-sites model (MSM)

The last four equations A-4 to A-7 of the simplified dynamical system are the same for both models (single site and multi-site, resp. SSM and MSM). For the MSM the first three equations are written:

$$ks_e.AE = kl_e. (p.A-AD-AE).E \quad A-20$$

$$ks_d.AD = kl_d. (p.A-AD-AE).D \quad A-21$$

$$\frac{(\alpha_F.(pA-AE) + \alpha.AE).f_1}{p.(A+B)} = d.A \quad A-22$$

The two first equations, once combined, give the same equation, A-11, in x and y as in the SSM. The third equations A-22 and A-3 are equivalent in both models when one notices that the free A population of the SSM corresponds to the number of A free sites, $(pA-AD-AE)/p$, of the MSM. f_1 in the SSM then corresponds to f_1/p in the MSM.

Since the dynamical equations on the slow manifold are the same we expect the attractors and their domains of existence to be equivalent which is verified by the computer simulations. The only difference concerns the evolution towards the attractors. At short time scales of order $(ks_d)^{-1}$ some differences are indeed observed, consistent with the approximations made in the MSM on the binding dynamics. Otherwise, at larger time scales the dynamics are identical, and even the transition lines from any given initial condition are very close. Their relative distance in the (ε, ks_d) plane is never more than 1 per cent of the parameters.