Environmental heterogeneity and the evolution of mutualism

Gergely Boza, István Scheuring*

Department of Plant Taxonomy and Ecology, Research Group of Ecology and Theoretical Biology, Hungarian Academy of Sciences and Eötvös University, Pázmány P. sétány 1/c, Budapest 1117, Hungary

Received 10 March 2004; received in revised form 20 July 2004; accepted 22 July 2004

Abstract

Mutualistic interactions among species are common in most ecosystems. Despite the ubiquitous presence of mutualism, its evolutionary origin and stability has remained enigmatic. Here, we study this problem by generalizing Doebeli and Knowlton’s spatial model for the evolution of mutualism [Doebeli, M., Knowlton, N., 1998. The evolution of interspecific mutualism. Proc. Natl. Acad. Sci. USA 95, 8676–8680]. We show that spatial environmental heterogeneity facilitates the evolution of mutualistic interactions under a wide range of plausible conditions. Our model exhibits complex spatial patch dynamics, with environmental heterogeneity reducing among patch migration and leading to increased phenotypic diversity. Moreover, under such environmental heterogeneity, we also observe that both mutualism and exploitative interactions can be permanently present in the same population.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Environmental heterogeneity; Evolution of mutualism; Intraspecific cooperation; Iterated prisoner’s dilemma; Variable investment

1. Introduction

Despite the prevalence of mutually beneficial interactions among species in many ecological communities, most previous studies have concentrated on antagonistic species interactions (Herre et al., 1999; Bronstein, 1994, 2001). Studies of mutualism are under-represented even in recent ecological and evolutionary work (see Law, 1985; Frank, 1994, 1996; Noé and Hammerstein, 1995; Doebeli and Knowlton, 1998; Ferriere et al., 2002; Hoeksema and Schwartz, 2003). Mutualism, defined as an interaction in which two or more species benefit each other, ranges from specific obligate symbiotic associations to facultative interactions among free-living species. While the important role of mutualism in the functioning of communities is widely acknowledged (Boucher et al., 1982), its origin and evolutionary stability has remained a challenging problem (Herre et al., 1999). To provide an evolutionary explanation for the widespread presence of mutualism, we need to answer two basic questions: (1) How can mutually beneficial interactions originate from non-mutualistic interactions? (2) What mechanisms prevent the spread of defectors (i.e., individuals that enjoy a fitness
benefit from the mutualistic act while not reciprocating it to the partner)?

Given the vast array of different kinds of mutualistic interactions, it is hardly surprising that most previous work has studied specific interactions by using various modeling approaches, including evolutionary game theory, population dynamics, or biological market models. For example, the classical game theory framework for studying cooperation among unrelated individuals is the Prisoner’s Dilemma Game (Trivers, 1971). In this game, partners can choose between a defective (non-mutualistic or cheater) and a cooperative (mutualistic) strategy. While cooperating with a cooperative partner yields a high fitness payoff to the performing individual, defecting against the cooperator is even more beneficial. Consequently, the defector is exploited by the defector and receives a low fitness payoff (so-called ‘suckers’ payoff). In contrast to this situation, the fitness payoff is slightly larger if both individuals defect. Thus, unless cooperation provides some extra advantage, defection is the only evolutionarily stable strategy (ESS) in indefinitely large populations (Maynard Smith, 1982), (Table 1). So, according to the Prisoner’s Dilemma Game, neither the origin nor the evolutionary stability of mutualism seems to be possible.

In their model, Doebeli and Knowlton (1998) introduced an interspecific Prisoner’s Dilemma Game where the payoffs can change according to the investments made by the partners. The level of investment $I_h$ by a host individual involves a cost $C(I_h)$ to the host and gives a benefit $B(I_h)$ to its symbiont partner. Thus, if the symbiont partner invests $I_s$, then the payoff for the host is $S(I_h, I_s) = B(I_h) - C(I_h)$. Similarly, the payoff of the symbiont is $S(I_s, I_h) = B(I_h) - C(I_h)$. (The authors refer to the mutualistic partners as ‘host’ and ‘symbiont’, and we will use the same terminology herein.) The payoffs resulting from the mutualistic interactions are summed over a series of Prisoner’s Dilemma Games with variable investments, where in each game the investment of a partner depends on its payoff received in the previous round of interaction. In this model, the success of a strategy within a given species depends on the fitness gains or losses accumulated over the repeated series of mutualistic interactions. Numerical simulations of this model revealed that mutualism cannot evolve if the partners live in “well-mixed” populations lacking spatial structure (Doebeli and Knowlton, 1998). The condition of a “well-mixed” environment means that the probability mutualistic interactions among different phenotypes is equal to the product of their relative frequencies. Doebeli and Knowlton (1998) considered a spatial version of this model, too, where host and symbiont populations were located on two separate lattices. Mutualistic interaction can occur between individuals located in the corresponding grid point on the different lattices and competition is limited to neighbouring conspecifics (Fig. 1). They have shown that mutualism can emerge from an initially neutral situation in this spatial model (Doebeli and Knowlton, 1998). The evolution of mutualism is particularly likely if dispersal is limited (Yamamura et al., 2004) or absent (Doebeli and Knowlton, 1998) and if there is some stochasticity in the competition or selection process. Nevertheless, Doebeli and Knowlton (1998) have emphasized that if the difference between the cost and benefit of mutualism is not great enough (although $B(I) - C(I) > 0$ valid for $0 < I < I_{crit}$), then

<table>
<thead>
<tr>
<th>Cooperate</th>
<th>Defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R$</td>
<td>$S$</td>
</tr>
<tr>
<td>$T$</td>
<td>$P$</td>
</tr>
</tbody>
</table>

The two strategies determine four outcomes of the interactions. The reward ($R$) of cooperation, the temptation ($T$) and punishment ($P$) of defection and the sucker’s payoff ($S$) are ordered as $T > R > P > S$. It is worth to play defect, independently what the opponent chooses ($T > R$ and $P > S$), thus defection is the only evolutionary stable strategy.
Mutualism could not evolve. Similarly, differences in the host and symbiont generation times have a negative impact on the evolution of mutualism. Mutualism only evolves if the competition–selection process is sufficiently stochastic, that is, there is a chance of the less fit phenotypes to survive to the next generation. Thus, in summary, the spatial model by Doebeli and Knowlton (1998) suggests that evolution of mutualism is possible, however, it is sensitive to the precise details of the ecological and demographical circumstances, and is less likely to evolve in environments where mutualism is more costly.

It is important to emphasize that Doebeli and Knowlton (1998) considered the habitats to be homogeneous and constant in time. Although various studies have investigated the role of spatial structure for the evolution of cooperation (Doebeli and Knowlton, 1998; Killingback et al., 1999; Yamamura et al., 2004), the effects of spatial heterogeneity on the emergence of mutualistic behavior have only rarely been studied.

Here we generalize the Doebeli and Knowlton model by including spatial heterogeneity. Interestingly, our model shows that, while the interspecific cooperation can be locally costly at some sites, mutualism evolves easily for most sites in space. Thus, although our results show that environmental heterogeneity generally facilitates the evolution of mutualism, we also find that a range of different strategies can co-occur, from stable mutualism with increased phenotype diversity to exploitation.

2. The model

Following Doebeli and Knowlton (1998) paper, we model the dynamics of the interacting species on two separate, two-dimensional lattice spaces. Individuals start the interaction with an initial investment, named initial offer \( a_h \) and \( a_s \). Thus, the host receives the payoff \( B(a_h) - C(a_h) \), and symbiont gets the payoff \( B(a_s) - C(a_s) \). Individuals modify the level of investment in the following rounds proportional to the payoff received from their partner previously, that is \( I'_h = a_h + b_h S(I_h, I_h), \) \( I'_s = a_s + b_s S(I_h, I_s) \), where \( b_h \) and \( b_s \) are the reward rates that determine how sensitively the focal pair of individuals change their investment in the following round \((I'_h, I'_s)\) by the effect of their opponent investment. In our simulations, negative investments were set to zero. This variable investment strategy resembles the Tit for Tat strategy in the Iterated Prisoner Dilemma Game (e.g., Axelrod and Hamilton, 1981), in that every individual starts with cooperation in the first round \((a_h, a_h > 0)\) and continues to cooperate or defect depending on the strategy that the partner played in the previous round.

The cost function increases linearly with the investment (Doebeli and Knowlton, 1998), e.g., \( C(I) = C_0 I \) (see Holland et al., 2002; Killingback and Doebeli, 2002 for other approaches). The benefit is a saturating function of \( I \), that is \( B(I) = B_0 [1 - \exp(-B_1 I)] \). After a fixed number of iterations (15), the payoffs of individuals are summed up, which yields the fitness of the individuals in host and symbiont populations, respectively. These fitness values determine the success of individuals during intraspecific competition. In this deterministic model, every individual on a given lattice point is compared to all its nearest neighbours (Fig. 1), and the one with the highest fitness replaces the local individual in the next generation. Thus, individuals with the locally most successful strategy (or phenotype) replace the less successful ones. The boundary conditions are periodic (i.e., lattices are folded into tori). Since the initial offer and the reward rate determine the strategy of the individual, the \((a \) and \( b)\) pairs specify the phenotypes of individuals. The phenotypes are inherited during the replication process in the competition step, but they can also mutate with a certain probability at each point of the lattice. The mutants were drawn from a bivariate normal distribution with a mean gained from the phenotype values of the mutated individuals.

Doebeli and Knowlton (1998) considered two ecological situations. In the first case, the generation times of the host and the symbiont were the same (symmetrical generation time (SG)). In the second case, they examined a situation with asymmetric generation times (AG), with shorter generation time for the symbiont. (The terms ‘host’ or ‘symbiont’ are arbitrary in the first case, since the interspecific mutualist interaction is absolutely symmetrical).

To generate spatial heterogeneity, we assumed that the parameter \( C_0 \), and hence the cost, varies from place to place. Consequently, investments are more costly on some points on the grid, while other local habitats are more favorable for mutualism. For the sake of
simplicity, we assume that there are only two types of points, points with high cost \(C_{\text{high}}\) and points with low cost \(C_{\text{low}}\). Let us further assume that these local habitats are distributed randomly and uncorrelated on both lattices. In the simulations, we varied the number of points with high cost and the cost difference between grid points being high and low cost. We started the simulations from extremely small, but positive levels of \(a\) and \(b\), and studied whether mutualism evolves from this neutral initial state.

3. Results

To investigate the effect of spatial heterogeneity on the evolution of mutualism, we used the deterministic version of the model. We chose a parameter set for which mutualism is present in a widely fluctuating manner in the SG model (Fig. 2(a)), but for which mutualism disappears in the AG model (Fig. 2(b)). We then introduced local grid points at which the cost of mutualism is even higher than in the corresponding homogeneous case.

If the ratio of high-cost patches is low or moderate (roughly 1–40%), the payoff curves are separated and stabilized in the SG case (Fig. 2(c)). However, there are occasionally abrupt changes in the payoff functions at the beginning of the simulations, and these transient phases can be very long (see Fig. 2(c)). Yet, after such a transient phase, one species (denoted as host) will generally exhibit a permanently higher fitness than the other species (symbiont), while the symbiont still has a fitness larger than zero. Consequently, the average benefit from the interaction is positive for both species. The evolutionary dynamics in our simulations break this symmetry, and one of the species consequently receives a much higher benefit from the mutualism than the other. Since the model is symmetrical and selection is deterministic, the initial conditions and stochasticity in the mutation process determine which species becomes the host (has higher fitness) and which the symbiont (has lower fitness), respectively.

In some patches (mostly on the high-cost grid points), payoffs may be close to zero, or even negative. This is shown in Fig. 3, which shows the same results as Fig. 2(c), but now with the average fitness values only depicted for the high-cost grid points. The fitness of the symbiont is negative or close to zero on average in these grid points, except for some mutant phenotypes that can remove transiently.

For the case of AG, mutualism also disappears after some time if the ratio of high-cost habitats is low, but this process is slower than in the homogeneous model (Fig. 2(b and d)). In the case of a highly heterogeneous environment (e.g., when cost is high in about half of the grid points), mutualistic interactions emerge for both SG and AG. Populations show smaller fitness fluctuations than in the homogeneous environment (Fig. 2(a, b, e and f)). In the SG model, the average fitness values of host and symbiont species are roughly the same (Fig. 2(e)). In the AG model, the host has a two–five times higher fitness than the symbiont (Fig. 2(f)); nevertheless, mutualism is still permanent (Fig. 2(b)). When the percentage of high-cost grid points is high (roughly 60–90%), mutualism shows an increasingly fluctuating dynamics again (Fig. 2(g and h)), but this system is still more stable than in the homogeneous case. Naturally, if the cost is high for all local grid points, then mutualism disappears in both models. The fitness differences observed in host and symbiont populations are largely due to differences in the mutualistic strategies. This can be demonstrated by measuring the evolution of \(a\) and \(b\) parameters in Fig. 2(c). It can be seen that species with higher fitness exhibits higher values of \(b\) and lower values of \(a\) than the species with lower fitness (Fig. 4).

We measured the average fitness (which is proportional to the level of mutualism) plotted against the proportion of high-cost grid points and the difference in the cost on high-cost and low-cost grid points in a systematic simulation experiment (Fig. 5). It can be seen from this simulation that mutualism is the most pronounced in the SG model if the environmental heterogeneity is the highest, but still permanent at any combination of studied parameter values (Fig. 5(a and c)). However, mutualism appears abruptly at a critical ratio of high-cost grid points in the AG model (Fig. 5(b and d)). Surprisingly, increasing the cost that an individual has to pay for the mutualistic investment in the high-cost patches has a little effect on the results. The outcome of the interactions is the same even if the cost is enhanced by magnitudes (see the \(y\) axes on Fig. 5).

Introducing heterogeneity in \(B_0\) (which measures the level of benefit of mutualistic interactions when investment is low), or considering the environmental
Fig. 2. The time evolution of average fitness in the SG and AG model. There are 100 symbiont generations within one host generation in the AG model. Host denoted by the black (red in online figure) and symbiont by the grey (green in online figure) line. (a) The homogeneous SG model and (b) the homogeneous AG model. (c and d) Heterogeneous models with low ratio of high-cost grid points ((c) SG and (d) AG model). The ratio of high-cost grid points is 20% that are distributed randomly in both lattices. (e and f) Heterogeneous models with medium level of high-cost grid points ((e) SG and (f) AG model). The ratio of randomly distributed high-cost grid points is 50%. (g and h) Heterogeneous SG models with high ratio of high-cost grid points ((g) SG and (h) AG model). The ratio of high-cost grid points is 80%. The other parameters were $B_0 = 7$, $B_1 = 2.5$, $C_{\text{low}} = 0.1$ and $C_{\text{high}} = 0.4$. The lattice size was $23 \times 23$, the mutation probability per site was 0.02 with variance 5% of the actual mean. Parameters $a$ and $b$ were set initially to 0.005.
heterogeneity present in correlated manner in the lattices, leads to a very similar conclusion, therefore, it is not shown here. Naturally, if $C_{\text{low}}$ is set to be higher (e.g., $C_{\text{low}} = 0.8$), then mutualism disappears in the homogeneous SG model too, but spatial heterogeneity raises it from the death similarly to that was shown in Fig. 2.

4. Discussion and conclusions

Doebeli and Knowlton (1998) showed that competition and fitness stochasticity is an effective stabilizing force for mutualism in a spatially explicit model setting. Here we generalize Doebeli and Knowlton (1998) results by showing that spatial heterogeneity can effectively stabilize mutually beneficial interactions in the deterministic version of Doebeli and Knowlton’s model. Importantly, this stabilization of mutualism even occurs when the heterogeneous environment is on average less profitable for the populations than in the homogeneous equivalent. It is also noteworthy that only a single heterogeneous grid point is sufficient for causing changes in the outcomes of the simulations and can, under special circumstances, protect mutualism from

Fig. 3. The average fitness values of hosts and symbionts in the sites where hosts and symbionts are located in high-cost grid points. Every parameter is the same as in Fig. 2(c).
vanishing (results not shown). The common feature of the mechanisms promoting mutualism in our model is an increase in phenotypic diversity. But why is such an increase in diversity important for the evolution of mutualism?

In spatial models for the evolution and maintenance of mutualism, different clusters of evolving phenotypes are formed, resulting from local interactions (Doebeli and Knowlton, 1998; Killingback et al., 1999). In our model, we observed the emergence of two qualitatively different pairs of clusters. For the first cluster, the initial offer is relatively high \(a_{\text{high}}\) and the reward rate low \(b_{\text{low}}\) in the host population, while the initial offer is low \(a_{\text{low}}\) and the reward rate is high \(b_{\text{high}}\) in the symbiont population. For the second type of cluster, the situation is the opposite (e.g., hosts have \(a_{\text{low}}\) and \(b_{\text{high}}\), while symbionts have \(a_{\text{high}}\) and \(b_{\text{low}}\); see Fig. 6). If the host (symbiont) individual has a higher reward rate than its symbiotic partner, then it is favorable for the host to make a low initial offer. In contrast, receiving a high initial offer is especially favorable for the symbiont (host) (Doebeli and Knowlton, 1998). The stochasticity inherent in the mutation process directs the \(a\) and \(b\) phenotype-pairs onto one of these two asymmetrical evolutionary paths. The locally most successful phenotype pairs spread and form phenotypically homogeneous pairs of clusters. This is because asymmetries in the initial

![Fig. 4. The time evolution of (a) initial offer \(a\) and (b) reward rate \(b\) in case of Fig. 2(c). Host indicated by the black and symbiont by the grey line as above.](image-url)
Fig. 5. The average fitness of host and symbiont in function of ratio of high-cost and low-cost grid points ($C_{\text{high}}/C_{\text{low}}$) and the percentage of high-cost sites. (a) The average fitness of the host in the SG and (b) in the AG model. (c) The average fitness of the symbiont in the SG and (d) in the AG model. The other parameters are the same as in Fig. 2.
offers and reward rates cause asymmetry in fitness. Individuals with the \((a_{low} \text{ and } b_{high})\) phenotype have a higher fitness than the individuals having a \((a_{high} \text{ and } b_{low})\) phenotype. The clusters spread, and after a while they come in contact with each other. Due to the asymmetries discussed above, the host and symbiont individuals migrate from the qualitatively different clusters to the grid points being on the boundaries of the clusters (Fig. 6). However, these ‘hybrid’ host–symbiont pairs generally have lower fitnesses compared to their neighbouring pairs; therefore, they cannot spread into the opposite cluster pairs. Consequently, the border of the clusters remains unchanged. While this scenario is very common, it may not be the general rule. For example, if the phenotypic asymmetries are sufficiently large between connecting clusters, then a hybrid phenotype pair can invade one of the cluster pairs, starting a new evolutionary path. These complex interactions can maintain a dynamical equilibrium among the transient cluster pairs, leading to the stabilization of mutualism. Stochasticity in the selection process enlarges the possibilities of transient states, and hence it improves the evolutionary maintenance of mutualism. This is only a qualitative verbal description of the spatial patch dynamics; there are currently no quantitative methods available to characterize cluster dynamics (Hauert, 2000).

The effects of environmental heterogeneity are similar to those of stochasticity. In contrast to the homogeneous case, where one phenotype can expand to form a large cluster, there exist at least four different phenotypes in the heterogeneous environment; one phenotype is successful on low-cost grid points, while another phenotype is most successful on high-cost grid points; the remaining two phenotypes are most successful on low-cost/high-cost and high-cost/low-cost grid point pairs, respectively. Furthermore, the high-cost grid points can act as ecological barriers to mixing among phenotypes. This is due to the following: first, because of the high cost of mutualism, the average fitness is generally lower at these sites as compared to low-cost grid points nearby. Thus, even a well-adapted phenotype can rarely win the local competition. Second, an invader phenotype from a neighbouring low-cost grid point cannot simply step over such a high-cost grid point, since the invader’s phenotype is generally not well-adapted to the new environment. Consequently, such a phenotype has a relatively low fitness as compared to its neighbours, and will thus be replaced by another phenotype with higher fitness in the following generation. Generally, this new invader receives a low fitness again in the high-cost grid point and will be replaced by a following invader. If the high-cost grid points take a shape of connected sets, they can act as barriers among the different clusters and can maintain boundaries for a very long times (Fig. 7). After the initial transient, mosaics of smaller clusters emerge (first column of Fig. 7), and there is a transient phase with the formation of larger patches with a correlated distribution on the host and symbiont lattice (middle column of Fig. 7). To demonstrate the barrier effect of connected sets of high-cost grid point, we replace a bar with high-cost grid points in the heterogeneous lattice. It can be seen that this ‘high-cost bar’ can act as a barrier for the different clusters over a long period of evolutionary time (Fig. 7, bottom). Consequently, in the heterogeneous case, this barrier maintains clusters of different mutualistic cluster pairs, while in the homogeneous model, mutualism disappears rapidly (left column of Fig. 7). If the ratio of high-cost grid points is increased, there is a great chance that low-cost grid points are surrounded by high-cost grid points (and vice versa). Isolation is even more effective in these clusters, so that they become essentially refuges for the phenotypical diversity. If the ratio of high-cost grid points is high (40–99%), the emergence of spatial refuges around the few low-cost grid points protects mutualism from declining in both the SG and the AG model. This result sharply contrasts

![Fig. 6. Schematic picture of invasions on the border of qualitatively different cluster pairs. Higher fitness is indicated by darker grey color from where the phenotypes arrive to the border.](image)
with intuition, since one would expect mutualism to be less favored in a bad environment. Interestingly, mutualism disappears abruptly at about 60% of the low-cost grid points in the AG model (Fig. 5). The percolation point is close to this value in the rectangular lattices. Consequently, we suspect that because all the low-cost grid points become connected at and above this threshold, then the high-cost grid points behave as less efficient barriers. Thus, the heterogeneous environment becomes similar to the...
homogeneous one. The less mutualistic and locally successful invaders can spread along the lattice freely, so mutualism is stopped at this point in the AG model.

In this paper, we show that environmental heterogeneity can maintain mutualism. Interestingly, the host with a slower evolution wins more from the mutualism than its symbiotic partner in the AG case. This result is explained by the so-called Red King effect recently. The symbiont with the faster life history offers below each other to receive benefits from the host, so they lose the ‘bargaining game’ and cannot maintain high fitness values (Bergstrom and Lachmann, 2003).

Numerous interspecific interactions can change from mutualism to extreme exploitation (e.g., parasitism) in changing environments (Bronstein, 1994, 2001). An interaction between the same two species can vary from exploitation to mutualism within the same community (Bronstein, 1994, 2001; Machado et al., 1996; Pellmyr and Leebens-Mack, 1999). Our results emphasize that although environmental heterogeneity may locally cause exploitation (or cheating), such heterogeneity is a crucial driving force for the stable maintenance of mutualism.

Acknowledgements

We thank C. Pál and T. Flatt for their helpful comments. I.S. is supported by the Hungarian National Science Foundation (OTKA, grant no. T037726).

References