Eur. Phys. J. B **10**, 181–186 (1999)

THE EUROPEAN PHYSICAL JOURNAL B

EDP Sciences

Model of populations colonizing a new habitat

I. Mróz[1] and A. Pękalski[2,a]

[1] Faculty of Chemistry, University of Wrocław, F. Joliot-Curie 14, 50-383 Wrocław, Poland
[2] Institute of Theoretical Physics, University of Wrocław, pl. M. Borna 9, 50-204 Wrocław, Poland

Received 11 December 1998 and Received in final form 17 December 1998

Abstract. We propose a model suitable for investigating the conditions under which a population, adapted to a given environment, may colonize a new neighbouring, spatially separated, habitat. We find out how similar the two regions must be for such a colonization to succeed and what will be the spatial distribution of genetic pools of the two populations after the process. It is found that between the two populations each adapted to the different region, a hybrid zone, characterized by increased heterozygosity, may be formed. The dependence of the shape of the hybrid zone on the external (environmental) parameters is determined. When the differences between the two regions are significant, the populations are separated by a depopulated zone. We show that the conditions in the colonized habitat influence the genetic pool of the population living in the first region. Computer simulations based on the standard Monte Carlo technique are used.

PACS. 87.10.+e General biophysics – 02.70.Lq Monte-Carlo and statistical methods – 07.05.Tp Computer modeling and simulation

1 Introduction

Speciation phenomena providing a good example of differentiation in nature have been intensively investigated experimentally and theoretically, by biologists (see *e.g.* [1,2]) and recently also by physicists (*e.g.* [3–5]). Many possible mechanisms of speciation have been described (*e.g.* [6]) but a satisfactory theory has not been presented yet. In some cases speciation occurs when a population of living organisms colonize an empty ecological niche and as a result the genetic pool of the population is changed.

When speciation is not complete, the individuals belonging to two populations can mate and produce hybrids that increase genetic and phenotypic variability of the population. In natural conditions hybrids often live in a relatively narrow area between the areas occupied by the two populations. The existence of hybrid zones corresponds to "a state of equilibrium" between natural selection against hybrids and a gene flow. The parameters of the zones, like their shape or density of hybrid populations were analysed in detail [7], but many problems still remain to be investigated. The role of changing or spatially heterogenous environment, was investigated *e.g.* by [8–10].

In this paper we propose a model of a colonization process that leads to separation of the initial population into two populations characterized by different genetic pools. We study the effect of migration of a part of the population into a different, spatially separated environment and possible separation of the population into two. In particular we investigate conditions under which the population living in a given habitat can persist and when it is able to colonize successfully the new habitat. We consider how various ways of migration influence the colonization process. Migration processes are often described using methods based on diffusion (*e.g.* [11]). Ecologists usually distinguish between two ways of migration [12]: "diffusive" migration means that individuals move through the environment gradually, step by step while "jump dispersal" refers to the situation when individuals can rapidly find themselves far from their previous habitat. Our approach allows us to investigate both situations.

The populations in our model generally stay in contact with each other during the colonization process. We investigate how the ecological factors may influence the possibilities of creation of a hybrid zone and its shape. Since natural hybrids are heterozygous at many loci (see *e.g.*, [13]) we identify the hybrid zone with the area of increased heterozygosity.

The present approach differs from that we took in our earlier paper [10] in considering here in much more detail the genetic structure of the populations and investigating possibilities of creation of hybrid zones.

We assume that alleles are either dominant or recessive and do not mutate. An influence of particular alleles on the fitness of a given individual is not fixed and depends on the ecological factors (features of the environments in which the population exist). It should be stressed that the evolution (dynamics) of our model is controlled mainly

[a] e-mail: `apekal@ift.uni.wroc.pl`

by the requirements of the two environments. Frequencies of all alleles occurring in the populations are moreover controlled.

2 The model

In our model a habitat is represented by a $L \times L$ square lattice. Each lattice site may be either empty or contain an individual belonging to a population. The total initial number of individuals is N ($< L \times L$). An individual is assumed to be diploid, *i.e.* its genotype has at each site two alleles, and the individual is characterized by: its location j, on the lattice, its age W_j and its genome containing P genes that code P phenotypic features. Denoting recessive and dominant alleles by a and A, respectively, a genotype $\mathbf{G}_j$ of each individual can be written as a P-dimensional vector of components equal either 0 (aa), 1 (aA) or 2 (AA):

$$\begin{aligned} \mathbf{G}_j &= \{g_j^1, g_j^2, ..., g_j^P\}, \\ g_j^\alpha &= 0, 1, 2, \\ \alpha &= 1, ..., P. \end{aligned} \tag{1}$$

In the computations we took $P = 10$. In our model we assume that a phenotype, $\mathbf{F}_j$, of an individual j, follows directly from its genotype, *i.e.*

$$\begin{aligned} g_j^\alpha = 0 &\Rightarrow f_j^\alpha = 0 \\ g_j^\alpha = 1, 2 &\Rightarrow f_j^\alpha = 1. \end{aligned} \tag{2}$$

Hence the phenotype may be written as a vector of length P, composed of 0's and 1's:

$$\mathbf{F}_j = \{f_j^1, f_j^2, ..., f_j^P\}, \quad f_j^\alpha = 0, 1, \tag{3}$$

where 0 and 1 mean that the αth phenotypic feature of the individual manifests itself as coded by the a or A allele, respectively.

Each individual has its own age counter, set at the birth at zero. An individual moves, via Brownian motion, through the lattice and when encountering another individual in one its nearest neighbouring (NN) sites, they may produce offspring. From the parents each offspring takes independently one allele for every locus, hence the features are inherited independently. Although it is clearly a simplification, it is a little more realistic than commonly assumed inheritance at the level of phenotypes (see *e.g.*, [14]).

There is one specific phenotype described by the bit string $\mathbf{F}_0$ that gives the best chance of survival and is henceforth denoted as the *model phenotype*. Such model phenotypes have been already used in earlier works [15,16,10]. The model phenotype is fixed and does not change in time. In this way the natural selection is realized in our model. Since we analyze possibilities of colonization of a new habitat, we divide the lattice into two parts. Each part has its own model phenotype constant in time and the two model phenotypes can differ from each other.

Natural selection is realized in our model by linking the survival probability of an individual with its adaptation to the environment it lives in. The adaptation, ad_j, is measured by the degree of agreement between the individual's phenotype and the one preferred by the environment (model phenotype) and agrees with the observation that fitness depends on many genes [13].

The algorithm governing the dynamics of the evolution of the model is the following:

1. At time t a site j is chosen at random. If it is occupied by an individual having a phenotype $\mathbf{F}_j$, we check its survival chances by comparing its phenotype with the model one $\mathbf{F}_0$

$$p_j(t) = P^{-1} \sum_{\alpha=1}^{P} F_j^\alpha(t) \,\mathsf{AND}\, F_0^\alpha, \tag{4}$$

   i.e., p_j is the fraction of bits which are equal in a bit-by-bit position of the two bitstrings $\mathbf{F}_j$ and $\mathbf{F}_0$.
   A random number $r \in [0, 1]$ is chosen from a uniform distribution. If $r > p_j(t)$, the individual is removed from the system. If not, the individual has a chance to mate. The probability of survival $p_j(t)$ is connected with viability of an individual and influences its fitness [17].
2. To mate the individual has to move to one, randomly chosen, NN site. If none is available, no mating is possible and the process returns to 1.
3. Accomplishing the move, checking is made for an individual (potential partner) in one of the NN sites after the move. If a partner is found, its survival probability, from eq.(4), is calculated. If either there is no partner or it did not survive, the process returns to 1.
4. A given number, (q), of offspring is located inside a square $LG \times LG$ centered at the position of the first parent. Genotypes of the offspring are constructed independently. For each locus two alleles, each selected randomly from one of the parents, are chosen at random. Then offspring's phenotypes are calculated following the rule (2).

The above steps, repeated as many times as there are individuals, comprise one Monte-Carlo Step (MCS) per particle.

After each MCS the age of all individuals is increased by one. If the age is equal to the assumed maximum value, the individual dies, *i.e.* it is removed from the system. The algorithm is repeated for the predefined number of MCS. Then the procedure is repeated with new, random, initial configurations. The results are accumulated and finally averaged over the number of trials.

With this algorithm in hand we divided initially the lattice into two equal parts of dimensions $L/2 \times L$, separated at the beginning by an impenetrable barrier. In one part, characterized by model phenotype $\mathbf{F}_{\mathrm{I}}$ a population with random positions and random values of its genotypes is located. The second part, characterized by a phenotype $\mathbf{F}_{\mathrm{II}}$, is empty. After some time, usually of the order of 5 MCS, the barrier is lifted and the population

may invade the empty niche. The invasion of the empty space is realized by displacements of individuals as well as by locating there offspring. In each part of the system the survival probability of an individual is calculated from equation (4) with the appropriate model phenotype **F**. Therefore, an individual adapted very well to one environment may be poorly adapted to the other one. Of course the adaptation of the individual is changed if it migrates to a different habitat.

When the model phenotypes are identical or very similar, the population has no problems in colonizing a new habitat and the genetic pools of the populations in the two regions do not differ significantly.

To get more detailed information on the colonization process we divided the dimension perpendicular to the barrier into thin (two lattice spacings) strips S_i, parallel to the barrier, and the following characteristics of the populations have been calculated:

1. the concentration, c_i, of individuals (density of the population) in the ith strip

$$c_i(t) = \frac{1}{n}\sum_{k=1}^{n} x_k(t), \quad k \in S_i, \tag{5}$$

where ($i = 1$, ..., number of strips), $x_i = 0, 1$ if the site i is empty or occupied and n is the number of sites in a strip;

2. average adaptation, ad_i, of an individual in the ith strip

$$ad_i(t) = \frac{1}{n}\sum_{k=1}^{n} p_k(t), \quad k \in S_i, \tag{6}$$

3. average number of heterozygotes, ht_k, per individual, in a strip

$$ht_i(t) = \frac{1}{nP}\sum_{k=1}^{n}\sum_{\alpha=1}^{P} g_k^{\alpha}(t)\left(2 - g_k^{\alpha}(t)\right), \quad k \in S_i, \tag{7}$$

4. average rate of recessive alleles, alR_i, per individual and number of alleles in a genotype, in a strip

$$alR_i(t) = \frac{1}{2nP}\sum_{k=1}^{n}\sum_{\alpha=1}^{P}\left(2 - g_k^{\alpha}(t)\right), \quad k \in S_i, \tag{8}$$

5. the last recorded quantity is the percentage of populations which survived the simulation.

It is straightforward to calculate also other quantities, like *e.g.* average age either in each strip or for the whole population. Our calculations show however that these data do not bring new information and therefore will not be reproduced here.

To reduce the number of free parameters of the model we have decided on setting the maximum age of an individual to 5 MCS. This means that for a typical run of 150 MCS there are more than 50 generations. In our model parents do not die after giving birth, hence the generations overlap. We have also set the maximum number of offspring to $q = 4$. Additional simulations, not reported here, convinced us that changing maximum age or q do not lead to drastic changes of the results, except if $q \leq 2$ when the populations die out quite fast. The initial density of the populations was $c(0) = 0.3$. Taking smaller values resulted in more populations being extinct, hence on working with worse statistics. The populations which did persist, quite fast reached in the first region a steady concentration $c \approx 0.8$, independent of the initial one. This is due to the fact that initial distribution of the genotypes was random, hence there was much "room" for improvement, what the population did during the first MCS. Larger initial concentrations also lead to the same final one. Typically the simulations were carried out on a 80×100 lattice, *i.e.* two parts 40×100 each, till 150 MCS and the results were averaged over 500 runs. This means that the data for each strip were averaged over some 80 000 entries. Hard boundary conditions were used, *i.e.* an individual could not move farther than the first or last row of the lattice. If a move towards a "forbidden" direction was selected, the individual stayed put. Longer simulations do not change the results significantly, indicating that our system has reached equilibrium.

The important parameters of the model are: model phenotypes, $\mathbf{F}_{\rm I}, \mathbf{F}_{\rm II}$, of the two regions and linear dimensions (LG) of the square area where the offspring are located.

3 Results

The choice of the model phenotypes is of course quite arbitrary. We have decided on working mostly with cases when at least one of the **F**'s has equal number of 0's and 1's.

As the reference case we have chosen for the first environment the following model phenotype

$$\mathbf{F}_{\rm I} = (0000011111) \equiv (0^5 1^5) \tag{9}$$

For the second region it was one out of the sequence

$$\begin{aligned}
&\mathbf{F}^1_{\rm II} = (1^5 0^5) \quad \text{(orthogonal to the first one)}\\
&\mathbf{F}^2_{\rm II} = (1^3 0^7) \quad \text{(2 common zeros with the first one)}\\
&\mathbf{F}^3_{\rm II} = (1^2 0^8) \quad \text{(3 common zeros)}\\
&\mathbf{F}^4_{\rm II} = (1^1 0^9) \quad \text{(4 common zeros).}
\end{aligned} \tag{10}$$

Orthogonal means that to each zero (one) in a given locus in one phenotype corresponds a one (zero) in the second phenotype. If to a zero (one) corresponds zero (one), it will be a common zero (one). In our model an environment more difficult to survive corresponds to a model phenotype with more 0's, since a 1 in a phenotype may come either from a 1 or 2 in the genotype, but a 0 in the phenotype must come from a 0 in the genotype. The first region extends to the middle of the system, *i.e.* till the 21st strip. The dependences of the population density, mean adaptation, percentage of heterozygotes, recessive allele, on the distance along the X-axis (perpendicular to

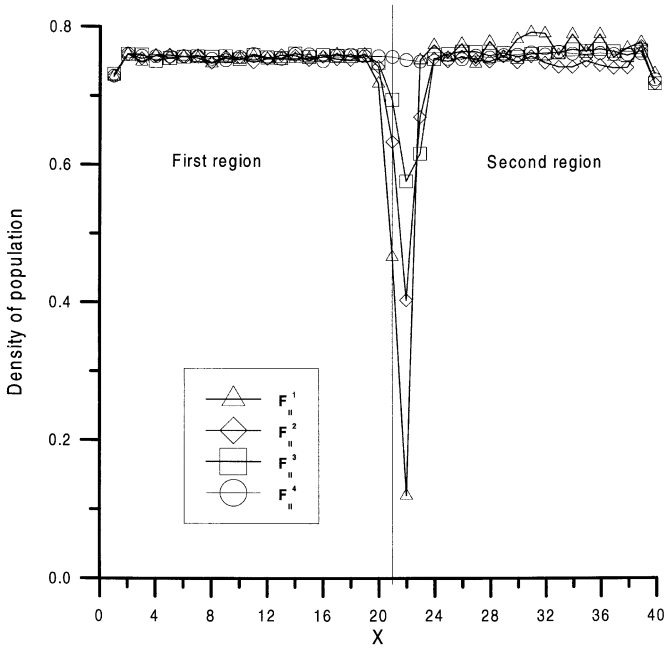


Fig. 1. Density of population *versus* distance (X-axis) perpendicular to the initial barrier located at the position 21 (see the text). The phenotype in the first region is $\mathbf{F}_{\rm I} = (0^5 1^5)$ and in the second region it is given by equation (10). LG = 4.

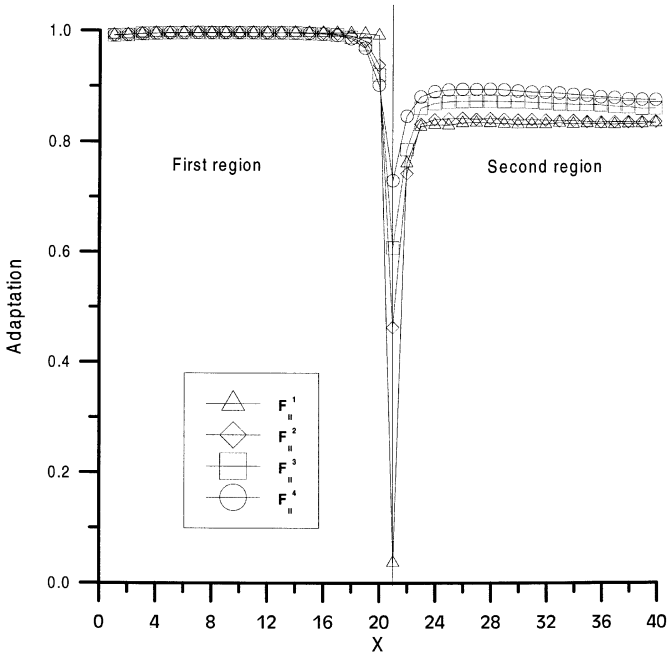


Fig. 2. Adaptation *versus* distance. Notation as in Figure 1.

the initial barrier), are shown in Figures 1-4 for $LG = 4$. In Figure 5 the percentage of surviving populations is presented.

It should be noticed that when the two model phenotypes are orthogonal, the two regions are separated by an interval almost completely depopulated (see Fig. 1), meaning that the two populations are truly separated. Any individual which happens to find itself on the border between the two regions (21st and 22nd strip) is so poorly adapted (see Fig. 2) that its chances of survival, hence mating, are indeed very poor. The width of this "depletion zone" widens with the number of common loci. In a more difficult environment a smaller number of populations persist, out of the ones that were tried, but those which do, reach relatively high level of adaptation, although clearly lower than, nearly perfect, adaptation in the first region (Fig. 2). The number of persisting populations depends on the new environment and it increases with the size of LG. The size of the area, (LG), over which the offspring can be located is, to some extent, measuring the inbreeding of the population.

The data in Figure 3 show an interesting feature. With increasing similarity between the two regions, the amount of heterozygotes in the second region decreases, but it increases in the first one and the maximum becomes more pronounced as well as shifted towards the inside of the first region. This means that when a population has a chance to colonize an adjacent habitat, the conditions there influence also the genetic pool, (the rate of heterozygosity, hence allelic frequency), of the first region. The rate is therefore determined not only by the conditions in a given environment but also by those in the neighbouring ones.

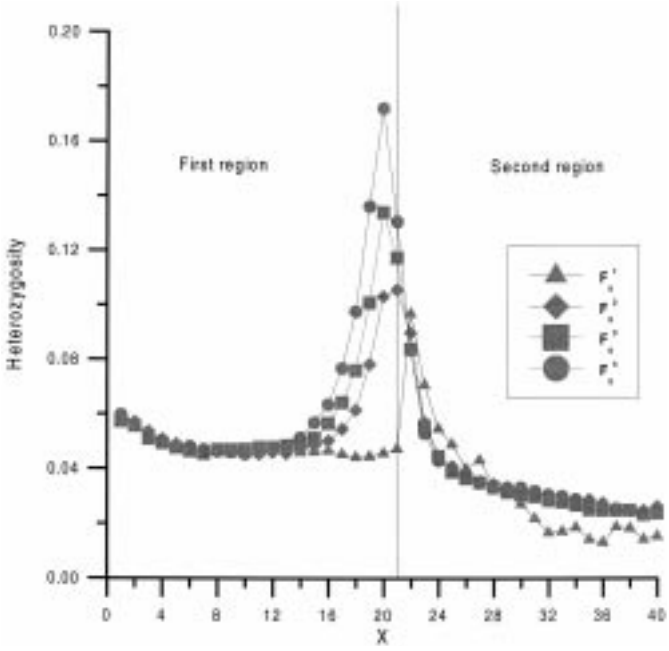
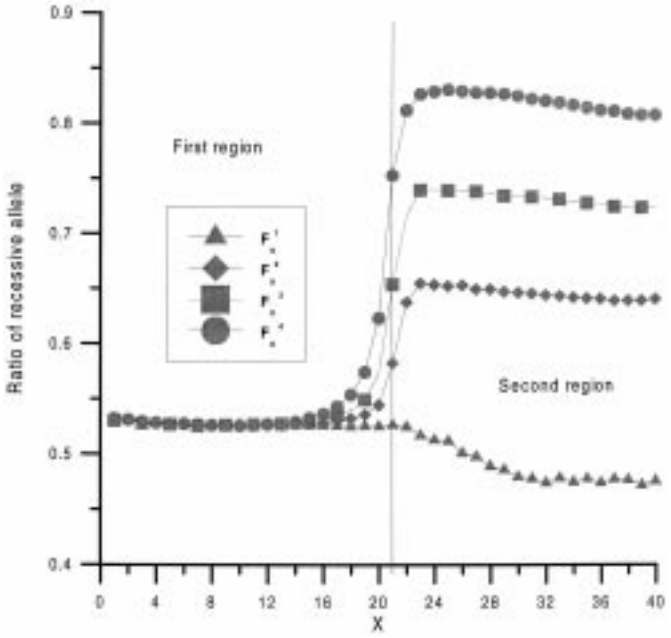
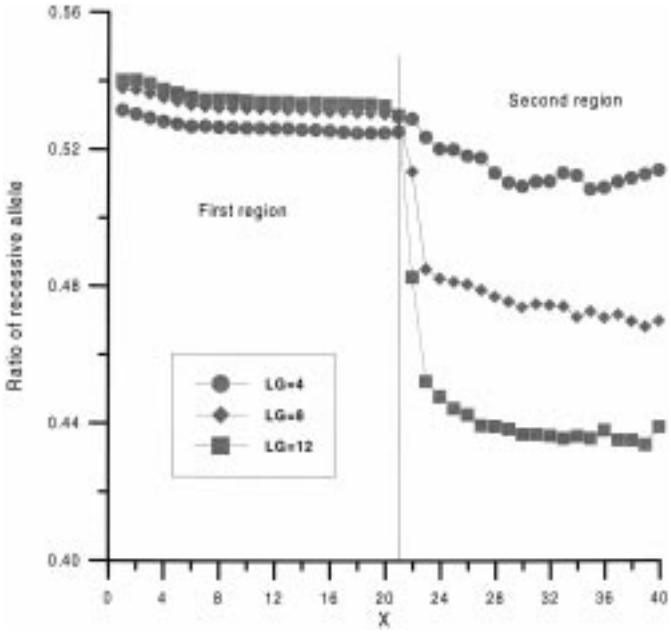


Fig. 3. Heterozygosity *versus* distance.

Such an effect, which does not depend on LG, is not observed for adaptation nor for concentration and the rate of the recessive allele. The width of the zone over which increased rate of heterozygotes is observed is increasing with the size of LG (decreasing inbreeding). The existence of the zone depends also on the model phenotypes; for the orthogonal ones $\mathbf{F}_{\rm I} = (0^5 1^5)$ and $\mathbf{F}_{\rm II} = (1^5 0^5)$ such a zone exists, while for the case $\mathbf{F}_{\rm I} = (1^{10})$ and $\mathbf{F}_{\rm II} = (0^{10})$ the zone disappears. If one considers $\mathbf{F}_{\rm I} = (0^{10})$ and $\mathbf{F}_{\rm II} = (1^{10})$ the zone appears again.



(a)



(b)

Fig. 4. Relative number of the recessive allele *versus* distance.

Figure 4 shows the distribution of the recessive allele. It is clear that individuals conforming to a model phenotype with more zeros have a larger percentage of the recessive allele. There is however one feature worth mentioning - the case of orthogonal phenotypes with equal number of zeros. The population which colonized the ecological niche has nevertheless a lower rate of the allele. This may be explained by the fact that the first region was populated by

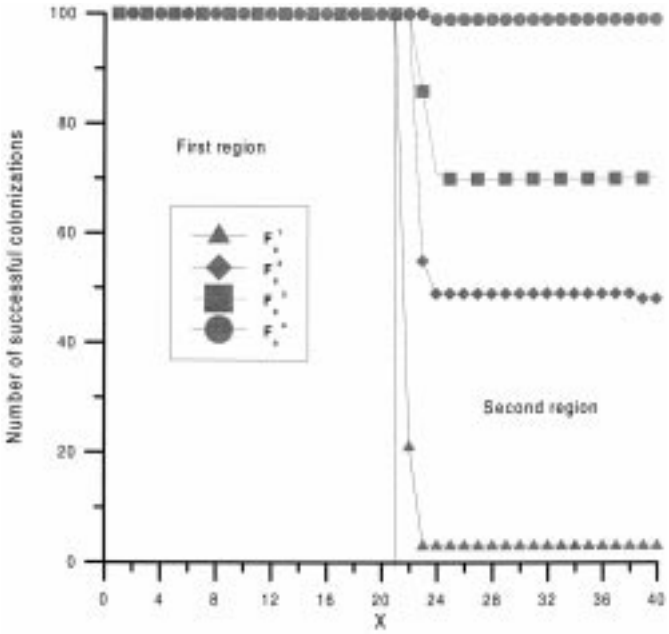


Fig. 5. Number of populations surviving after 100 MCS *versus* distance. Data from 100 runs. $LG = 4$.

a set of individuals with random genotypes, whereas the second region by an already shaped population.

It is clearly seen from Figure 5 that the degree of similarity between the two regions is crucial for the colonization to succeed. This result corresponds to the analysis presented by Pease *et al.* [8], in which the slow rate of environmental changes allow the investigated population to persist, otherwise it goes extinct. The role of the size of environmental changes in the extinction risk was stressed by Bürger and Lynch [9].

We have also investigated the cases when the initial environment was characterized by the phenotype $\mathbf{F}^1 = (0^{10})$ or by $\mathbf{F}^2 = (1^{10})$ and the final one by an appropriate orthogonal one. We have found that very few colonization attempts are successful in the first case (only 1 out of 500) but the populations which survived in the second region reached quite high mean adaptation and density, comparable to that in the first region.

In the second case, *i.e.* starting with "easier" environment, lead to adaptation and density in the difficult (*i.e.* characterized by more zeros in the model phenotype) region nearly equal to the initial ones. Also here the percentage of successful populations was below 10%.

4 Final remarks

The obtained results stress that minimum initial population densities are important for persistence of the populations in one habitat. The probability of successful colonization is connected with initial mean adaptation of the colonizing population. Increasing dissimilarity between the two regions makes the second region harder for initial colonization to occur but does not strongly affect the level

of mean adaptation for the population which successfully colonized the new habitat.

Stabilization of density of a population, which has been obtained using our model, is typical for closed populations in which fertility or mortality of individuals depend on their concentration [18]. According to our model mortality of newborn offspring (there may be no place for them on the lattice) and probability of creating offspring follow this rule. The rules of migration defined in our model are important if one takes into account possibilities of colonizing a new habitat by offspring. If the area where the offspring are located (LG) is small, they can migrate to the other habitat more gradually than if LG is large. The first situation is similar to the "diffusive" way of colonization while the other one corresponds rather to "jump dispersal".

In our model an individual migrates through the lattice randomly and it may stay at its maternal environment as well as move to the other one. We have therefore no habitat choice. This is clearly a simplification, rarely found in nature, only for some invertebrata [19]. We have decided to restrict our model to this simple approach in order to get a clear picture of the considered mechanisms. Habitat selection has been considered in theoretical models and seems important for population conservation [8], differentiation and speciation phenomena [20].

We have shown that obtaining a distinct hybrid zone, defined by a large degree of heterozygosity is not a rule. The shape of the zone is determined by the used model phenotypes, but not only by their similarity; it is also important to what extent the heterozygotes are tolerated in the first region.

Preliminary results have been presented at a national topical conference - Zwierzyniec, September 1998.

References

1. J.A. Endler, *Natural selection in the wild* (Princeton University Press, Princeton, New Jersey, 1986).
2. R.H. Tamarin, *Principles of genetics*, Fourth edition (W.C. Brown Publishers, Dubuque Iowa, 1993).
3. B. Derrida, P. Higgs, J. Phys. A: Math. General A **24**, L985 (1991).
4. A.O. Sousa, S. Moss de Oliveira, Int. J. Mod. Phys. C (1999).
5. K. Malarz, T. Tieggemann, Int. J. Mod. Phys. C **9**, 481 (1998).
6. *Speciation and its consequences*, edited by D. Otte, J.A. Endler (Sinauer, Sunderland, Mass 1989).
7. E. Mayr, *Populations, species and evolution* (Harvard University Press, Cambridge, Massachusetts, 1970).
8. C.M. Pease, R. Lande, J.J. Bull, Ecology **70**, 1657 (1989).
9. R. Bürger, M. Lynch, Evolution **49**, 151 (1995).
10. I. Mróz, A. Pękalski, K. Sznajd-Weron, Phys. Rev. Lett. **76**, 3025 (1996).
11. T. Nagylaki, Proc. Natl. Acad. Sci. USA **75**, 423 (1978).
12. E.C. Pielou, *Biogeography* (J. Wiley, New York, 1979).
13. J.M. Szymura, N.H. Barton, Evolution **45**, 237 (1991).
14. M. Doebeli, Evolution **50**, 535 (1996).
15. M. Eigen, Naturwissenschaften **58**, 465 (1971).
16. J. Thoms, P. Donhaue, N. Jan, J. Phys. I France **5**, 935 (1995); J. Thoms, P. Donhaue, D. Hunter, N. Jan, J. Phys. I France **5**, 1689 (1995).
17. T. Nagylaki, *Introduction to theoreticalpopulation genetics*, (Springer-Verlag, Berlin, Heidelberg 1992).
18. Ch.J. Krebs, *Ecology* (Harper-Collins College Publishers, 1994).
19. G. Fraenkel, D.L. Gunn, *The orientation of animals: Kineses, taxes and compass reaction* (Oxford University Press, Oxford, 1940).
20. T. Kawecki, Proc. R. Soc. Lond. B **263**, 1515 (1996).