A Computational Model of Speciation in Non-Uniform Environments Without Physical Barriers.

Ezequiel A. Di Paolo
School of Cognitive and Computing Sciences
University of Sussex, Brighton BN1 9QH, U.K.
Email: `ezequiel@cogs.susx.ac.uk`

Abstract

A simple computational model of parapatric speciation is introduced and analysed within the framework of the Dobzhansky-Muller model of accumulations of incompatible mutations. It is shown that speciation can occur without the need of physical barriers in environments showing non-uniform local conditions. A population of ancestor agents is placed in an artificial environment, their fitnesses depending on the balance of energy intakes and costs caused by movement and heat dissipation during their lifetime. Phenotypic features such as thermal resistivity are encoded genetically and the agents are able to interbreed locally. It is observed that the accumulation of mutations cause the formation of neighbouring populations showing localized adaptation but still able to interbreed with the ancestor populations. It is argued that this is a case of polytypic species or "subspecies". However this does not necessarily happen between populations separated by one or more subspecies, and speciation can be said to occur in this case, which can be viewed as the emergence of allopatricity without any previous physical barriers.

1 Introduction

The formation of a new biological species, as a phenomenon, arises from the interplay of many different factors acting in as many different levels and timescales, and as a result it proves to be a highly complex process to explain satisfactorily as a whole. Traditional models study the effect of geographical factors on populations' gene pools and explain under what circumstances isolating mechanisms might evolve between related subpopulations. However only a few verbal arguments address the qualitative problems imposed by the effects of population dynamics, ecological factors, and localized interactions between the individuals and the environment. This fact suggests that explanations focusing only on certain dimensions of the process will in general fail to grasp completely its complex dynamics.

The purpose of the computational model presented here is precisely to try to capture some of the emergent features of speciation that are missed for this reason by purely analytical/stochastic models and verbal arguments. These usually try to describe the speciation process at a population level. By means of computer simulations many features can be modelled at a lower level (in this case at the level of the individual) and effects at the collective level can be expected as a result.

Computational models of this kind have many advantages, not the least important of which are their flexibility and their capability of expanding the range of questions that can be explored; questions that are very hard to answer by more traditional analytical means. Such questions include the effect of environmental variations in space and time and the effect of the distribution of resources.

However it is unreasonable to expect purely computational models to replace other kinds of explanations due to the difficulty (or even impossibility) of translating all the relevant features of the natural world into a computer program. It is the hope of the people working with these kinds of model to be able to support and extend existing arguments not fully supported by the existing empirical observations, possibly by adding qualitative considerations arising from an approximation to the dynamics of the real case.

2 Modes of Speciation

In biological research a species is traditionally defined as a *"group of actually or potentially interbreeding populations which are reproductively isolated from other such groups"* (Mayr, 1963). In other words a biological species can be characterized by the genetic differences that prevent genetic flow to and from another species. Despite its objectivity the use of this definition can meet several practical problems as discussed by Barton (Barton, 1988), and in general morphological features are used to make distinctions between species, although it is possible for two organisms to be morphologically very similar and yet be unable to exchange genes. However, when it comes to understanding the mechanisms by which two populations sharing a common ancestor become reproductively incompatible, the biological species definition provides an objective ground.

The problem of speciation is interesting because it is counter-intuitive to think of populations moving from one adaptive peak to another without passing through a valley of low fitness and being eliminated by natural selection. However if the gene flow between two populations of a given species is somehow diminished or interrupted for a sufficiently long time, then both populations may be able to accumulate a certain number of non-maladaptive but incompatible mutations. This is known as the Dobzhansky-Muller model of speciation (Dobzhansky, 1936, 1951; Muller, 1962; Orr, 1995). According to this model when two populations starting with identical genotypes at certain loci (aa,bb) become isolated (the simplest case is with two allopatric populations, i.e. geographically isolated), an A mutation may appear in one of them and be fixed, that is to say that Aabb and AAbb genotypes are viable, and also a B mutation may appear and remain fixed in the other population. Then, although, Aabb, AAbb, aaBb and aaBB genotypes may be perfectly viable and fertile, the A and B alleles may produce a deleterious effect together in the same genotype, resulting in hybrid inviability or infertility (Orr, 1995).

Of course this does not has to happen, but as shown by Orr when more loci are substituted the interactions become more complicated and the probability of incompatibility increases faster than linearly with this number. Incompatibilities are more likely to happen between substituted and “untested” alleles.

As mentioned above, the simplest isolating mechanism, is when the two populations are allopatric. This usually happens when some kind of geographical accident (a river, a mountain, an island, etc.) acts as a barrier between populations of a given species. In this case gene flow is physically interrupted as it is impossible or highly unlikely for individual members to cross the barrier. Depending on the relative sizes of the isolated populations it is possible to identify two extremes in allopatric speciation: when the populations are large and similar in size (the dumbbell model) and when a strong disparity exists in population numbers. The unbalanced genetic pool in the “founder” population in the latter case provokes many interesting effects leading to rapid speciation, also called peripatric speciation (Mayr, 1954).

It is still possible, however, for speciation to occur even if a certain level of gene flow is allowed to exist between populations, thus relaxing the assumption of purely allopatric populations. A species that is continuously distributed in a large habitat may be faced with different selection pressures in different regions of this habitat, thus favouring localized selection and providing a mechanisms for (at least partial) reproductive isolation. This is mainly due to the low fitness of individuals moving out of their local environment. This case is known as parapatric speciation. Whether this is not allopatric in a general sense is a semantic issue. However, some people remain doubtful about how many of the supposedly parapatric speciation cases observed in nature are not the result of previous isolation of populations that expanded their ranges later in evolutionary history (see Mayr, 1988).

Non-geographical types of speciation, with unimpaired gene flow in a randomly mating population, are also possible although they require new explanations other than those presented by the model of accumulation of incompatible mutations. This kind of (sympatric) speciation will not be contemplated in the present work.

Even if speciation does *not* happen, the mechanism of accumulated mutations is sufficient to explain the occurrence of populations (allopatric or distributed over qualitatively different continuous regions) characterized by a genetic continuity, and at the same time showing evidence of local adaptations. This is the concept of a *polytypic species* (Mayr, 1963, 1988; Dobzhansky, 1951), sometimes also called *Rassenkreis*. In this case, although reproductive isolation between "neighbouring" populations has not been established, gene flow is largely prevented between distant populations which may then become sufficiently separated (genetically) and reproductively incompatible.

The effects and mechanisms described above are fairly complicated to model and describe in analytical ways, and this is the motive for working with a computational model which, although simple, is able to capture many of their dynamic interactions and emergent properties.

3 The Model

The design of a computational model as a tool for investigating the dynamics of the speciation process should be based both on general simple assumptions about the world and on its potential flexibility for experimentation. It is not our intention to emulate specific details but, as with most models in scientific research, to capture the essential components relevant to the phenomena being modelled in order to obtain non-obvious results. As such, many features of the natural world, such as the conservation of energy and

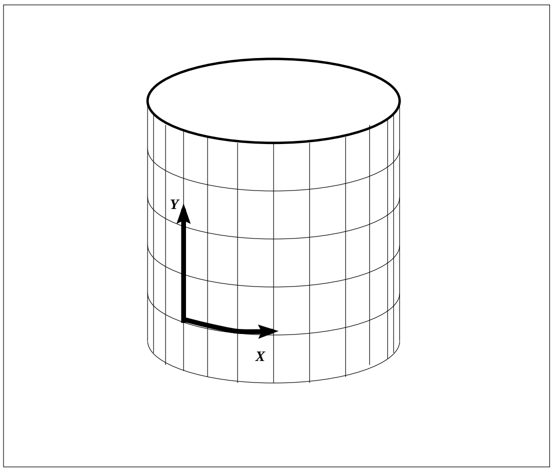

Figure 1: Cylindrical geometry of the environment.

the second law of thermodynamics, have been respected, while other features such as genetic encoding and agent behavior have been simplified for the sake of clarity in the analysis of the results. These are, of course, working assumptions, and no claim is made about their ultimate relevance in the process of speciation. One of the advantages of a computational model of this kind, as discussed later, is the flexibility in the relaxation of assumptions in order to expand it and compare new results with previous ones.

3.1 The Environment

In this model a population of agents "live", "reproduce" and "die" in an artificial environment. As the intention is to study general cases of speciation and not allopatric speciation in particular; much thought was given to the design of an environment in which cases of parapatric speciation are allowed. The environment can be described as a grid of 2-D cylindrical geometry with a much larger dimension in the aperiodic coordinate (y) than in the periodic one (x), as shown in Figure 1. Both dimension are much larger still than the agent's average single movement, (typical values are: $Agent'sMovement = 1; X_{max} = 100; Y_{max} = 1000$).

The whole environment is subject to a continuous flow of energy. This energy is stored in evenly distributed reservoirs from which the agents may extract a fixed amount of energy per unit of time for the costs of moving around, reproducing, and maintaining their homeostasis. The reservoirs are replenished at a fixed rate up to a saturation value, but the energy "used" by the agents does not return to the reservoirs and cannot be "used" again. It this way both the first and second laws of thermodynamics are observed.

Agents are born in the environment with a certain amount of initial energy (which, with the exception of the first generation, is provided by their parents), and this energy level is decreased every time the agents move by an amount proportional to the distance travelled. Also in every time step a certain amount of energy is transferred from the agent to the environment as heat dissipation. In order to calculate this quantity the environmental temperature, the agent's body temperature and the agent's "thermal resistance" are modelled.

Environmental temperature is a local feature of the environment, not necessarily uniform or constant, but continuous over space and time. In most of the experiments performed this parameter followed a smooth constant climatic gradient over the y coordinate, which varied linearly from a minimum, $T_{min} = 10$, to a maximum, $T_{max} = 30$.

3.2 The Agents

Each agent is characterized by a genotype which contains all the information for specifying its ultimate adult phenotype. However, no developmental mapping from the first to the latter has been included in the model and the translation is a direct one. This simplification may have certain consequences (see Discussion). The genotype is a simple haploid binary encoding of the phenotypic features. Each creature is modelled as a regular polygon of N sides with a skin thickness d, both specified genetically (see Figure 2). Six bits are used to encode the number N and four for d. N ranges from 3 ("000000" translates to "0 + 3") to 66 ("111111" or "63 + 3"); while d ranges from 0 to 15. These characteristics determine the thermal resistance of the agent according to a conduction-convection model of heat transfer. Neglecting boundary effects, the rate of heat dissipated $q"$ is:

$$q" = hP\frac{T_{body} - T_{env}}{1 + Bi}$$

where P is the perimeter of the creature, h the skin-air convection coefficient and $Bi = hd/k$ is the Biot number, k being the thermal conductivity of the skin.

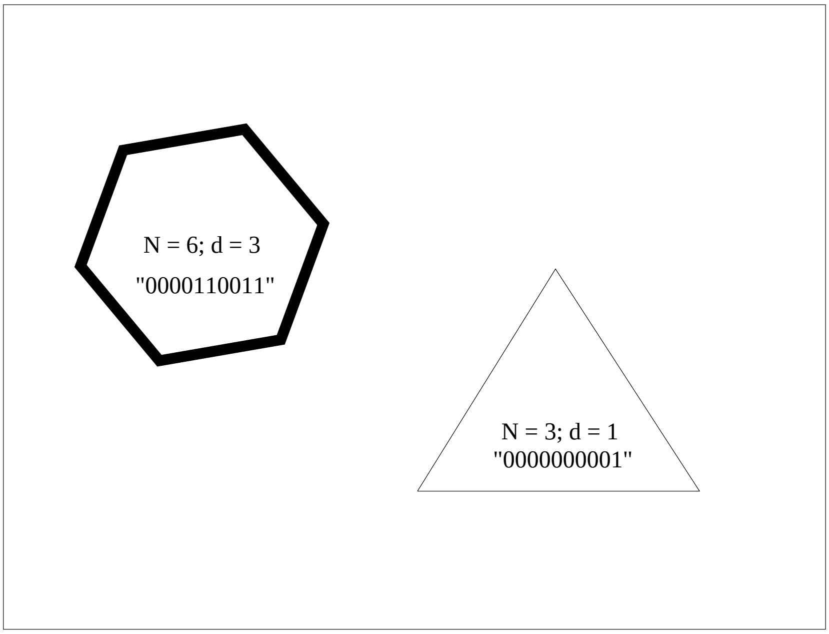

Figure 2: Examples of agents.

Some phenotypic features, however, remain constant for all the agents. These are the body volume, the internal temperature (usually 30), the skin thermal conductivity and the skin-air convection coefficient. If the body volume V for all agents is constant, the perimeter P will depend only on N according to:

$$P = 2\sqrt{VN\tan(\frac{\pi}{N})}$$

In this way thermal resistivity is only controlled by the parameters N and d.

When the energy level of an agent has reached a certain threshold value, the agent is ready to reproduce. For this it must find a mate. The only criterion for mate suitability is locality. The agent will only mate with other agents in its vicinity. The mating process is only modelled at a genetic level with the use of uniform crossover and random mutation operators like in a simple genetic algorithm. A reproduction energy cost is discounted from the first parent and transfered to the offspring, who initially shares the same environmental cell. The whole reproduction process is highly localized, an assumption that seems to be necessary for speciation to occur.

An agent continues to draw energy from the environment and spending it in movement and reproduction until it dies. There are two possible causes of death. The first occurs when the energy level reaches a minimum value (usually 0) and it is the result of a combination of low environmental resources and inability to maintain homeostatic equilibrium. The other cause of death is by direct rupture of homeostasis by low heat dissipation. This will happen when the environmental temperature is too hot and the value of dissipated energy falls below a minimum. Then a surviving agent will tend to be found in areas neither too cold nor too hot for it, in order to keep a good energy balance. No explicit requirement to fitness is used in this model and as a result of the local and individual character of reproduction and death, populations are allowed to vary in size.

The agents move in a random walk and this is the only behavior included in the model. There is no direct interaction between agents (except at reproduction) and they may even occupy the same cell without any behavioral change. There is, however, an indirect interaction that can be observed as emergent at the population level. Whenever agents accumulate in a localized region they tend to extract energy from the same reservoir, so that the chances of survival depend indirectly on the population density in the area they are inhabiting. This may work as a pressure for agents to migrate to other, perhaps more inhospitable regions but richer in resources and, hopefully, adapt to their new environments.

4 Simulation Results: Parapatric Speciation

As was mentioned above, the agents are placed in an environment characterized by a smooth linear temperature gradient along the y coordinate. A population consisting mostly of identical individuals known to be fit at the mean environmental temperature is uniformly distributed around the "center" of the world (mean value of y). These individuals were produced by allowing a random initial population of agents to evolve in an environment with this constant temperature. As expected many different genotypes survived. This is because both phenotypic traits being modelled contribute independently to individual fitness, allowing for more than one combination in the general case. One of these resulting combinations was chosen as the

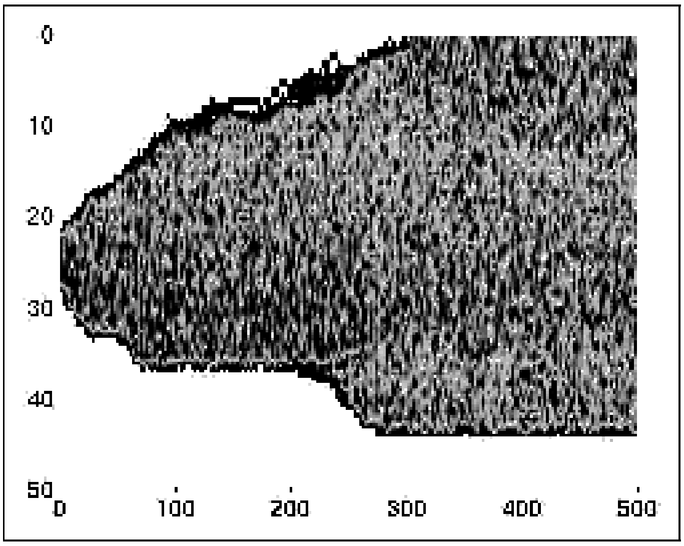

Figure 3: Whole population distribution. In this and the following figures the vertical axis corresponds to the aperiodic (y) axis of the environment (scale 1:20), and the horizontal axis corresponds to time steps (scale 1:20).

ancestor species.

In the non-uniform environment the population density is maintained but the environmental resources are reduced. This is done by decrementing the rate of replenishment of the energy reservoirs, thus providing the selective pressure for individuals to move to other areas with more resources. The first observed effect is an abrupt reduction in the size of the population. Gradually after that the area covered by the population starts to increase as a result of the diffusive effect of the random walk. This process of expansion follows a rule of proportionality to the square root of time.

In the run shown in Figures 3 to 8 (one of many yielding similar results) a population of 100 individual with N = 11 and d = 3 (11-3) is initially placed in the world between y = 450 and y = 550. In these figures the horizontal axis represents time, the vertical axis the y coordinate and the gray and black dots indicate that a number of individuals are occupying a cell with vertical position y at time t (darker dots meaning less individuals). Four major events are observed in this simulation. Around t = 1000 a subpopulation 3-3 appears towards the hotter side of the region occupied

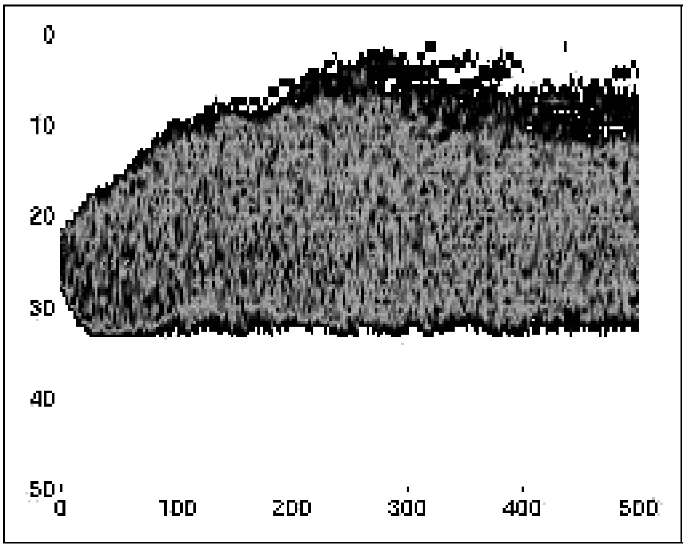

Figure 4: Distribution of ancestor population, 11-3, originally distributed between y = 450 and y = 550.

by the population 11-3, and begins to expand. This cannot yet be called a case of speciation for reasons explained below. It is important to notice that both populations share a significant amount of territory. Then at t ≈ 5000 a population 3-1 appears towards the hotter side of the area occupied by the agents 3-3 and a very short time after that another population, this time 7-1, is born between these two. The colder side of 11-3 is most of the time being conquered by this population, but at t ≈ 5000 a population 11-7 fills the coldest regions of the world.

Can any of these events be classified as a case of speciation? Only some of them. A simple genetic analysis can show that no two neighbouring population are reproductively isolated. As it turns out any offspring whose parents belong to different neighbour populations will itself belong to one of them. As an example let us consider the case between the populations 11-3 and 11-7. The respective genotypes are “0010000011” and “0010000111” and it is easy to see that any uniform recombination of these will result in the offspring having one or the other but not any different genotype (except when a mutation occurs). So in this case it is more correct to speak of a polytypic species with a genetic continuity, (this is simply a result of the

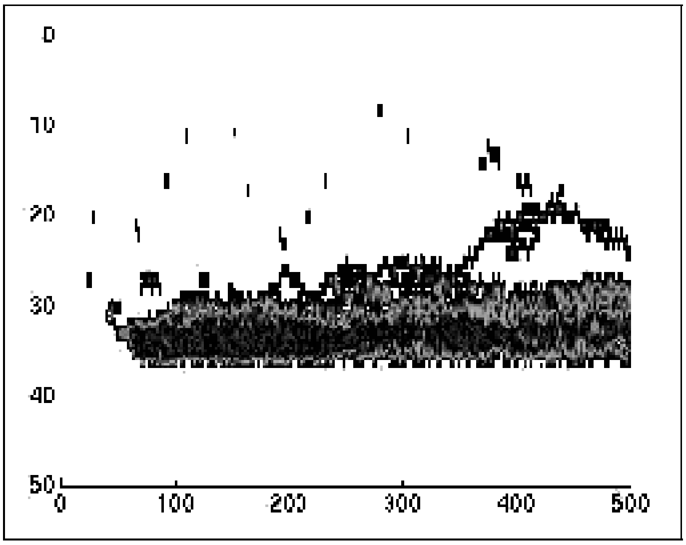

Figure 5: Distribution of population, 3-3, appearing first at t ≈ 1000. Note that there is a zone of coexistence with population 11-3.

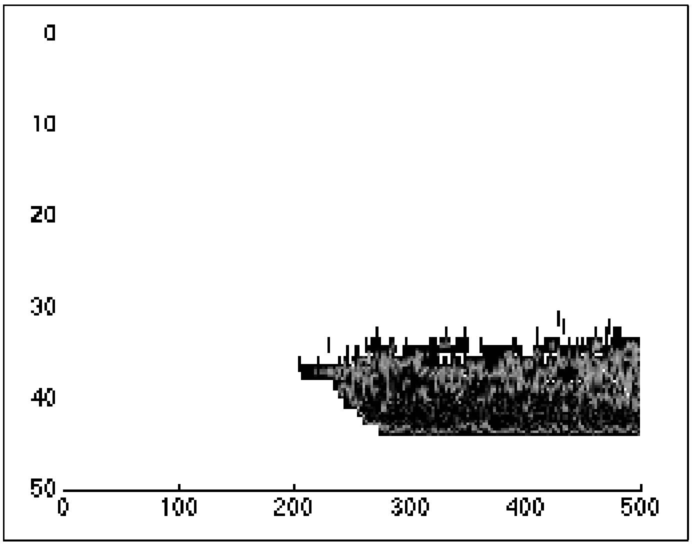

Figure 6: Distribution of population, 3-1, appearing around t ≈ 5000.

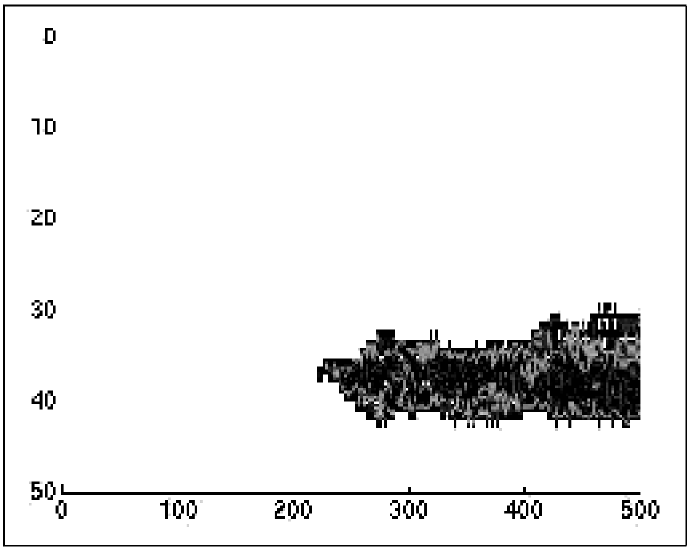

Figure 7: Distribution of population, 7-1, appearing around t ≈ 5000 between population 3-3 and 3-1. It probably derives from 3-1 as the genetic distance is shorter with this population than with 3-3

type of genetic encoding used in this model, but it will be argued that this is the general case in parapatric speciation).

On the other hand, because they are adapted to different local conditions, it is extremely hard for two individuals of non-neighbouring populations to mate simply because the high costs associated with travelling to regions other than their own. Gene flow is severely diminished, although not strictly interrupted, and non-neighbouring populations become allopatric *de facto*. This is what happens with the pairs 11-3/7-1, 11-7/7-1 and 11-7/3-1. The degree of geographical isolation permits further accumulation of complementary mutations, and these pairs, which in principle are able to reproduce should the populations re-encounter, have a good chance of producing individuals with very low fitness [1]. In other words, the requirements of genetic continuity allow for incompatible mutations to appear in non-neighbouring populations. In this model due to the simplicity of the genetic encoding

---

[1] Genetic continuity may persist between non-neighbouring population, as in the case of 11-3 and 3-1.

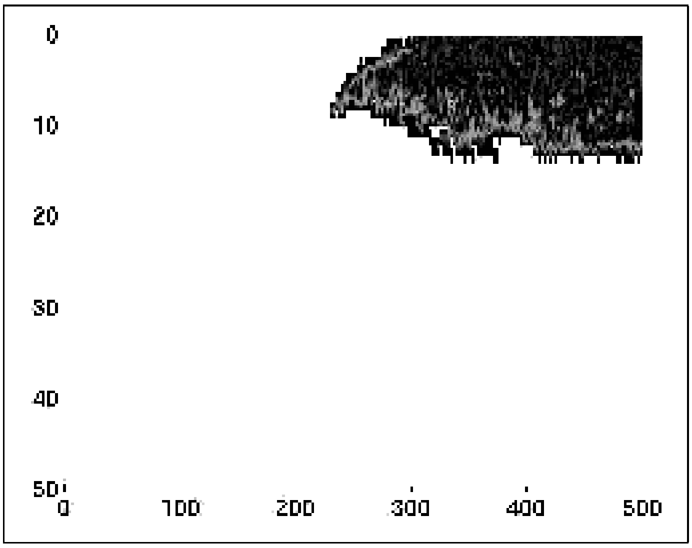

Figure 8: Distribution of population, 11-7, appearing around t $\approx$ 5000, towards the hotter side of the ancestor population.

and the lack of a developmental process, this “hybrid” genotype will in fact be viable, but it will have a low probability of existing beyond the first generation. In this case it makes sense to speak of different species.

It is interesting to notice that if speciation is to occur parapatrically, without any real physical barriers between populations, then there are two necessary requirements for the first mutations in the ancestor genotype to survive, namely: (1) these mutations must translate into an adaptive or neutral change for the local environment and (2) mutated genotypes must be able to replicate when combined with the ancestor genotype in order to achieve a critical number of individuals that carry the new mutation. This is simply because it is highly unlikely for an individual carrying a new genotype for the first time to find other individuals in the same condition and give birth to a new interbreeding population by themselves. This is independent of how well adapted the new genotype really is, if it is not able to be replicated, and the only way to do this is by “using” the ancestor genotype, it will not form a new population. The expected result then is that neighbouring populations will tend to be able to interbreed, but due to their localized adaptiveness they will remain distinguishably distributed in the environment.

5 Discussion

The model presented in this paper does not include any mechanisms of development from genotype to phenotype. It is to be expected that the true nature of biological speciation: the inviability of infertility of hybrids, will not be replicated in these simulations. As stated by Dobzhansky (Dobzhansky, 1951):

> "In the course of evolution, the function of a gene in the development may undergo such changes that the gene may subtend developmental processes other than those with which it was previously concerned. If the gene functions in two or more races or species, the gene system may become no longer compatible in hybrids." (p. 207).

However, on account of an effect that may be considered similar to this one when viewed over evolutionary periods, namely the low fitness of hybrids, speciation is achieved within this computational model. It is possible, in future implementations, to include some simple developmental rules, such as a checking of genotype compatibility, or "fertility genes", or more complex behaviors, such as mate choice, that may provide a closer approximation to the real case. It is the assumption of the present work that such changes to the model will not provide significant qualitative differences in the evolutionary dynamics.

Although most cases of speciation in nature are thought to be allopatric, it was the purpose of these experiments to concentrate on the possibility of parapatric speciation and in this way illustrate verbally supported arguments (Barton, 1988) from a dynamical systems point of view. One interesting effect highlighted by this approach which, although not new, is usually overlooked, is the need for newly mutated genotype to be able to crossbreed with the ancestor genotype if a new population is to be formed, and then, by definition, constituting only a case of "subspecies" or polytypic species. It is important to notice that this shows that better adaptation at the individual level does not necessarily mean better survival value of the genes. Between a neutral mutation capable of being replicated with the ancestor genotype and an adaptive mutation unable to do so, the first will have more chances of being propagated. This dynamical effect prevents the formation of species that share part of their ancestor's habitat. Only when two population have become "separated" by a third one able to interbreed with both, there is the possibility of speciation and this may be viewed as the emergence of allopatricity without prior physical barriers. This adds some qualitative considerations to the Dobzhansky-Muller model as presented by Orr (Orr, 1995); although in principle only a pair of incompatible mutations is sufficient for speciation to occur, this is only valid in the case of initially allopatric population (the case covered in that model). In the absence of physical barriers the minimum number of accumulated mutations to prevent interbreeding will tend to increase due to the requirements imposed by the genetic continuity between neighbouring populations, thus providing a kind of phyletic inertia. After the above facts have been taken into account the Dobzhansky-Muller model can also explain parapatric speciation in non-uniform environments.

The necessity for this previous "step" of subspecies formation in parapatric speciation is not obvious *a priori* from the verbal argument alone. It is an interesting conclusion that, even though mutations are the result of a random process, surviving mutations will depend on the environmental pressures *and* on historical facts as well, such as the kind of preceding ancestor. The ancestor population will constraint the formation of new populations with mutated genotypes not only as a "starting search point", but also as a reference for genetic continuity. In the example described in the previous section many adaptive variations can be made to the 11-3 population by means of a single mutation (specially towards the colder regions of the world), however only a few are able to perpetuate themselves by "using" the ancestor genotype to replicate.

Experiments with this computational model not only support the argument for parapatric speciation in the framework of the Dobzhansky-Muller model, but provide an interesting qualitative extension to both as well.

As in the above discussion, many explanations in evolutionary biology are, of necessity, of a historical nature, and, as such, are better expressed in models that are able to capture the dynamics of the process being explained. This computational model has some advantages viewed in this perspective over other, analytical or verbal, models. Although not included in this report, it is possible to very easily simulate temporal variations in the environment of any kind, including catastrophic, very short term, phenomena. This allows the study of many complex processes that are very hard to model analytically and very counter-intuitive to understand verbally. Examples of these are the influence of dynamical features of the environment in speciation.

On the other hand, flexibility may turn out to be disadvantageous as the researcher may be "carried away" very easily, and be tempted to include many features in a given model making the analysis of results very difficult. The modelled mechanisms and processes need to be considered with care and assumptions about what is being included must be justified. One good heuristic method is to build the model incrementally, adding new features only once the dynamics of the previous stages are well understood. Then it is possible to have some basis of comparison as to the contribution of different ingredients in the final result. Such will be the case with this model, when other features (e.g. development mechanisms) are added.

6 Acknowledgements

I wish to thank Phil Husbands for his comments on previous versions of this article. This work is financially supported by the *Consejo Nacional de Invetigaciones Científicas y Técnicas de la República Argentina* and the Argentine Ministry of Education.

Reference

Barton, N. H. (1988). Speciation. In A.A. Myers, P. G. (Ed.), *Analytical Biogeography*, pp. 185 – 218. Chapman and Hall, London.

Dobzhansky, T. (1936). Studies on hybrid sterility. ii. localization of sterility factors in drosophila pseudooscura. *Genetics*, *31*, 113 – 135.

Dobzhansky, T. (1951). *Genetics and the Origin of the Species* (3rd edition). Columbia University Press, New York and London.

Mayr, E. (1954). Change of genetic environment and evolution. In Hardy, J. A., & Ford, E. (Eds.), *Evolution as a Process*, pp. 157–180. Allen and Unwin, London.

Mayr, E. (1963). *Animal Species and Evolution.* Belknap Press, Cambridge, Mass.

Mayr, E. (1988). *Towards a New Philosophy of Biology.* Belknap Press of Harvard University Press.

Muller, H. J. (1962). *Studies in Genetics: the selected papers of H.J. Muller*, chap. Bearing of the Drosophila work on systematics. Indiana, U. P.

Orr, H. A. (1995). The population genetics of speciation: The evolution of hybrid incompatibilities. *Genetics*, *139*, 1805 – 1813.