

Genetic Footprints of Speciation

Pleuni Pennings

*Section of Evolutionary Biology, Department of Biology II,
Ludwig-Maximilians-University, Munich, Germany*

Goal

To set up an individual-based model that allows for different speciation processes to be studied in one modeling environment, and to use this tool to study how patterns of variation in non-coding genetic sequences are influenced by different speciation processes.

Background and motivation

One of the main aims of evolutionary biology is to explain the species diversity that we see today and in the fossil record. Speciation apparently takes place often enough to give rise to a high species diversity, but not so often that we cannot distinguish species anymore. Understanding the process of speciation is therefore a central theme in evolutionary biology. The question of speciation can be split up into the following subquestions (where subquestion (4) can be seen as the main question): (1) Which processes can lead to speciation and what mechanisms drive these processes? (2) Under what circumstances do these mechanisms operate? (3) How will these mechanisms shape observable (genetic, ecological, geographic or other) patterns? (4) What mechanisms and processes have shaped and shape the species diversity that we observe today?

Concerning subquestion (1) Dieckmann et al (2004b) suggest a classification of speciation processes that uses three axes of differentiation: ecological, mating, and spatial differentiation, and also three mechanisms that can drive the differentiation: selection, drift, and external factors. A speciation process is then characterized by the route through the thus-defined three-dimensional space, and by the mechanisms that drive (different parts of) the route. The route to speciation in sexual organisms always starts at the origin of the three-dimensional space (no differentiation in any of the three directions) and ends with at least mating being differentiated. Also, after (successful) speciation, the two new species should be differentiated with respect to at least one of the other two axes, geography or ecology to allow their persistent coexistence. This follows directly from what is maybe the best established law of ecology, the law of competitive exclusion: two species can not coexist if they use exactly the same resources. The classical notions of allopatric and sympatric speciation can be incorporated into this classification. Allopatric speciation starts with an abrupt move on the spatial axis driven by an external factor. Sympatric speciation involves no move on the spatial axis at all. Adaptive speciation involves at least some movement that is driven by selection. It should be clear that, at least in theory, there are many routes to speciation.

The answers to subquestions (2) and (3) can both lead to an answer to (4). For example, if we knew exactly under what circumstances (in terms of parameter ranges) speciation by ecological differentiation occurs (i.e. the answer to subquestion (2)), we could estimate the relative importance of this process for speciation by determining the appropriate parameter values from data. In the same way, if we knew the resulting

genetic pattern of speciation by ecological differentiation (i.e. the answer to subquestion (3)), we could collect the right (genetic) data from pairs of sister species and determine how many of these pairs show the pattern caused by ecological differentiation and thereby estimate the relative importance of this process. Ideally, theoretical (and experimental) studies should provide answers to (2) and (3), and those answers should guide the direction of data collection.

Research on speciation is traditionally much more focused on subquestion (2) than on subquestion (3). Maybe this is simply because, until recently, no one could have imagined the amount of data that is available now or will be available in the near future. What are "observable patterns" has changed dramatically in recent years, especially regarding patterns of DNA sequence variation. Yet, this altered range of options has not yet been exploited for the study of speciation. This project is aimed at finding ways of using patterns of DNA sequence variation (hereafter called genetic patterns) to infer speciation processes. Apart from the availability of large amounts of data, there are two other reasons why now is a good time to look at these genetic footprints of modes of speciation. First, any work in this direction can build upon a framework for individual-based simulations that has been developed by Michael Doebeli, Ulf Dieckmann, and others in recent years. Second, with the power of today's computers, the complexity of the needed simulations should no longer constitute a problem.

Two lines of research are worth noting here. One is the work that was started by Barraclough and Vogler (2000) to try and exploit our knowledge of geographic patterns to infer the predominant speciation process in a clade. The other is the extensive work on how postzygotic isolation can evolve (see Orr (1995), Kondrashov et al. (2002), and Welch (2004)). These two lines of research do not overlap with what I suggest in this proposal but they are based on the same question: what can we learn about past speciation processes by looking at observable patterns today?

Research questions

In this project I aim to look at the effects of different speciation processes on genetic patterns. If I can identify these effects, this should allow me to infer (speciation) process from (genetic) pattern. The single main question is:

- How do different speciation processes shape genetic patterns?

With respect to the speciation processes, this main question can be refined:

- Which aspects of a speciation process have an effect on observable genetic patterns?

It could be, for example, that either the mechanism responsible for speciation or the geographic aspect of the speciation process is most important. Looking at it from the side of the genetic pattern, the main question can also be refined:

- Which aspects of DNA sequence variation are influenced by a speciation process?

This could for example be the number of segregating sites or the shape of the coalescent tree. And it could be necessary to look at many loci, or at rather specific loci. I will perform different neutrality tests (such as the HKA test (Hudson et al. 1987) and Tajima's D test (Tajima 1989)) on data harvested from the simulations to see if such standard tests may point to relevant differences.

Methods and work plan

Methods

The methods that I will use for this project are (i) individual-based simulations to generate data and (ii) standard population genetic methods to analyze them.

Steps

I will carry out the following steps. The first two steps will mainly consist of programming a model that incorporates all the necessary aspects.

1. Allow for different speciation processes to occur in one modeling environment so that direct comparisons between the different processes will be possible (see below for details of the models).
2. Add neutral loci, with appropriate recombination and mutation rates. This has already been modeled by Sergei Semovski and Yurji Bukin from Irkutsk; it should thus be easy to include these features in the model.
3. Run the simulations with different parameter values and harvest the information from the neutral loci. Some careful experimental design will be required in order to be able to explore different parts of the parameter space, while still generating enough data for each combination of parameter values.
4. Analyse the data from step 3 using tools from population genetics.

Model

Initially I will compare (a) strict sympatric speciation due to ecological differentiation in a non-spatial model (as described in Dieckmann and Doebeli 1999) with (b) strict allopatric speciation followed by reinforcement after secondary contact. The model should be built in such a way that scenario (a) and (b) can be run with only small changes to the parameter settings.

The model will describe a population of hermaphrodites that reproduce sexually. Character values will be determined by many additive diploid diallelic loci. There will be three important traits in the model: an ecological trait that determines resource use, a mating trait that determines assortativeness, and a marker trait that determines the preference in case of assortative mating. A fourth trait (a so-called internal trait) will be added later. There will be a resource that is distributed following a Gaussian distribution. In this model, the ecological trait value of the population will first evolve towards the optimum of the resource distribution, i.e., to where the carrying capacity has its maximum. At this optimum, the population may find itself at a fitness minimum if the curvature of the carrying capacity at its maximum is less than that of the competition function. Simply said, under these circumstances it pays off to be different from the rest because individuals gain more from avoiding competition than they lose because further away from the resource maximum there is (obviously) less resource.

As long as mating is random, recombination will always prevent evolutionary branching because it will create intermediate phenotypes. However, if there is linkage disequilibrium between the marker trait and the ecological trait, there is selection pressure for mating to become assortative. This is the case if, for example, birds that eat large seeds have slightly greener feathers and birds that eat smaller seeds have more yellow feathers (because of drift in a finite population). If a mutation now caused a bird to prefer to mate with a similar looking bird, this mutation would give the bird a selective advantage, and assortativeness could evolve, eventually leading to reproductive isolation.

In scenario (b) the population will be split in two by a virtual geographic barrier. Individuals will be assigned to one or the other subpopulation at random. To avoid evolutionary branching within the two identical subpopulations, the parameter determining the 'width' of the competition function will be enlarged. To allow for speciation to occur without ecological differentiation, a so-called internal trait has to be added to the model. This trait has no absolute fitness optimum, but offspring survival is highest when the distance between the trait values of the parents is zero. This trait will also be governed by several diallelic loci with additive effects. Since it would be more realistic if this trait would not be bound in its values, I could allow effect size of the different loci to evolve (in which case the loci do not remain diallelic, but infinitely many alleles could evolve). The mean value of this trait will change due to mutation and drift, not selection, while the variance of the trait value within a population is restricted by selection. If the original population is split in two and the two subpopulations evolve independently, the internal trait will (after some time) have different mean values in the two populations. The distance between those mean values will (in expectation) increase with time. If the two populations come into secondary contact, there will be selection for prezygotic isolation. And if the marker trait (that did not play a role until then) is in linkage disequilibrium with the internal trait, assortative mating can evolve (in the same way as in scenario (a)), so that the two populations will stay reproductively isolated. The internal trait has already been modeled by Géza Meszéna from Budapest, so it should be easily added to the model. In scenario (a) the addition of the internal trait after evolutionary branching will probably not change the process much, although it may slightly facilitate speciation.

Scenario (a) and (b) can both be seen as a two-patch system, in (b) there is no migration between the two patches, whereas in (a) the migration rate is so high that the two patches are effectively one. In reality the migration rate can also take intermediate values, and I will allow for this in my model.

In addition to scenario (a) and (b), it will be interesting to look at the effect of spatial resource heterogeneity (SRH), in times of the two patches differing in resource distribution. Day (2000) has analyzed the effect of SRH on evolutionary diversification in a two-patch system. He finds that, if migration rates are not too high, SRH facilitates evolutionary diversification. Spatial heterogeneity will constitute a third relevant axis of parameter space, the first two being migration and the ratio between the width of the resource distribution and the width of the competition curve.

Neutral loci will also be added to the model. These loci will be linked to the other loci, with a small probability of recombination in each generation. They will consist of a stretch of nucleotides that can mutate according to a simple mutation model (the one-parameter Jukes-Cantor model to start with, giving equal probabilities to every possible single nucleotide mutation). Simulations should run for a while to reach mutation-drift equilibrium, before evolutionary branching will be allowed in the first scenario, or before the population will be split in the second scenario. After the simulations have run, genetic sequences from all individuals will be collected and analyzed using standard population genetic methods.

Schedule

First month: The four steps described before should be carried out one after the other. It will be useful to do a round of these four steps relatively soon to get a first impression of how the model behaves and what needs to be changed.

Second month: With this knowledge I will make changes to the program (improving on steps 1 and 2) before starting a second round of simulations (step 3) and analyses (step 4).

Third month: I will use the last month of the summer program mainly to start writing down the results of this project.

Intermediate result: If the model with two scenarios (a and b) works (that is, step 1 of the work plan is carried out successfully;) I can obtain some intermediate results. Being able to simulate different modes of speciation in one model will allow me to directly compare the behavior of the two models. It would, for example, be interesting to determine whether the time needed for speciation is different in both models. These kinds of relatively simple comparisons have been complicated in the past by the fact that models for distinct modes of speciation differed too much in their set-up.

Relevance and link to ADN's research plan

In ADN, much previous research has been devoted to speciation. This project will build on and extend this work.

Expected output and publications

The study described here is expected to result in a co-authored paper in an international journal and will be integrated as a chapter in my PhD thesis.

References

- Barracough, T.G. and Vogler, A.P. (2000). Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist*, 155: 419-434.
- Day, T. (2000). Competition and the effect of spatial resource heterogeneity on evolutionary diversification. *American Naturalist*, 155: 790-803.
- Dieckmann, U. and Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature*, 400: 354-357.
- Dieckmann, U., Tautz, D., Doebeli, M. and Metz, J.A.J (2004). Epilogue. In *Adaptive Speciation*. Dieckmann U., Doebeli, M., Metz, J.A.J. and Tautz, D. (eds). pp. 380-394. Cambridge University Press.
- Hudson, R.R., Kreitman, M. and Aguadé, M. (1987). A Test of neutral molecular evolution based on nucleotide data. *Genetics*, 116: 153-159.
- Kondrashov, A.S., Sunyaev, S. and Kondrashov, F.A. (2002). Dobzhansky-Muller incompatibilities in protein evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 14878-14883.
- Orr, H.A. (1995). The population-genetics of speciation - the evolution of hybrid incompatibilities. *Genetics*, 139: 1805-1813.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123: 585-595.
- Welch, J.J. (2004). Accumulating Dobzhansky-Muller incompatibilities: reconciling theory and data. *Evolution*, 58: 1145-1156.