

New Techniques for Estimating the Timing of Speciation from Molecular Data

Yurji Bukin

*Limnological Institute, Siberian Branch of the
Russian Academy of Science, Irkutsk, Russia*

Goal

To examine the reliability of using pairwise differences between mtDNA sequences (and of other methods of molecular analysis) for inferring the timing of past speciation, using individual-based evolutionary models.

Background and motivation

Today a host of different molecular data exists for many organisms. With the help of molecular data we can study gradual evolution, speciation, and others population processes. Even though different molecular methods exist for this purpose, there is a problem of assessing the reliability of existing techniques and of devising new techniques where necessary. In this study I will focus on methods based on pairwise differences between mtDNA sequences and compare these with alternative methods.

We use a numerical evolutionary model for the study of speciation processes. Modern computing greatly facilitates the investigation of such individual-based models. These models correspond more closely to natural populations of organisms than do other, simpler models. Simulating the molecular evolution of neutral “nucleotide” sequences in addition to that of selected characters gives us an opportunity of investigating the applicability and reliability of available phylogenetic techniques. The neutral sequences considered in our model are transferred to offspring via the maternal line, which is analogous to the transfer of mitochondrial DNA (mtDNA).

Phylogenetic evolutionary trees can be constructed using molecular sequences extracted at the end of simulation runs; these demonstrate the possibility of reconstructing evolutionary history. Numerical indicators can assess the genetic exchange between subpopulations. Also environmental change, resulting in changing population sizes, can be traced based on neutral molecular data (Strimmer and Pybus, 2001, Emerson et al., 2001). The individual-based model allows us to estimate the reliability of these and alternative methods (Semovski and Sherbacov, manuscript, Emerson et al., 2001). In addition, questions about the speed of molecular evolution can be addressed by studying the influence of model parameters on the speed of the ‘molecular clock.’

An important method of molecular analysis is based on the distribution of pairwise differences between mtDNA sequences. This technique has been used to study the timing of speciation events and of abrupt environmental changes. We have compared this method with alternative methods of molecular analysis (Semovski and Sherbacov, manuscript, Semovski et al., 2004, Semovski et al., 2003). We used parametric bootstrapping to estimate the statistic significance of observed differences in the distribution of pairwise DNA differences (Semovski et al., 2003).

This method may also allow detecting the effects of interspecific interactions on evolving neutral mtDNA sequences (Sankoff and Nadeau, 2000). In particular, the interaction between organisms of different species may be responsible for differences between the speeds of their molecular evolution. To explore these questions, and to provide a test bed for the comparison of alternative molecular methods, I will focus in this study on a model of predator-prey coevolution.

Research questions

My study will address three interrelated sets of issues:

- I will compare the method based on the distribution of pairwise differences between mtDNA sequences with a method for reconstructing phylogenetic trees. In particular, I will assess their differential performance in estimating the timing of past speciation events. In addition, I will compare the reconstructed phylogenetic trees of evolution with the ‘real’ phylogenetic trees directly recorded in our numerical simulations.
- I will study the reliability of the method based on the distribution of pairwise differences between mtDNA sequences. In particular, we must understand when we may use this method safely and when its results are not to be trusted. In this context, spurious effects resulting from finite sample size have to be explored.
- I will study the influence of interspecific interactions on the speed of molecular evolution. This may suggest methods for recalibrating molecular clocks when studying coevolutionary dynamics.

Methods and work plan

The first set of instruments for my work contains methods for reconstructing phylogenetic trees. There are established algorithms for inferring phylogenetic trees from DNA data.

The second instrument for my work is the distribution of pairwise differences between mtDNA sequences. I have devised a computer program to construct distributions of these differences. This program constructs the distribution and estimates its reliability. It can be applied to the output of individual-based evolutionary models.

The third instrument for my work is an individual-based model of predator-prey coevolution. The deterministic approximation of the population dynamics in this model is given by the following equations (Doebeli and Dieckmann, 2000),

$$\frac{dN_1(x)}{dt} = r_1 N_1(x) \left[1 - \int C_1(x-x') N_1(x') dx' / K_1(x) \right] - p N_1(x) \int C_2(y'-x) N_2(y') dy' ,$$

$$\frac{dN_2(y)}{dt} = r_2 N_2(y) [1 - N_2(y) / K_2] + pa N_2(y) \int C_2(x'-y) N_1(x') dx' ,$$

where $N_1(x)$ and $N_2(y)$ are the population sizes of prey and predator, x and y are their phenotypes, C_1 defines competition between prey individuals, and C_2 defines the ability of predators to harvest prey. The intraspecific competition in the prey population causes

its phenotypic divergence. Such divergence may cause the divergence of predators (Doebeli and Dieckmann, 2000). All individuals in my model possess ecologically neutral “nucleotide” sequences, as described in the preceding section.

First, I will investigate different evolutionary scenarios with the help of the individual-based model. I will then use the final DNA sequences extracted from the model to compare the method based on the distribution of pairwise differences between mtDNA sequences with other methods and to assess the reliability of the former. Molecular data collected throughout simulation runs will give me the possibility to study the speed of the molecular clock and to construct the ‘real’ phylogenetic trees.

Relevance and link to ADN’s research plan

In this project we aim at building a framework for the study of speciation and coevolution based on selected and neutral molecular data. This will give us useful insights for studying these phenomena in natural populations. The planned program of research will extend previous studies carried out in the context of ADN’s research focus on *Adaptive Speciation*. In particular, the coevolutionary model studied here is based on the one proposed by Dieckmann and Doebeli (1999).

Expected output and publications

This work will be the final part of my PhD thesis and is expected to be published in a jointly authored paper.

References

- Dieckmann, U. and Doebeli, M. (1999). On the origin of species by sympatric speciation, *Nature*, 400:345-357
- Doebeli, M. and Dieckmann, U. (2000). Evolutionary branching and sympatric speciation caused by different types of ecological interactions, *The American Naturalist*, 156:S77–S101
- Emerson, B.C., Paradis, E., Thebaud, Ch. (2001). Revealing the demographic histories of species using DNA sequences, *Trends in Ecology and Evolution*, 16(12): pp. 707-716
- Irwin, D.E. (2002). Phylogeographic breaks without geographic barriers to gene flow, *Evolution*, 56(12):2383-2394.
- Sankoff, D. and Nadeau, J. (2000). *Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment, and the Evolution of Gene Families*. Kluwer Academic Publishers, Dordrecht, pp. 525-536.
- Semovski, S.V. and Sherbacov, D.Y. (manuscript). Genetic traces of environmental variations in ancient lakes.
- Semovski, S.V., Verheyen, E., Sherbacov, D.Y. (2004). Simulating the evolution of neutrally evolving sequences in a population under environmental changes, *Ecological Modelling*, in press.
- Semovski S.V., Bukin, Y. S., Sherbacov, D. Y. (2003). Speciation and neutral evolution in one-dimensional closed population, *Int. J. Mod. Phys.*, 14(8): pp 973-983.
- Strimmer, K. and Pybus, O.G. (2001). Exploring the demographic History of DNA sequences using the Generalized Skyline Plot, *Molecular Biology and Evolution*, 18(12):2298-2305