



The probability of speciation on an interaction network with unequal substitution rates



Peter Olofsson^{a,b,*}, Kevin Livingstone^c, Joshua Humphreys^c, Douglas Steinman^a

^a Department of Mathematics, Trinity University, United States

^b School of Engineering, Jönköping University, Sweden

^c Department of Biology, Trinity University, United States

ARTICLE INFO

Article history:

Received 5 March 2016

Revised 13 April 2016

Accepted 24 April 2016

Available online 10 May 2016

Keywords:

Bateson–Dobzhansky–Muller interactions

Interaction networks

Reproductive incompatibility

Speciation

ABSTRACT

Speciation is characterized by the development of reproductive isolating barriers between diverging groups. A seminal paper of a mathematical model of speciation was published by Orr (1995), extended by Livingstone et al. (2012) to incorporate interaction networks. Here, we further develop the model to take into account the possibility of different substitution rates for network nodes of different connectivity. Mathematically, this amounts to sampling nodes from an undirected graph where the inclusion probability for a given node depends on its degree (number of connecting edges). We establish formulas for the rate of speciation and identify a crucial parameter that is a measure of the deviation from simple random sampling.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

The Bateson–Dobzhansky–Muller (BDM) model describes how fixation of mutations in allopatric populations could produce inviability or sterility in hybrid offspring, without the mutations lowering fitness within either population. Briefly, the BDM model starts with an ancestral (diploid) population of genotype *aabb*; in one population, the *A* allele arises and becomes fixed, while in the other population, *B* arises and is fixed. The resulting hybrid from the *AAbb aabb* cross would have genotype *AaBb*, and as *A* and *B* have never been “tested” together, they could interact epistatically to cause a genetic incompatibility. The accumulation of such Bateson–Dobzhansky–Muller incompatibilities (BDMIs) can cause permanent isolation, and hence speciation.

In [7], Orr introduces a mathematical model for the rise of BDMIs. In this model, two diverging lineages fix alleles at *K* loci between them, and each new allele has a probability *p* of being incompatible with an allele (derived or ancestral) from the other lineage at one of the *K* – 1 loci at which substitutions have occurred. It is demonstrated how the expected number of incompatibilities increases as a function of *K*², a phenomenon referred to as “snowballing,” (see [6]) and how the probability of speciation also depends on *K* through *K*². Here and subsequently, speciation is simply defined as the occurrence of at least one BDMI.

In [5], we elaborated upon Orr’s model by recognizing that most nodes in a real biological network are connected to only a small number of other nodes, while very few nodes act as central hubs with a large number of interactions (see [2]). We modeled an interaction network as an undirected graph where nodes are loci and edges are existing interactions, each edge leading to a BDMI with probability *p*. The crucial parameters of the network turn out to be the number of nodes, *N*, and the number of edges, *N_E*, which are used to define the density $\alpha = N_E / \binom{N}{2}$ (Orr’s model would thus correspond to a complete graph with $N_E = \binom{N}{2}$ and $\alpha = 1$).

The rise of *K* substitutions is modeled by randomly sampling *K* nodes from the network, resulting in a subgraph where each edge (existing interaction) may lead to a BDMI, with probability *p*. Thus, the more edges, the more likely a BDMI and also note that the number of edges is a random variable with range $\{0, 1, \dots, \binom{K}{2}\}$, whereas in Orr’s model we always get $\binom{K}{2}$ edges. It is demonstrated that, to a first order of approximation, the formula for the probability of speciation contains α as a parameter but is otherwise similar to Orr’s. Effectively, the parameter *p* (probability of a single BDMI) is replaced by the product αp .

The sampling in [5] is done randomly, that is, every node is equally likely to be included in the sample of size *K*. Biologically, this translates into all alleles being equally likely to become substituted; however, as is evident from the literature on the subject, this assumption is questionable. For example, in [3], it is observed that the connectivity of well-conserved proteins in the protein–protein interaction (PPI) network for the yeast *S. cerevisiae* is

* Corresponding author.

E-mail address: polofso@trinity.edu (P. Olofsson).

negatively correlated with their rate of evolution, and [4] reports similar findings for yeast, and also for two other species. Indeed, in [4] it is pointed out that there is a “consistent reduction in evolutionary rate for essential proteins in all three species: essential genes in the protein interaction network evolved at 70% the rate of nonessential genes (yeast: 70.5%; worm: 71.4%; fly: 70.1%)”. To allow for this possibility, we now extend our model from random sampling to sampling where a node has an inclusion probability that may depend on its degree (number of interactions). Although the biological motivation is the aforementioned negative correlation, our model is not restricted to such assumptions but allows for any kind of dependence of sampling probabilities on connectivity.

2. Mathematical model

We consider a network with N nodes such that there are N_i nodes of degree i for $i = 1, 2, \dots$, where we note that the theoretical upper limit for i is $N - 1$. We define the *degree distribution* as the sequence $(q(1), q(2), \dots)$ where

$$q(i) = \frac{N_i}{N} \quad (1)$$

so that the *mean degree* is

$$\mu = \sum_i iq(i) \quad (2)$$

which can be thought of as the expected degree of a randomly chosen node. Let N_E denote the number of edges in the network and define the *density* of the network as

$$\alpha = \frac{N_E}{\binom{N}{2}} \quad (3)$$

As every edge connects to 2 nodes, we get the relation

$$2N_E = \sum_i iN_i \quad (4)$$

and combining this with (1) and (2) yields

$$2N_E = N\mu \quad (5)$$

and as $\binom{N}{2} \approx N^2/2$, we also note that

$$\mu \approx N\alpha \quad (6)$$

relations we will make use of later.

Now sample K nodes in a way such that a given node of degree i has a probability $f(i)$ of being chosen when picking a single node. The case of random sampling, as in [5], corresponds to a uniform distribution: $f(i) = 1/N$ for all i . Let D be the degree of a node chosen according to the $f(i)$ and note that, as there are $Nq(i)$ nodes of degree i , we get

$$P(D = i) = Nq(i)f(i) \quad (7)$$

and hence expected value

$$E[D] = \sum_i iNq(i)f(i) \quad (8)$$

As this is the mean when choosing according to the $f(i)$, and μ is the mean when choosing randomly, we define the ratio

$$r = \frac{E[D]}{\mu} \quad (9)$$

which can be thought of as a measure of deviation from random sampling. Note that random sampling gives $E[D] = \mu$ so that $r = 1$. If nodes of higher degrees are less likely to be chosen (less prone to substitution), we typically get $r < 1$.

3. Results

Let X be the number of edges that we get when sampling K nodes according to the $f(i)$. The key observation used to compute both speciation probability and expected time until speciation is the following proposition, proved in the [Appendix](#):

Proposition 1.

$$E[X] \approx \frac{K^2}{2} r^2 \alpha$$

Note: With random sampling, as in [5], the result is that

$$E[X] = \binom{K}{2} \alpha \quad (10)$$

$$\approx \frac{K^2}{2} \alpha \quad (11)$$

which is the case when $r = 1$. In [7], both r and α equal 1 and $X = \binom{K}{2}$.

It is interesting to notice that, at least to a first-order approximation, the dependence of $E[X]$ on the network is light, in the sense that the network itself enters only via the density α and the sampling only via the parameter r . Thus, there are many different network topologies and many different sampling schemes (substitution mechanisms) that can lead to the same speciation rate.

Other than the usual snowballing – quadratic rather than linear dependence of $E[X]$ on K – we now also notice a quadratic dependence on r , which if $r < 1$ amounts to a “snowballing in reverse,” indicating that the speciation rate might be sensitive already to small reductions in r .

A formula for the probability of speciation (at least one incompatibility) after K substitutions now follows from [Proposition 1](#):

Proposition 2. Let p denote the probability of an incompatibility and let S denote the event of speciation after K substitutions. Then

$$P(S) \approx 1 - (1 - p)^{K^2 r^2 \alpha / 2}$$

To investigate the effect of the sampling scheme on the speciation rate, we compare $P(S)$ for a few different shapes of the $f(i)$. In general, let

$$f(i) = cr(i) \quad (12)$$

where r is a function determining the general shape of f and c is a normalizing constant. By (7)

$$c = \frac{1}{N \sum_i q(i)r(i)} \quad (13)$$

which by (8) gives the expected degree as

$$E[D] = \frac{\sum_i iq(i)r(i)}{\sum_i q(i)r(i)} \quad (14)$$

whence we can compute the speciation probability by (9) and [Proposition 2](#).

As an example, we will again use the PPI network for *S. cerevisiae* and investigate 3 general shapes of a decreasing f : linear, polynomial, and exponential. As the yeast network appears to be well described by a power law (see [9]), we take

$$q(i) = ai^{-b}, \quad i = 1, \dots, N - 1 \quad (15)$$

where the parameter b gives the exact shape and a is a normalizing constant. Different values of b in the range 1.5–2.5 have been reported (see [1]) and as we are mainly concerned with differences

Download English Version:

<https://daneshyari.com/en/article/4499869>

Download Persian Version:

<https://daneshyari.com/article/4499869>

[Daneshyari.com](https://daneshyari.com)