1 **Short communication:**

2 **Local speciation can be incorporated into neutral theory of biodiversity**

3 Youhua Chen

4 Department of Renewable Resources, University of Alberta, Edmonton, AB, Canada, T6G 2H1

5 Email:haydi@126.com

6
7 **Abstract**

8 In this report, I incorporate local speciation into neutral theory of biodiversity to predict
9 species abundance distribution patterns. By fitting the local-speciation model to the abundance
10 data of tree species in Barro Colorado Island of Panama, the results showed that the new model
11 performed better than the original neutral model without local speciation. Moreover, the estimated
12 local speciation rate ($\nu' = 0.0107$) was found to be ten-fold larger than the estimated background
13 speciation rate ($\nu = 0.0011$), indicating that local speciation should not be ignored in neutral
14 theory. Given that species across different local habitats and landscapes may present
15 heterogeneous speciation rates, it is valuable to incorporate local speciation into neutral theory to
16 study spatiotemporal patterns of ecological assemblages.

17
18 **Key words:** neutral theory, biodiversity maintenance, speciation, statistical ecology, numerical
19 optimization

Introduction

Since Stephen Hubbell published his book (Hubbell, 2001), neutral theory of biodiversity has represented an important advance for modern community ecology. Neutral theory has been proposed to describe spatiotemporal patterns of ecological communities (Alonso and McKane, 2004) and has been successfully applied to describe species abundance distribution (SAD) patterns. In addition to numerical simulation originally employed in Hubbell (2001), analytical calculation formulas have been extensively developed over the past decade (Volkov et al., 2003, 2007; Etienne and Alonso, 2005; Etienne, 2007).

Three biogeographical processes, namely speciation, immigration and extinction, are major factors shaping insular diversity patterns (Whittaker et al., 2008). However, their relative importance is not equivalent. In most cases, immigration and extinction are the dominant mechanisms stemmed from island biogeography theory (MacArthur and Wilson, 1967; Chen, 2015a). By contrast, speciation is often assumed to be small and slow (Heaney, 2000). However, some other works found that local speciation can contribute substantially to island diversity (Losos, 2010). In particular, on large islands, ecologists believed that radiation or phylogenetic diversification could increase island diversity (Heaney, 2000; Losos and Schluter, 2000). This is because the opening of ecological opportunities (or empty niches) in these islands (Algar and Losos, 2011). Based on these empirical observations, local speciation has been integrated in island biogeography models (Chen and He, 2009; Chen, 2015a).

However, in the context of neutral theory of biodiversity, the role of local speciation is unclear. In addition to drift, background speciation, immigration and extinction (Volkov et al., 2003; Chen, 2015b), Can local speciation contribute to the prediction and interpretation of SAD patterns in

empirical data sets? To address this question, in the present study, I explicitly incorporate local speciation into the sampling formulas for neutral theory and test the fitting power of new sampling formulas on empirical SADs by comparing it to the original formula without the involvement of local speciation.

Methods and Materials

A neutral model with local speciation

In this section, I deduce the sampling formula for the neutral model with local speciation. The model has an origin from the original neutral model, the corresponding stochastic master equation of which is governed by the following birth-death process:

$$dP_n / dt = A_{n-1}P_{n-1} + C_{n+1}P_{n+1} - (A_n + C_n)P_n \quad (1)$$

Where P_n denotes the probability of a species at time t having abundance n . The coefficients A_n and C_n are the birth and death rates for the probability when the species has the abundance n . For the proposed local-speciation neutral model in the present study, they are calculated as follows,

$$\begin{cases} A_n = \frac{J-n}{J} \frac{n}{J-1} (1-m-v') + \frac{J-n}{J} m\omega \\ C_n = \frac{n}{J} v' + \frac{(J-n)n}{J(J-1)} (1-m-v') + \frac{n}{J} m(1-\omega) \end{cases} \quad (2)$$

Without further notation hereafter, J denotes the local community size. m is the regional immigration rate of species from the metacommunity. ω is the relative abundance of species in the metacommunity. The new parameter presented in equation (2) is the local speciation rate v' , which reflects the influence of local speciation on the stochastic birth-death process of the neutral model. When there is no local speciation (i.e., $v' = 0$), the transition probabilities in equation (2)

become identical to those presented in previous studies (Vallade and Houchmandzadeh, 2003; Volkov et al., 2003).

Here I further set the following quantities,

$$\begin{cases} r = \frac{J-1}{1-m-v'}m \\ s = \frac{J-1}{1-m-v'}v' \\ \lambda = \frac{J-1}{1-m-v'}m\omega \end{cases} \quad (3a)$$

Where r here represents the strength of regional dispersal, while s represents the influence of local speciation. Their relation is given by $s/r = v'/m$. This ratio thus depends on the relative strength of local speciation and regional immigration. Based on these definitions, it can be further seen that,

$$\begin{cases} \lambda = \omega r \\ s = r \frac{v'}{m} \end{cases} \quad (3b)$$

Then, by using the quantities in equations (2-3), equation (1) at equilibrium can be solved as,

$$P_n(\omega) = \binom{J}{n} \frac{(r\omega)_n (s+r(1-\omega))_{J-n}}{(s+r)_J} \quad (4)$$

This is the typical form of local sampling formula established previously (Vallade and Houchmandzadeh, 2003; Alonso and McKane, 2004; Etienne and Alonso, 2005). Equation (4) uses the notation for Stirling number as $(a)_n = a(a+1)\dots(a+n-1)$. So, the expected species with abundance n in the local sample is given by,

$$E[S_n | \theta, r, s, J] = \int_0^1 \binom{J}{n} \frac{(r\omega)_n (s+r(1-\omega))_{J-n}}{(s+r)_J} \times \theta \frac{(1-\omega)^{\theta-1}}{\omega} d\omega \quad (5)$$

Where θ denotes the fundamental biodiversity number (Hubbell, 2001; Alonso and McKane, 2004; Volkov et al., 2003). The background speciation rate v mentioned above is

integrated in this parameter as $\theta = 2J_M v$ (Hubbell, 2001; Alonso and McKane, 2004). Here J_M is the metacommunity size.

Alternatively, one can use an approximate sampling equation presented in Volkov et al. (2003, 2007), in which the probability of a species with abundance n is given by,

$$P(n) = \binom{J}{n} \frac{\Gamma(n+\lambda)}{\Gamma(1+\lambda)} \frac{\Gamma(J+r-\lambda+\frac{r}{m}v'-n)}{\Gamma(\frac{r}{m}v'+r-\lambda)} \frac{\Gamma(\frac{r}{m}v'+r)}{\Gamma(\frac{r}{m}v'+J+r)} \quad (6)$$

$$= \binom{J}{n} \frac{\Gamma(n+r\omega)}{\Gamma(1+r\omega)} \frac{\Gamma(J+r-r\omega+s-n)}{\Gamma(s+r-r\omega)} \frac{\Gamma(s+r)}{\Gamma(s+J+r)}$$

Where $\Gamma(\bullet)$ denotes the Gamma function as $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$. The second equality in (6) used the quantities presented in equation (3b).

The expected value of species with abundance n in the community is then calculated approximately as,

$$E[S_n | \theta, r, s, J]$$

$$= \theta \int_0^1 \binom{J}{n} \frac{\Gamma(n+r\omega)}{\Gamma(1+r\omega)} \frac{\Gamma(J+r-r\omega+s-n)}{\Gamma(s+r-r\omega)} \frac{\Gamma(s+r)}{\Gamma(s+J+r)} e^{-\omega\theta} d\omega \quad (7)$$

When $s=0$ (thus $v'=0$, no local speciation), the above equation (7) is identical to the formula presented in Volkov et al. (2003). The integral in this equation can be done numerically so to compute the expected number of species with abundance n .

97

Maximum likelihood method for estimating parameters

To fit the model to empirical data sets, I employ the maximum likelihood model used in the previous studies (Alonso and McKane, 2004; Etienne and Alonso, 2005) as,

$$L\{Data | \theta, r, s, J\} = p(1)^{S_1} \cdots p(a)^{S_a}$$

$$p(n) = \frac{E[S_n | \theta, r, s, J]}{\sum_{i=1}^J E[S_i | \theta, r, s, J]} \quad (8)$$

$$-Log(L\{Data | \theta, r, s, J\}) = -\sum_{i=1}^a S_i Log(p(i))$$

Here, $p(n)$ is the normalized probability of a species with n individuals.

$E[S_n | \theta, r, s, J]$ is the expected number of species with n individuals calculated from

equation (7). a is the maximum abundance found in the local community. S_k ($k = 1, 2, \dots, a$) is

the observed species number with abundance k in the empirical data set.

An empirical test

Now it is ready to test the local speciation model and compare it to the original neutral model

(i.e., the model without local speciation by setting $s=0$). Tree abundance data for Barro Colorado

Island (BCI) permanent forest plots in Panama

(<http://ctfs.arnarb.harvard.edu/webatlas/datasets/bci/>) were used as an empirical test as previous

studies (Alonso and McKane, 2004; Etienne, 2007). Because it is time-consuming to compute

Stirling number in equation (5), I estimated the parameters (θ , r , and s) using the maximum

likelihood model derived from the equation (7). The integral was calculated numerically using the

standard trapezoidal rule.

For comparing alternative models and choosing a better model to characterize SAD pattern of

BCI tree species, I utilize the modified Akaike information criterion (AICc) (Akaike, 1974;

Haining, 2003; Chen, 2013) accounted for small samples as follows,

$$AICc = -2(m-1)Log(L) + 2k \frac{m-1}{(m-1)-k-1} \quad (9)$$

Where $\text{Log}(L)$ is the log-transformed likelihood value as in equation (8), m is the number of data points used for maximizing the likelihood as above and k is the number of parameters presented in the model.

Results

The maximum likelihood estimation of parameters could be best achieved at $\theta = 47.6179$, $r = 1879.94$, and $s = 252.401$ (Table 1). This local speciation model had $\text{AICc} = 2612.126$. In contrast, for the model without local speciation (by mandatorily setting $s = 0$ in equation (7)), the fitted parameters became $\theta = 46.625$ and $r = 2288.86$, respectively (Table 1). These estimated values were very similar to those reported in the previous studies (Volkov et al., 2003; Etienne, 2007). The corresponding AICc for the model without local speciation was 2521.345 (Table 1).

One can see that, the introduction of local speciation parameter s did not influence the fitted value of the parameter θ , but changed the fitted value for r to some degree (Table 1). Moreover, the corresponding AICc implied that local speciation model performed better than the model without local speciation in fitting SAD bins for BCI tree species (Table 1). At last, the comparison on the fitted curves of both models showed that local speciation model fitted better for large-abundance classes in BCI tree SAD data set (Fig. 1).

On the basis of the fitted parameters and using the relations of $r = Jm/(1-m-v')$ and $s = Jv'/(1-m-v')$, one can obtain $v' = 0.0107$ and $m = 0.08$. As a comparison, the background speciation rate v derived from the relation of $\theta = 2Jv$ is 0.0011. Here J is used to approximate metacommunity size J_M , which should be much larger (thus v should be much

smaller). Therefore, local speciation rate v' was at least 10-fold larger than the background speciation rate v .

Discussion

There are many attempts at integrating different biological mechanisms into neutral models so as to further consolidate the theory. For example, some recent works (Rosindell et al., 2010; Etienne and Rosindell, 2012) introduced protracted speciation into the neutral theory. In another paper (Etienne and Haegeman, 2011), random fission speciation was incorporated into the theory as well for addressing the problems encountered in the point mutation model. Based on the empirical test and model comparison (Table 1 and Fig. 1), the present work goes a step further by incorporating local speciation into the neutral theory.

Normally, speciation in ecological communities is always assumed to be lower than the immigration rate (Heaney, 2000). However, as I showed here, introduction of local speciation has a better predictive power than the original model without local speciation (Table 1 and Fig. 1). Thus, integration of local speciation can be possible, especially when applied to communities with high radiation-driven endemic degree of species (Rosindell and Phillimore, 2011). Moreover, the new model could serve as a null model to examine empirical data sets whether local speciation should be involved as a possible mechanism. Based on the empirical test on BCI tree data, local speciation model was supported (Table 1) and local speciation rate was surprisingly high in the data: the local speciation rate v' could be 10-fold larger than the background speciation rate v .

In conclusion, the introduction of local speciation in neutral theory performed better than the original neutral model in fitting empirical SAD patterns. Speciation might be better understood as

a gradual, continuous and locally heterogeneous process in structuring ecological communities, as evidenced from the present study and preceding works (Rosindell et al., 2010; Rosindell and Phillimore, 2011; Etienne and Rosindell, 2012).

Acknowledgements

Smithsonian Tropical Research Institute is acknowledged for making the BCI data sets available to the public. The work is supported by the China Scholarship Council.

References

- Akaike, H., 1974. Information theory as an extension of the maximum likelihood principle, in: Petrov, B., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 276–281.
- Algar, A., Losos, J., 2011. Evolutionary assembly of island faunas reverses the classic island-mainland richness difference in *Anolis* lizards. *J. Biogeogr.* 38, 1125–1137.
- Alonso, D., McKane, A., 2004. Sampling Hubbell’s neutral theory of biodiversity. *Ecol. Lett.* 7, 901–910.
- Chen, X., He, F., 2009. Speciation and endemism under the model of island biogeography. *Ecology* 90, 39–45.
- Chen, Y., 2015a. An island biogeography model for beta diversity and endemism: the roles of speciation, extinction and dispersal. *Int. J. Biomath.* 8, 1550011.
- Chen, Y., 2015b. Impacts of dispersal limitation on temporal biodiversity patterns in a neutral model. *Ecol. Model.* 305, 22–28.
- Chen, Y., 2013. An autoregressive model for global vertebrate richness rankings: long-distance dispersers could have stronger spatial structures. *Zool. Stud.* 52, 57.
- Etienne, R., 2007. A neutral sampling formula for multiple samples and an “exact” test of neutrality. *Ecol. Lett.* 12, 608–618.
- Etienne, R., Alonso, D., 2005. A dispersal-limited sampling theory for species and alleles. *Ecol. Lett.* 8, 1147–1156.
- Etienne, R., Haegeman, B., 2011. The neutral theory of biodiversity with random fission speciation. *Theor. Ecol.* 4, 87–109.
- Etienne, R.S., Rosindell, J., 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Syst. Biol.* 61, 204–213.
- Haining, R., 2003. Spatial data analysis. Theory and Practice. Cambridge University Press, Cambridge.

- Heaney, L., 2000. Dynamic disequilibrium: a long-term, large-scale perspective on the equilibrium model of island biogeography. *Glob. Ecol. Biogeogr.* 9, 59–74.
- Hubbell, S.P., 2001. The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32) (Monographs in Population Biology). Princeton University Press.
- Losos, J., 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism. *Am. Nat.* 175, 623–639.
- Losos, J., Schluter, D., 2000. Analysis of an evolutionary species-area relationship. *Nature* 408, 847–850.
- MacArthur, R., Wilson, E., 1967. The Theory of Island Biogeography. Princeton University Press., Princeton, N.J.
- Rosindell, J., Cornell, S.J., Hubbell, S.P., Etienne, R.S., 2010. Protracted speciation revitalizes the neutral theory of biodiversity. *Ecol. Lett.* 13, 716–727.
- Rosindell, J., Phillimore, A., 2011. A unified model of island biogeography sheds light on the zone of radiation. *Ecol. Lett.* 14, 552–560.
- Vallade, M., Houchmandzadeh, B., 2003. Analytical solution of a neutral model of biodiversity. *Phys. Rev. E* 68, 061902.
- Volkov, I., Banavar, J., Hubbell, S., Maritan, A., 2007. Patterns of relative species abundance in rainforests and coral reefs. *Nature* 450, 45–49.
- Volkov, I., Banavar, J., Hubbell, S., Maritan, A., 2003. Neutral theory and relative species abundance in ecology. *Nature* 424, 1035–1037.
- Whittaker, R., Triantis, K., Ladle, R., 2008. A general dynamic theory of oceanic island biogeography. *J. Biogeogr.* 35, 977–994.

Tables and Figures

Table 1. Parameter comparison for the neutral models with versus without local speciation fitted on the BCI tree species abundance data set.

Models	Parameters estimated	AICc
model without local speciation (s is fixed to 0 in equation 7)	$\theta = 46.625$, $r = 2288.86$	2612.126
model with local speciation (s is unknown and should be estimated in equation 7)	$\theta = 47.6179$, $r = 1879.94$, $s = 252.401$	2521.345

Fig. 1. Fitted curves for the neutral models with (dashed red line) versus without (solid black line) local speciation on the BCI tree species abundance data set. Abundance classes are arranged following Preston's method as previous studies (Volkov et al., 2003, 2007).