From genetic neighbourhood to local community: Estimating a key parameter of the Unified Neutral Theory of Biodiversity

A partir del vecindario genético al vecindario de comunidad: Estimación de un parámetro clave de la Teoría Neutral Unificada de la biodiversidad

BARRY J. RICHARDSON1, * & ELIZABETH T. ARIAS-BOHART2

1CSIRO Ecosystem Sciences, GPO Box 1700, Canberra, ACT 2601, Australia
2University of California, Essig Museum of Entomology, 1101 Valley Life Sciences Bldg. #4780, Berkeley, CA 94720, California, USA

*Corresponding author: barry.richardson@csiro.au

ABSTRACT

Issues related to the size and nature of what is termed a “local community” in a continuous metacommunity under unified neutral theory are explored. Developments in the parallel field of neutral genetics theory allow the size of genetic neighbourhoods in the context of isolation by distance processes to be calculated. These suggest a way to estimate the geographical size of an equivalent form of local community in ecology. It is derived from the estimated spatial autocorrelation value at zero geographical distance and this is obtained using the distribution of the relative frequencies of species in field samples separated by varying geographical distance. A data set of 62000 beetles collected for other purposes, and consisting of beetle samples collected by fogging 52 trees in Nothofagus spp. forest and 24 trees in Araucaria forest distributed in southern Chile were used to demonstrate the proposed methods. Local communities had estimated effective community sizes ($J_e$) of ~12000 individuals in each forest type for predators and ~30000 for phytophages/xylophage beetle communities. These numbers can be used in available software to estimate the number of species and the distribution of specimens between species in a local community. The geographical size of a local community can be used in other ways. Firstly, because it is related to the distance moved between birth and breeding, it provides information on the geographical scale of community processes; it gives a measure of the scales over which speedy recolonization is possible and scales over which breaks in the habitat cannot be easily crossed. Secondly, replicate samples geographically separated by less than the area of a local community are not independent, but pseudoreplicate measures of the characteristics of the metacommunity.

Key words: Araucaria, beetle communities, effective community size, estimating parameter J, Nothofagus.

RESUMEN

Se exploran los problemas relacionados con el tamaño y la naturaleza de lo que se denomina una “comunidad local” en una metacomunidad continua en el contexto de la Teoría Neutral Unificada. Avances en el campo de la Teoría Neutral de la Genética permiten calcular el tamaño de un vecindario genético en el contexto de aislamiento a través de procesos de distancia, lo cual sugiere una manera de estimar los tamaños equivalentes de una comunidad local (el vecindario de la comunidad ecológica). Ellos se deriva estimando valores de autocorrelación espacial a la distancia geográfica cero, lo cual se obtiene usando la distribución de frecuencias relativas de las especies de las muestras de campos, separadas por distintas distancias geográficas. Para demostrar los métodos propuestos, consistentes muestras de coleópteros fueron colectadas con la fumigación de 52 árboles en bosques de Nothofagus y 24 árboles en bosques de Araucaria, los cuales estaban ampliamente distribuidos en el centro sur de Chile. También se obtuvo un set de datos de 62000 coleópteros para otros propósitos. El tamaño efectivo estimado de las comunidades de coleópteros locales para cada tipo de bosque es alrededor de 12000 individuos para depredadores y 30000 para fitófagos/xilófagos. Por otra parte, la extensión geográfica de la comunidad de un barrio puede ser utilizados de otras maneras. En primer lugar, porque se relaciona con la distancia recorrida entre el nacimiento y la crianza, lo que proporciona información sobre los procesos de la comunidad a escala geográfica; entrega medidas de escalas donde la recolonización rápida es posible y escalas en las cuales las sobre que barreras no son fáciles de pasar en un hábitat. En segundo lugar, las réplicas de muestras geográficamente separadas por el tamaño de la comunidad de un barrio, no son independientes, representando pseudoréplicas de la metacomunidad característica.

Palabras clave: Araucaria, comunidad de coleópteros, estructura de la comunidad, estimación del parámetro J, Nothofagus.
INTRODUCTION

Over the past decade explorations of the conceptual basis of community ecology have been enriched by the development of a theoretical model called the unified neutral theory of biodiversity and biogeography (hereafter, unified neutral theory). This seeks to predict the number of species and relative distribution of individuals between species in a community (Hubbell 2001, Chave 2004, Bell et al. 2006, Leigh 2007) assuming only stochasticity in the birth and death of individuals and the functional equivalence of species are relevant. Unified neutral theory as developed by Hubbell can be used to make mathematical predictions of the number of species and the shape of the species relative abundance curve of both the metacommunity and of a ‘local community’ within this metacommunity (Hubbell & Borda de Agua 2004).

In theory, predictions of the shape of the metacommunity curve can be made from knowledge of the underlying variables, the fundamental biodiversity number ($\theta$), migration rate (m) from the metacommunity into local communities, and community size ($J$; number of individuals). $\theta$ and m can be estimated using maximum likelihood methods and the observed species abundance curves (Etienne 2007, 2009). As well as examining the fit of the theoretical curve for the metacommunity to field data, it is possible to predict the shape of the curve for a local community (Hubbell & Borda de Agua 2004) from the shape of the curve for the metacommunity and the effective size of the local community, if this is known.

The origin of unified neutral theory is the mathematically equivalent theory in population genetics, the neutral theory of molecular evolution (Kimura 1983). Two main models of the geographical structure of a species’ local populations were used in this development: firstly, a metacommunity structure consisting of a series of interconnected but distinct local (‘island’) populations, and, secondly, a continuous population in which gene flow was restricted by isolation by distance processes. The island model was used by Hubbell (2001) in the development of his form of the unified neutral theory and will provide the framework used here. The isolation by distance model in a continuous community is also relevant to the study of structuring in widespread natural communities (Chave & Leigh 2002). Of critical interest is the effective size of a local community in both scenarios as it is an essential parameter, for example in modelling community abundance curves.

In genetics the issue of estimating the size of a local population was addressed by Wright (1946, 1969) and Malecot (1948), who developed the concept of a ‘genetic neighbourhood’. This was defined as the ‘population of a region in a continuum from which the parents of individuals born near the centre may be treated as if drawn at random’. That is, it is related to the distance individuals move between birth and breeding (in most animals, more generally, between the same point in the life cycle in consecutive generations). Wright showed that in a two dimensional continuum it was an area with a radius that could be measured as the variance of the distances moved by individuals between birth and breeding. When the size of such an area is multiplied by the density of individuals within it, the result is an estimate of $N_e$, the effective population size, a key concept in neutral genetics theory. These concepts, though developed in the field of population genetics, can be transferred to community ecology within the framework of unified neutral theory as the mathematical basis is the same (Hubbell 2001). In this context, $N_e$ would become $J_e$, following Hubbell’s terminology, the effective size of a local community within the metacommunity, and the geographical size of a ‘genetic neighbourhood’ becomes the ‘geographical size of a local community’.

Thus, in unified neutral theory, while $J_m$, the size of the metacommunity, remains unchanged, $J_e$, the size of a local community or patch under an island/metacommunity model is replaced by $J_e$, the effective size of a local community at a particular location in a continuous community. Mathematically, $J_e$ is equivalent to $N_e$ in the development of the original genetic theory (Hu et al. 2006), though ‘community size’ (actually sample size) has been used as if equivalent, but incorrectly, in field studies of continuous communities.
Measuring the size of a local community

Actually measuring the geographical size of a local community by measuring the variance of the distances moved by individuals of many species between birth and breeding is very difficult, if not impossible. Other approaches to measuring neighbourhood size have been developed in genetics (e.g., Broquet & Petit 2009, Guillot et al. 2009), and these can be transferred to community ecology as the mathematical assumptions are the same (Hu et al. 2006). One approach is based on the observation that the isolation by distance processes that underlie the structuring of communities result in geographically adjacent samples being more similar than would be expected if they were samples taken at random across the geographical range of the metacommunity. As a consequence, increased spatial autocorrelations will be found between samples of species as the geographical distance separating the samples is reduced.

While various ways of estimating the geographical size of a genetic neighbourhood from this relationship have been explored (Epperson & Li 1996, Epperson et al. 1999, Hardy & Vekemans 1999, Rousset 2000, 2008, Vekemans & Hardy 2004, Epperson 2005, 2007), recently an equation that predicts the size of a neighbourhood from the level of autocorrelation (as Morans I) in allele frequencies in the ‘zero’ geographical distance class has been developed (Epperson, 2005) by using the formula: ln (Nb) = 1.11-1.13 × ln (I).

The number of classes does not affect the I value for the zero class however the number of pairs and the geographical size of this class can radically affect the estimate of I and its CI. Zero class pairs should not be geographically separated by more than the diameter of a local community area. A pilot study using sampling points selected to give a series of logarithmically increasing distance classes may be necessary to give a first estimate the diameter of a local community.

The value of Morans I for the zero distance class was then used to calculate the geographical size of the local community as described by Rousset (2008, after Epperson, 2005) by using the formula: ln (Nb) = 1.11-1.13 × ln (I).

Where I is the value of Moran’s I in the zero distance class and Nb is the area of the local community. In the demonstration study, where the sampling unit was a logged tree, local community area would be ‘number of trees’ (actually the area covered by this many trees). In most studies, it would be ‘number of quadrats’ (actually the total area of this number of quadrats). Local community size (J) was then calculated as density (specimens per tree or quadrat) times local community area (number of trees or quadrats in the area).

Values for the fundamental biodiversity number (θ) and fundamental dispersal number (I) were estimated as described by Etienne (2007) from the data sets using maximum likelihood methods via the software provided as a supplement to Etienne (2007). This package assumes constant local population size and migration rates: more recent software packages take account of variation in these parameters (Etienne 2009). Using the combined sample size (N) and the value of I, m can be estimated from the formula: m = 1/(I × N-1) (Etienne 2007).

Given these estimated values for θ, m, and J, programs developed by Hubbell & Borda de Agua (2004), allow the number of species and the shape of the species abundance curve to be predicted for the metacommunity and thence for the local community. The
local community program also provides confidence limits for the estimates. It was assumed that 14,000 generations had passed since the last ice-age (Ashworth & Hoganson 1993), and the metacommunity size was 990,000,000 (the largest the software allowed, though this is a gross underestimate) and this, along with a presumed turnover of half of the individuals per year, were used in the modelling. Sensitivity analysis using the ± 95 % CI values for \( J_e \) and 10 % variation in metacommunity size showed only minor effects on the apparent distribution of individuals between species (results not shown). The predicted distribution of specimens between species was then compared with the field data.

RESULTS

The relationship between Morans I and geographical distance for predators in the Nothofagus beetle fauna is shown in Fig. 1. It can be seen that, as predicted, I is most positive in the first (0-1 km distance) class. A similar shaped curve (not shown) was found for each of the other three sample sets. The zero distance means and confidence values for I are given in Table 1, as are the estimates of the number of trees in a local community and the effective size of \( J_e \). The confidence intervals are wider in the Araucaria forest estimates because of the smaller number of pairs available. The numbers of trees in a local community are similar in all four cases, as are the estimated sizes of \( J_e \) for predators in the two forest types. Because of differences in density (specimens per tree), the number of specimens in a local community for phytophages/xylophages is lower in the Nothofagus forest than in Araucaria forest.

The estimated values for \( \theta \) and \( m \) for each community are given in Table 1. It can be seen that the value of \( \theta \) for the Araucaria communities is less than half that found for the Nothofagus communities whilst the estimates of migration rates from metacommunities into local communities are similar in the two forest types but, as expected (Arias et al. 2008), higher for predators than for phytophages and xylophages.

A comparison of the observed species abundance curve for the metacommunity for predators living on Nothofagus and the shape of the curve predicted from \( \theta \) shows they are similar (Fig. 2). The shape of the predicted local community species abundance curve was then calculated based on the shape of the predicted metacommunity curve and the estimated values of \( J_e \) and \( m \). These results are also shown in Fig. 2. It can be seen that the predicted structure of the local community

\[ \text{Fig. 1: Relationship between spatial autocorrelation (as Moran's I) and geographical distance for the predator beetle communities of Nothofagus forest. Mean and 95 \% C.I. shown. Information for the zero distance class is summarised in Table 1.} \]

Relación entre la autocorrelación especial (según Moran’s I) y distancia geográfica para la comunidad de coleópteros depredadores de los bosques de Nothofagus. Promedio y 95 \% C.I. Información para la distancia clase cero que se resume en la tabla 1.

\[ \text{Fig. 2: Fit of observed predator community data for Nothofagus forest to predicted metacommunity and local community species abundance curves. These were developed by modelling using the values for } \theta, m \text{ and } J_e \text{ given in Table 1 and the programs of Hubbell & Borda de Agua (2004).} \]

Ajuste de los datos observados de la comunidad de coleópteros depredadores para los bosques de Nothofagus, y de las curvas de abundancia de las comunidades locales. Estos fueron desarrollados con el modelado utilizando los valores de \( \theta \), y \( J_e \) m en la Tabla 1 y los programas de Hubbell & Borda de Agua (2004).
differs from that of the metacommunity and of the field data. Unfortunately the sampling sites, which were selected for other reasons, did not include a set of trees from within the same local community that would provide sufficient specimens to allow a species abundance curve for an actual local community to be estimated. It should be noted that, though the predicted metacommunity curve was similar to that for the field data, the predicted number of species was a five fold over-estimate (Richardson & Arias-Bohart 2011). Similar patterns were found in the results for all four analyses.

DISCUSSION

The method proposed here to estimate the geographical size of a local community works as expected and is straight-forward to apply. The estimates have narrow confidence limits (Table 1) allowing them to be used in modelling and practical management (see below). As methods of estimating $J_e$ as well as $\theta$ and $m$ are now available, it is possible, using the programs provided in Hubbell & Borda de Agua (2004), to calculate the metacommunity and local community species abundance curves predicted under unified neutral theory in a continuous

<table>
<thead>
<tr>
<th></th>
<th>Nothofagus forest</th>
<th>Araucaria forest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predators:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fundamental biodiversity number ($\theta$)</td>
<td>122.3</td>
<td>46.0</td>
</tr>
<tr>
<td>Fundamental dispersal Number ($I$)</td>
<td>7.68</td>
<td>6.21</td>
</tr>
<tr>
<td>Total individuals collected ($N$)</td>
<td>9 979</td>
<td>3801</td>
</tr>
<tr>
<td>Migration rate ($m$)</td>
<td>0.0008</td>
<td>0.0016</td>
</tr>
<tr>
<td>Moran’s I (0 distance) ± 95 % C.I.</td>
<td>0.0582</td>
<td>0.0616</td>
</tr>
<tr>
<td>Local community area (± 95 % C.I.) (as number of trees)</td>
<td>75 (60-101)</td>
<td>71 (41-205)</td>
</tr>
<tr>
<td>Beetles per tree</td>
<td>169</td>
<td>158</td>
</tr>
<tr>
<td>$J_e$ (Effective local community size ± 95 % C.I.) (as beetles /local community area)</td>
<td>12675 (10140-17069)</td>
<td>11218 (6478-32390)</td>
</tr>
<tr>
<td><strong>Phytophage and Xylophage:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fundamental biodiversity number ($\theta$)</td>
<td>99.4</td>
<td>31.37</td>
</tr>
<tr>
<td>Fundamental Dispersal Number ($I$)</td>
<td>9.18</td>
<td>7.89</td>
</tr>
<tr>
<td>Total individuals collected ($N$)</td>
<td>20579</td>
<td>11181</td>
</tr>
<tr>
<td>Migration rate ($m$)</td>
<td>0.0004</td>
<td>0.0007</td>
</tr>
<tr>
<td>Moran’s I (0 distance) ± 95 % C.I.</td>
<td>0.0615</td>
<td>0.0609</td>
</tr>
<tr>
<td>Local community area (± 95 % C.I.) (as number of trees)</td>
<td>71 (57-92)</td>
<td>72 (38-349)</td>
</tr>
<tr>
<td>Beetles per tree</td>
<td>349</td>
<td>465</td>
</tr>
<tr>
<td>$J_e$ (Effective local community size ± 95 % C.I., as beetles /local community area)</td>
<td>24779 (19893-32108)</td>
<td>33480 (17670-162285)</td>
</tr>
</tbody>
</table>
community (Fig. 2) and the number of species expected in each case. The methods used here therefore would allow the null predictions of the neutral model to be tested against field data. It is noteworthy that, though the theoretical species abundance curve was similar to that of the field data, the predicted number of species was grossly in error (modelled 1451, estimated using Chao 1362).

There are, however, further issues to be considered when comparing the predicted and observed curves. All the relative species abundance curves based on field data showed a very large excess of species represented by single specimens when compared to equivalent curves for most other groups (e.g., Hubbell 2001). Similar results have been found, however, in studies of large tropical arthropod faunas (e.g., Morse et al. 1988, Stork 1997). The causes of this anomaly relate to failures to meet the underlying assumptions made when collecting and interpreting such field data sets. These failures include: undersampling bias, community disequilibria, and combining data for species that are not ecologically equivalent (Coddington et al. 2009, Richardson & Arias 2011). It is likely, for example, that the sample from Nothofagus forest (already ~46000 individuals) would need to be five times larger before the number of species and the shape of the species accumulation curve stabilised (Richardson & Arias 2011). These issues would need to be taken into consideration when collecting data sets to test neutral theory predictions.

The use of spatial autocorrelations to calculate neighbourhood size in this way has been considered in the genetic literature and there are at least two further issues that need to be considered in designing such studies. Firstly other factors may affect the structure of a community with increasing geographical distance, for example clinal variations in rainfall or temperature. Such factors would increase the value of I in the zero distance class, reducing the predicted geographical size of a local community. Secondly, Rousset (1997) argues that there are theoretical reasons for restricting the geographical distances between the furthest apart samples in such an analysis to approximately twenty times the size of a neighbourhood. While the Chilean forests used might be considered to constitute single entities, the sample sets used in the present study to explore the methodology were collected for other reasons and were separated by distances of up to 600 km. The experimental design was not therefore suitable for other than a demonstration study and to obtain first estimates of local community size. In the future, a targeted study of such communities would need to be based on a series of samples collected over much shorter distances, distributed so as to give reasonable numbers of comparisons in each distance class, including the zero distance class and each of a sufficient size to remove the problem of undersampling (Coddington et al. 2009, Richardson & Arias 2011). It also will need to include sufficient material from a single local community to provide the curve necessary to test the validity of the species abundance curve predicted for a local community. The distribution of sample sites should also be chosen so as to reduce the effects of confounding variables like temperature, rainfall or soil type. The result of such a study would likely be larger local community sizes (i.e. lower values for I) for Chilean forests than those estimated here.

The geographical size of a local community has several attributes that can be conveniently used in practical ways unrelated to the estimate of species abundance curves. Firstly, because it is related to the distance moved between birth and breeding, it provides information on the geographical scale of community processes; it gives, for example, a measure of the scales over which speedy recolonization is possible and scales over which breaks in the habitat cannot be easily crossed. Such information may also be useful in deciding the size of areas that should be burnt in hazard reduction burns or in setting cutting coupe sizes and the gaps between them. Secondly, replicate samples geographically separated by less than the size of a local community, are not independent estimates of characteristics of the metacommunity but of a local community. That is, if samples are taken from within the same local community they are pseudorePLICATE measures of the characteristics of the metacommunity. If samples are taken further apart than the geographical size of a local community, they are not replicate measures of the characteristics of a particular local community. Similarly the distances separating samples when estimating
beta diversity need to take account of local community size.

The design of sampling regimes and the estimation of neighbourhood size from spatial structuring are matters actively being explored in genetics at present and the results will be relevant to the design of community studies. At this stage no clear conclusions have been reached, though, no doubt, matters will be clarified over the next few years (Broquet & Petit 2009, Guillot et al. 2009).

ACKNOWLEDGEMENTS: We would like to thank all those who assisted in the fogging, sorting and identification program in Chile. Brian Epperson and Olivier Hardy provided helpful advice on methods for analysing Moran’s I and John Busby, Saul Cunningham, and Simon Ferrier made useful suggestions. Christine Richardson prepared the figures. We are very grateful to Dr. Richard M. Bohart for financial support. This work was also supported, in part, by the National Science Foundation (Grant #0445413 to ET Arias and KW Will).

LITERATURE CITED


Associate Editor: Matías Arim
Received December 22, 20010; accepted October 22, 2011