

The neutral theory of biodiversity: do the numbers add up?

SEAN NEE

Ashworth Laboratories, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK

Introduction

In the 1960s the new technique of gel electrophoresis revealed what was then considered to be an astonishing amount of molecular variation in natural populations, with about 30% of genetic loci being polymorphic. It was thought at the time that natural selection – the dominant principle in evolutionary biology – could not account for such high levels of variability – the costs of selection would be too high. This prompted the proposal of the neutral theory of molecular evolution (Kimura 1968; King & Jukes 1969; Kimura 1983), which postulated that most molecular evolution did not involve natural selection at all: in this model, selectively neutral mutations arose and their frequencies simply fluctuated at random, as is inevitable in a finite population.

Tropical forests pose a similar puzzle for ecologists. The enormous numbers of species seem to challenge our classical notions that coexistence requires that each species has its own unique niche: how can so many species construct unique niches from such a small number of requirements – sun, water, a patch of ground? The same puzzle in a different context became known as the ‘paradox of the plankton’ (Hutchinson 1961). As in molecular evolution, so too in ecology a neutral theory has been proposed which may resolve our puzzle (Bell 2001; Hubbell 2001) and is attracting much attention. The theory’s advocates have more in mind than tropical trees, but I will restrict myself to trees to make the discussion specific. This was the context in which the theory was first proposed (Hubbell 1979) and one of the main biological areas where we have a puzzle apparently requiring a radical solution. Hereafter I will refer to the neutral theory of biodiversity as NTB.

The underlying stochastic theory is the same in the two areas, although it has been studied for much longer in population genetics, as neutral advocates point out (Hubbell 2001; Volkov *et al.* 2003). This means we can import many results from the population genetics literature, simply interpreting the biology in a different way. In this paper I will import an important result concerning the time-scale of the neutral process, which raises serious difficulties for the NTB as an explanation of tropical forest diversity. This difficulty has been noted before (Leigh 1999; Lande *et al.* 2003), but does not appear to be as widely known as it should. Nonetheless,

as I will discuss, neutral theory can still provide useful null models for the interpretation of data.

In the first section, I gather together some basic theoretical results that are common to both population genetics and NTB. This is for two reasons. First of all, it serves to emphasize that we really are talking about the same theory, giving us the confidence to bring into the body of NTB an important result from population genetics concerning the time-scale of the neutral process. This will be done in the third section. Secondly, people may find it useful to have these basic results gathered together in one place – they are currently somewhat scattered about. Having emphasized the underlying identity of the theory, I then briefly explore the implications of the biological differences between the worlds of molecular evolution and biodiversity for the likely future development of the theory in the biodiversity context.

BASIC THEORY

In NTB, all species are ecologically equivalent. Species arise at random when an individual ‘mutates’ to become a new species, a rare occurrence that occurs at a per-individual rate u . This is analogous to a genetic mutation creating a new allele in a population. Abundances of all species simply fluctuate at random, as individuals randomly die and are replaced by the offspring of another individual, chosen at random, regardless of species. This is analogous to genetic drift. There are N individuals of all species in total, constituting what is called the ‘metacommunity’ in NTB and the ‘population’ in population genetics. The metacommunity could be the tropical forest of Central America perhaps. NTB as visualized by Hubbell (2001) connects small, local communities to the metacommunity by migration: these may be thought of, for example, as the investigator’s study area. I am going to ignore this extra, lower layer of theory: see, for example (Volkov *et al.* 2003). Of course, there are many scenarios that one could study: Bell (2001), for example, explicitly models individuals diffusing over a landscape.

Pick two trees at random at generation t . Define f_t as the probability that they are identical: this is Wright’s coefficient of inbreeding in population genetics (Crow & Kimura 1970) if we were discussing alleles instead of trees. If they are the progeny of the same tree in the previous generation, which occurs with probability $1/N$ under NTB, then $f_t = 1$. If they are the progeny of

different trees, which has probability $1 - 1/N$, then the probability they are identical is simply f_{i-1} . Finally, in order to be identical, neither of them can be new species, which occurs with probability $(1 - u)^2$. Putting this all together we have (Crow & Kimura 1970; Hubbell 2001):

$$f_i = (1 - u)^2 \left(\frac{1}{N} + \left(1 - \frac{1}{N}\right) f_{i-1} \right). \quad \text{eqn 1}$$

At equilibrium, assuming u is small, we have

$$f \approx \frac{1}{2Nu + 1}, \quad \text{eqn 2}$$

and we recognize $2Nu$ as the ‘fundamental biodiversity number’ in NTB, usually denoted as θ in both population genetics and NTB. The fact that it arises at so many points in the theory should not be surprising as it quantifies the flux of new species or mutations into the metacommunity or population. The probability that two trees chosen at random will be different is:

$$1 - f \approx \frac{\theta}{\theta + 1}. \quad \text{eqn 3}$$

Clearly, it makes no sense to suppose $2Nu$ is small if one wishes this to be meaningful for tropical forests, so N must be enormous, as we would like to imagine. However, as we will soon see, this creates problems for NTB as an explanation of forest diversity.

But before going on to those, let us collect some further basic results and introduce an idea from population genetics that is new to NTB: the ‘effective number of species’, S_e (Crow & Kimura 1970). This is the number of species in the forest, S , discounting/devaluing those that are very rare, consisting of, for example, just one or two individuals. Let p_i be the relative abundance of the i th species. Then,

$$S_e = \frac{1}{\sum_1^S p_i^2} = \frac{1}{f} = 2Nu + 1. \quad \text{eqn 4}$$

If all species have the same relative abundances, then $S_e = S$, otherwise $S_e < S$. Clearly, S_e is Simpson’s diversity index and we have achieved a simple derivation of the predicted index under NTB even without knowing the actual distribution of abundances predicted by NTB. S_e has some other attractive theoretical features as well, meaningful for NTB, and these are discussed in the population genetics literature (Crow & Kimura 1970).

The derivation of the distribution of abundances will not be given here. The distribution is (Crow & Kimura 1970: Eqn 9.6.11):

$$\Phi(x) = \frac{\theta(1 - x)^{\theta-1}}{x}, \quad 0 < x < 1. \quad \text{eqn 5}$$

The interpretation of this quantity is as follows: for a very small increment, which we will denote by dx , the product $\Phi(x)dx$ is the number of tree species with relative abundance between x and $x + dx$.

Numerous results can be derived from this distribution. For example, Ewens (1972) showed that the expected number of species, $E(S, n)$, in a sample of size n individuals is given by the elegant formula:

$$E(S, n) = \theta \sum_{i=0}^{n-1} \frac{1}{\theta + i}. \quad \text{eqn 6}$$

See also (Kimura 1983; Hubbell 2001). This can be approximated by (Abramowitz & Stegun 1972: Section 6.3):

$$E(S, n) \approx \theta \ln \left(1 + \frac{n}{\theta} \right), \quad \text{eqn 7}$$

which is the expected number of species in a sample of size n drawn from a population with the log-series distribution (Leigh 1999; Hubbell 2001) with θ in the place of the parameter denoted as α in the ecological literature (e.g. Williams 1964). Distribution equation 5 is the continuous version of the log-series distribution, the former giving the relative abundances, between 0 and 1, and the latter giving the actual abundances, 1, 2, 3, etc.

Leigh (1999: Appendix 8.2), in a *tour de force* of concise synthesis, derives the log-series distribution for the NTB, in both discrete and continuous forms, as well as the sampling result, equation 6, all in a single page! Note: the fundamental biodiversity number in Leigh is Nu rather than $2Nu$. This is because Leigh is studying a model with overlapping generations, which might be considered more appropriate than the model of discrete, non-overlapping generations which is used to derive the actual mathematical results in NTB and much of population genetics.

As stated earlier I am only discussing the metacommunity. Explicit expressions for the patterns of relative abundance of species in a local community connected by migration to the wider world are given in Volkov *et al.* (2003).

As an aside, I note that the connection between the log-series distribution and neutral models, first given prominence by Hubbell (2001), does not seem to have been recognized in the population genetics literature or in the literature seeking a stochastic theory generating the log-series distribution (Kendall 1948; Boswell & Patil 1971). One would have thought that R. A. Fisher, who derived the distribution (Fisher *et al.* 1943), as well as being one of the creators of population genetics (Fisher 1958: first published in 1930), would have seen the connection. Perhaps the fact that by 1943 Fisher had long since ceased to work in population genetics, concentrating on inventing statistics instead, accounts for this.

TIME-SCALE OF NTB

It is clear that the neutral theory of molecular biology and NTB have the same mathematical structure. This is good, as we can plunder population genetics at will

for results that have accumulated over decades. Here we ask about the time-scale of NTB.

What is the average age, $a(p)$, of a species of tree that has increased from a very low abundance at the time of its origin by mutation/speciation and is currently found at relative abundance p ? On a time-scale of generations, this average age is given by (Kimura 1983: Eqn 8.35):

$$a(p) = -2N \left(\frac{p}{1-p} \right) \ln(p) \quad \text{eqn 8}$$

Because the process of NTB is symmetrical with respect to time (if you watched a film of the species abundances fluctuating randomly over time, how would you tell if the film was running backwards?) this is also the average time to extinction of a species at abundance p .

If a tree species has a relative abundance of 30% then $a(p) = N$ generations. If a tree generation is 30 years (Leigh *et al.* 1993), then the age of the species is $30N$ years. How many trees are there in the 'meta-community', e.g. Central America? If there are one billion, then the age of the species is 30 billion years! Even rare species are, on average, impossibly old: a species with a relative abundance of 1% dates back to the origin of the angiosperms in the Cretaceous.

So NTB has a serious problem with time-scales as pointed out by Lande *et al.* (2003) and Leigh (1999). In molecular evolution, the enormous time-scales involved mean that if you observe a neutral polymorphism in a species, it probably predates the origin of that species (Kimura 1983). As Lande *et al.* (2003: p. 166) put it, reversing the temporal perspective: 'the time scale for extinction of common species in a neutral community is on the order of N generations. For extremely abundant communities, such as oceanic plankton, tropical insects, and even tropical trees, this predicts that extinction of common species is not expected to occur within the age of the earth, whereas species observed in the fossil record become extinct within a few to several million years'.

One possible way out of this dilemma is to observe that the relevant N in equation 8 is not today's tree number, but the harmonic mean of the tree numbers over time and, when this number fluctuates, the harmonic mean is dominated by the low numbers (this result is from the population genetics calculation of the 'effective population size' – see any population genetics text book, such as Crow & Kimura (1970)). So perhaps we can invoke severe historical reduction in the size of the forest to solve the time scale problem.

This reasonable presumption has been invoked in molecular evolution as well, but in the context of explaining why some very abundant species such as fruit flies have such *low* levels of molecular diversity (Kimura 1983: Chapter 9). So, invoking this solution creates the new problem – why is there such high diversity in tropical forests, and we are right back where we began. Lande *et al.* (2003) suggest fixing this by postu-

lating elevated speciation rates during these periods of forest contraction into refugia which has, of course, been suggested (e.g. Huston 1994). But we are now wandering far from NTB and into the mechanics of the evolutionary process in our quest to understand tropical forest diversity.

Another possible way out is to retreat from a very pure NTB and allow the existence of many classes of trees, each class having its own niche (shade tolerant/drought resistant, etc.) but species dynamics *within* each class being governed by NTB. This would lower N , thus addressing the time-scale problem, and elevate diversity, albeit by restoring the classical niche framework of understanding. I do not know whether the numbers can be made to add up or not under this 'fix', but again we are retreating a long way from NTB as an explanation of the diversity.

I began by recalling that the neutral theory of molecular evolution was proposed when it was discovered that there is so much molecular diversity – polymorphism – in species. But we should also recall that many genetic loci are not polymorphic and those that are typically have few alleles (corresponding to 'species' in NTB). In humans, the closest genetic analogues for the enormous diversity of the tropical forest are the major histocompatibility complex (MHC) loci, which have hundreds of alleles. These code for proteins that present pathogens to the cells of the immune system. Although there is disagreement about the precise details (De Boer *et al.* 2004) it is clear that infectious diseases are providing selective advantages to rare variants at these loci. While I am not discussing alternatives to NTB for explaining tropical diversity, it would be perverse to ignore this striking parallel to the Janzen/Connell framework for understanding tropical diversity which postulates that, because of specialist pests and predators, rare species are at an advantage – see Leigh (1999) for a thorough discussion of the latter.

Discussion

Neutral theory, and its variants, plays an absolutely central role in understanding molecular evolution: it is not yet clear whether or not NTB will play as central a role in ecology (Nee & Stone 2003). Genomes are divided into a huge number of essentially independently evolving parts and we have enormous quantities of precise data. Neutral theory explains the rapid evolution of pseudogenes, third positions in codons, less constrained protein genes and much, much more. Stochastic models, such as those that have revolutionized the analysis of phylogenetic relationships (Hillis *et al.* 1996), do not actually have to postulate neutrality or, indeed, any particular biology. Probability models rarely do: the outbreak of war is certainly not a 'random' event, but the numbers of wars in a given year is very well described by the Poisson distribution (Nee 2003). Nonetheless, these models are most naturally wedded to neutral theory and have been stimulated by it.

In ecology we have the equivalent of a single genetic locus (trees) and limited data. Although I have been emphasizing here the unity of the stochastic theories in the two areas, the biological differences are important too, as we have just seen. Some tests of neutrality in molecular biology require comparisons both within and between species separated by suitable intervals. Even if the forests of Madagascar are to African forests as mice genomes are to human genomes, we do not have many such 'species' (forests to compare) in ecology. However, the biological differences may well work to the advantage of NTB in some cases. For example, the population genetics literature makes many predictions about gene genealogies (the family tree of alleles at a genetic locus) which cannot be tested directly because there is too little information to construct the genealogies. However, the analogue in NTB consists of the species' phylogenies, and these *can* be constructed.

Returning to the particular focus of this paper, even if we ultimately conclude that NTB is unsuitable as an explanation of tropical forest diversity, it will surely continue to be useful as a null model in appropriate contexts, although, as always, there are disagreements about choices of null models (Gotelli & Graves 1996; McGill 2003; Volkov *et al.* 2003). Its utility in this regard is well illustrated by a study exploiting one of the great natural experiments in ecology: the flooding of Lake Gatun in Panama. The Panama canal is only a 'canal' in parts, mostly it is a artificial lake. The flooding of this lake nearly a century ago turned several hill tops into islands. Leigh and colleagues (Leigh *et al.* 1993) used NTB as a null model to see if they needed to invoke biology to understand the changing vegetation composition on these islands. They did.

Acknowledgements

I am grateful to Bert Leigh for explaining to me why the fundamental biodiversity number is sometimes Nu and sometimes $2Nu$. Nick Colegrave and two anonymous referees made helpful comments on the manuscript.

References

Abramowitz, M. & Stegun, I.A. (1972) *Handbook of Mathematical Functions*. Dover, New York.
 Bell, G. (2001) Neutral macroecology. *Science* **293**, 2413–2417.
 Boswell, M.T. & Patil, G.P. (1971) Chance mechanisms generating the log arithmetic series distribution used in the analysis of number of species and individuals. *Statistical Ecology* (eds G.P. Patil, E.C. Pielou & W.E. Wates), pp. 99–130. Penn State University Press, Philadelphia, PA.

Crow, J.F. & Kimura, M. (1970) *An Introduction to Population Genetics Theory*. Harper & Row, New York.
 De Boer, R.J., Borghans, J.A.M., van Boven, M., Keesmis, C. & Weissing, F.J. (2004) Heterozygote advantage fails to explain the high degree of polymorphism of the MHC. *Immunogenetics* **55**, 725–731.
 Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoretical Population Biology* **3**, 87–112.
 Fisher, R.A. (1958) *The Genetical Theory of Natural Selection*. Dover, New York.
 Fisher, R.A., Corbet, S. & Williams, C.B. (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology* **1943**, 42–58.
 Gotelli, N.J. & Graves, G.R. (1996) *Null Models in Ecology*. Smithsonian Institution Press, Washington, DC.
 Hillis, D.M., Moritz, C. & Mable, B.K. (1996) *Molecular Systematics*. Sinauer, Sunderland, MA.
 Hubbell, S.P. (1979) Tree dispersion, abundance and diversity in a tropical dry forest. *Science* **203**, 1299–1309.
 Hubbell, S.P. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ.
 Huston, M.A. (1994) *Biology Diversity*. Cambridge University Press, Cambridge.
 Hutchinson, G.E. (1961) The paradox of the plankton. *American Naturalist* **95**, 137–145.
 Kendall, D.G. (1948) On some modes of population growth leading to R.A. Fisher's logarithmic series distribution. *Biometrika* **35**, 6–15.
 Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature* **217**, 624–626.
 Kimura, M. (1983) *The Neutral Allele Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
 King, J.L. & Jukes, T.H. (1969) Non-Darwinian evolution. *Science* **164**, 788–798.
 Lande, R., Engen, S. & Saether, B.E. (2003) *Stochastic Populations Dynamics in Ecology and Conservation*. Oxford University Press, Oxford.
 Leigh, E.G. (1999) *Tropical Forest Ecology*. Oxford University Press, Oxford.
 Leigh, E.G., Wright, S.J., Herre, E.A. & Putz, F.E. (1993) The decline of tree diversity on newly isolated tropical islands – a test of a null hypothesis and some implications. *Evolutionary Ecology* **7**, 76–102.
 McGill, B.J. (2003) A test of the unified neutral theory. *Nature* **422**, 881–885.
 Nee, S. (2003) The unified phenomenological theory of biodiversity. *Macroecology: Concepts and Consequences* (eds K.J. Gaston & T.M. Blackburn). Blackwell Publishing, Oxford.
 Nee, S. & Stone, G. (2003) The end of the beginning for neutral theory. *Trends in Ecology and Evolution* **18**, 433–434.
 Volkov, I., Banavar, J.R., Hubbell, S.P. & Maritan, A. (2003) Neutral theory and relative species abundance in ecology. *Nature* **424**, 1035–1038.
 Williams, C.B. (1964) *Patterns in the Balance of Nature*. Academic Press, London.

Received 7 April 2004; revised 7 July 2004; accepted 21 July 2004