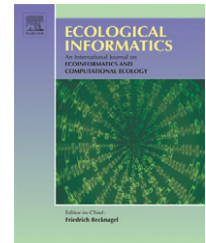


available at www.sciencedirect.comwww.elsevier.com/locate/ecolinf

A coalescence approach to spatial neutral ecology

James Rosindell^{a,*}, Yan Wong^{a,1}, Rampal S. Etienne^{b,2}

^aInstitute of Integrative and Comparative Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK

^bCommunity and Conservation Ecology Group, University of Groningen, Box 14 9750 AA, Haren, The Netherlands

ARTICLE INFO

Article history:

Received 16 January 2008

Received in revised form

19 April 2008

Accepted 4 May 2008

Keywords:

Coalescence

Neutral

Spatial ecology

Simulation models

ABSTRACT

Neutral models in ecology have attracted much attention in recent literature. They can provide considerable insight into the roles of non-species-specific factors (e.g. stochasticity, dispersal, speciation) on community dynamics but often require intensive simulations, particularly in spatial settings. Here, we clearly explain existing techniques for modelling spatially explicit neutral processes in ecology using coalescence. Furthermore, we present several novel extensions to these methods including procedures for dealing with system boundaries which enable improved investigation of the effects of dispersal. We also present a semi-analytical algorithm that calculates the expected species richness in a sample, for any speciation rate. By eliminating the effect of stochasticity in the speciation process, we reduce the variance in estimates of species richness. Our benchmarks show that the combination of existing coalescence theory and our extensions produces higher quality results in vastly shorter time scales than previously possible: years of simulation time are reduced to minutes. As an example application, we find parameters for a spatially explicit neutral model to approximate the species richness of a tropical forest dataset.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Advances in the analysis of neutral evolutionary processes have taken advantage of fast algorithms based on coalescence techniques (Felsenstein, 2004; Kingman, 1982; Wakeley, 2006; Notohara, 1990). These are primarily used to model the evolution of DNA sequences but can be applied more generally to asexually reproducing individuals, words and other patterns resulting from a process of lineage branching (Blythe and McKane, 2007). Hubbell's "Unified Neutral Theory" (2001) was formulated as an ecological analogue to these evolutionary processes, and stimulated much debate (Alonso et al., 2006; Gewin, 2006; Leigh, 2007; Holyoak and Loreau, 2006; Hubbell, 2005) by making the assumption that every individual behaves in the same way regardless of its species. Recent studies of

neutral ecological processes have exploited coalescence techniques, both analytically (Etienne, 2007; Etienne and Olf, 2004; Etienne, 2005) and computationally (Chave and Leigh, 2002; Chave et al., 2002), but no detailed explanation has yet been provided of the implementation and benefits for ecological simulations. We aim to both enable, promote and further develop the use of these methods in ecology.

In this paper, we explain the precise details of the coalescence method of simulation in ecology and provide several novel and powerful extensions. We use quantitative benchmarks to show that unmanageably large simulations for traditional methods are rendered straightforward with our approach. The exact methods we describe have already led to advances in the understanding of species–area curves from neutral models (Rosindell and Cornell, 2007). We illustrate our

* Corresponding author. Tel.: +44 7968 278 436.

E-mail addresses: bjyjl@leeds.ac.uk (J. Rosindell), yan@pixie.org.uk (Y. Wong), r.s.etienne@rug.nl (R.S. Etienne).

¹ Tel.: +44 1865 726111.

² Tel.: +31 503 632 230.

algorithms by finding the manifold in parameter space that gives rise to an empirically observed species richness, an analysis which would be impossible using traditional methods.

2. Neutral models in ecology

Neutral models in ecology make the controversial assumption of ecological equivalence among individuals regardless of their species identity. Despite the evident falsity of this assumption, they have done remarkably well at reproducing empirical data. Neutral models can, therefore, provide insight into the roles of non-species-specific (or trait-neutral) factors such as stochasticity, dispersal or speciation on community dynamics. It is clear that non-neutral processes exist in ecosystems, but neutral theory helps us to ascertain whether these asymmetries strongly influence the bulk patterns of biodiversity. A second use of neutral models is as a null model; a sound foundation on which to build and test more complex ideas. Neutral models centre around mechanistic birth and death processes in a population of individuals, which is a significant improvement on earlier null models that were just statistical distributions whose parameters were meaningless (Hubbell, 2001).

Hubbell's (2001) original formulation of a neutral model, and most recent variations (Chave and Leigh, 2002; Chave et al., 2002; Rangel and Diniz, 2005), make the zero-sum assumption that births and deaths balance, so that the total number of individuals in a community is fixed. While this assumption is not essential (Etienne et al., 2007a) it does lead to convenient computational techniques. The population genetics literature distinguishes two simple zero-sum birth–death models. The “Wright–Fisher” process proceeds in uniform time steps, in which all individuals simultaneously die and are replaced by a selection of their offspring (chosen by some stochastic process). This is commonly used in neutral genetic models and is best suited to populations with non-overlapping generations, such as annuals. Alternatively, in the “Moran” process (Moran, 1962), each uniform time step involves only a single individual dying and being replaced. This is more appropriate when generations overlap in time, for example in the dynamics of tropical tree communities. The equilibrium abundance distributions are in fact identical for both processes, but the temporal dynamics are different (Blythe and McKane, 2007; Etienne and Alonso, 2007). Although both processes can be modelled using coalescence (Wakeley, 2006), for simplicity in this paper we assume a Moran process throughout. Speciation is another important component of most neutral ecological models: as is conventional, we combine this with the birth–death process via a fixed probability of speciation ν that an offspring represents a new species.

Ecological communities are strongly influenced by spatial structure, and this too should be reflected in neutral models. Hence a major advance was the development of spatially implicit models such as the “local community model” (Hubbell, 2001) where a local community is linked to a regional “meta-community” by dispersal, and local individuals and immigrants have different chances of recruiting offspring. More complex spatial structures have also been studied for island systems (Ulrich and Zalewski, 2007). However, such models remain unrealistic for the common cases where there are no clear

boundaries which define local communities. For this reason, spatially explicit versions are important. These require an explicit dispersal kernel, for example given by fat-tailed (Chave and Leigh, 2002; Chave et al., 2002), Gaussian or “square” (Rosindell and Cornell, 2007) distributions. The methods we describe are suitable for any of these spatial structures, but are particularly helpful in the spatially explicit case.

Ecological neutral models are usually simulated from an arbitrary starting distribution, by repeatedly applying the rules of the model until the system reaches a steady-state (Hubbell, 2001). In common with simulations from many fields of research, there are several difficulties with this forward approach: we are never certain when the steady-state has been reached, and boundary conditions in the spatially explicit case can cause unwanted bias in results (Hubbell, 2001). The traditional solutions to this are computationally wasteful: simulations are run for much longer than strictly necessary to ensure a steady-state, and to reduce the influence of boundary conditions results are only taken from a tiny central sample of a much larger simulated population (Hubbell, 2001). A considerable amount of effort is expended in computations that are discarded, because they involve individuals either outside the sample area or in the distant past where they cannot influence the final result. Our simulation algorithm, based on coalescence, does not suffer from these problems: it derives results for the exact limit of infinite time equilibrium, and can be used to simulate both finite and infinite landscape areas.

The standard coalescence approach takes a sample of members from a larger population, and is retrospective rather than prospective. Instead of proceeding forwards in time, it goes backwards in time, tracing the lineages of all sampled individuals until they either speciate or coalesce. In our spatially explicit models, the sample consists of all individuals occupying a “survey area”. This survey area should not be confused with the larger landscape area which, ecologically, represents the complete community, from which the survey has been sampled.

As we trace time backwards, lineages coalesce. This means that forwards in time, coalescence is an instance where one lineage divides into two branches. For this to be feasible, the process that updates any one lineage as time passes must not depend on the other lineages. This is stronger than the Markov property (where the process must depend only on the current state of the system, and not on the state at previous times); fortunately many neutral processes meet this requirement.

3. Basic concepts of coalescence with worked example

For simplicity we illustrate the process with a miniature yet spatially explicit example in one dimension. Imagine a population consisting of eight individuals on a line (a landscape area of eight), we wish to simulate the equilibrium species abundance distribution of three adjacent individuals (a survey area of three).

Fig. 1a shows one possible outcome of a forward simulation of this starting at generation $t = 10$. The eight individuals in the population are each present in cells labelled A–H, with one individual in each cell. In each generation, a randomly chosen individual dies and is replaced either by a new species or by the

offspring of a neighbouring individual (nearest neighbour dispersal). By tracing the ancestral lineages of the three individuals in the sample area at time t , it is possible to identify the species to which they belong. The individuals in cells D and E are the same species as they share a common ancestor in generation $t-7$, whereas the individual in cell F is a distinct species. In this way, we can use the ancestry of individuals to reconstruct the species abundance distribution of the sample at equilibrium.

We now show how coalescence simulations produce equivalent results. Consider the individuals in D, E, and F at time t , we wish to discover their species. Each of these individuals occupies a cell that was left empty by a previous death. Working backwards, we must decide in which of D, E, or F the most recent death occurred. Under neutrality, this does not depend on the species previously present in the cell so, as with the forwards method, we choose any of the three cells with equal probability. In this case cell D was chosen as the position of the most recent death. We now need to discover the parentage of the individual that reoccupies cell D. This new individual either appeared as a result of dispersal from neighbouring cells (probability $1-\nu$) or is a new species (probability ν). The neutrality assumption means that we can reverse the dispersal process, using the same dispersal kernel to work out the parent. The dispersal kernel must be independent of species identity because the species of each individual is not known at this point. This would never be required in a neutral model because no species-

specific traits will exist by definition. In our case, the individual in cell C is randomly chosen as the parent of the individual in cell D, giving the same result as in Fig. 1a.

Continuing the simulation backwards, we can now ignore cell D, following only the lineages in cells C, E, and F. Again, a previous death may have occurred in any with equal probability. In this case, it occurs in cell E, the individual in this cell being born from cell D. We now follow cells C, D, and F. The next most recent death event in C, D, or F, randomly occurs in cell F. In this case, speciation occurs (probability ν) with the reoccupying individual being of a new type to the simulation (although we refer to this as speciation it could equally be interpreted as immigration (Etienne et al., 2007a,b)).

After speciation, the number of lineages we need to trace decreases by one: we now only follow the individuals in cells C and D. We allocate the next most recent death at random to cell D. On this occasion speciation does not occur, and the individual in cell D is born from a parent in cell C. From this we deduce that the two lineages have a common ancestor, a coalescence event. Hence the individuals in our original sample cells D and E are of the same species and the simulation is complete. We have produced identical results to those from the forward simulations (Fig. 1a), with the common ancestry of individuals in the system represented by multi-branched “coalescence trees”. The same principles apply to larger, more realistic simulations in higher dimensions and with more complex dispersal processes

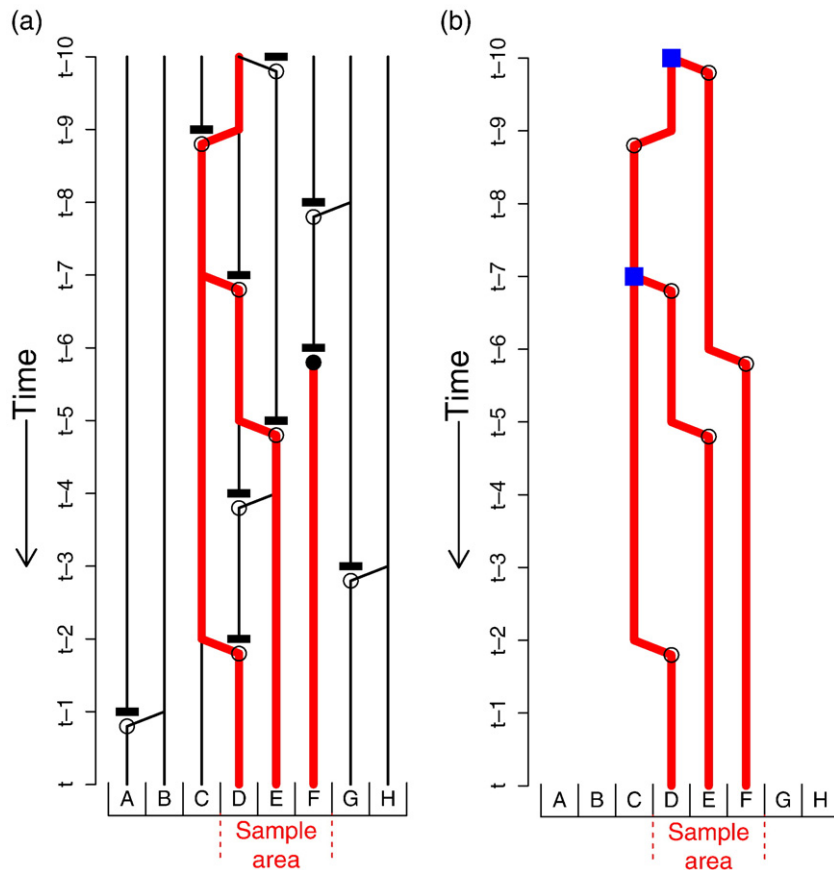


Fig. 1 – Panel (a) shows Schematic of a forwards simulation. The ancestry of the individuals in the sample area is marked with ticker lines. Death events are marked as bars and birth events are marked as circles. If a birth is a speciation event, then the corresponding circle is solid, otherwise we use open circles. Panel (b) shows the coalescent tree without speciation. Filled squares represent coalescence events and open circles represent births, the same notation is used in the Appendix.

(Chave and Leigh, 2002; Chave et al., 2002). In two dimensions the process is analogous to simulating a number of particles on a grid whose positions define the survey area (Chave and Leigh, 2002; Chave et al., 2002; Felsenstein, 2004; Holley and Liggett, 1975). Each particle in this simulation represents a lineage in the ecological model. Each time step, a random particle (lineage) is chosen to take a single step of a random walk (death/rebirth elsewhere). The direction and distance of this step is decided by the dispersal kernel. When particles collide (coalescence), they aggregate. With each move there is a small chance of an aggregate being removed from the simulation (speciation). Once all aggregates have been removed, the simulation terminates and the distribution of aggregate sizes gives the species abundance distribution.

4. Simulation efficiency

The coalescence method displays greatly improved efficiency over the forward algorithm. In a forward simulation, every time step is associated with a birth–death event, represented in Fig. 1a by a pair of symbols, a bar and a circle. In any reasonable implementation, the majority of the computational burden is associated with these events. Using the coalescence approach means that many of these birth–death events need not be considered for a number of reasons.

Firstly, the coalescence method means we need only consider individuals within our survey area at time t . In contrast, the forward method tracks lineages outside the survey area, because they may later enter the survey area. This enables us to omit calculations associated with time steps $t-1$ and $t-3$ and also saves significantly on computer memory. Secondly, we do not need to track lineages later doomed for extinction. Forward simulations provide no way of knowing which lineages will later become extinct. In our example, this allows us to omit time step $t-4$. Thirdly, with the forward method, it is unclear how many time steps are required to reach the steady-state. With the backwards method we know that we can stop once a single lineage is left in the simulation. Time steps $t-8$, $t-9$ and $t-10$ in our diagram represent the undefined additional effort required to ensure a steady-state is reached. This third advantage of coalescence extends beyond computational efficiency and affects the accuracy of the results themselves. In the forward method, attempts to cut the simulation time could easily result in data being collected before the steady-state is reached. We have observed this by a statistical comparison of our own forward and coalescence simulations, especially where the speciation rates are small.

Thus far, we have explained in detail the specifics of existing methods that have previously not been fully documented in the ecological literature. In the following sections we describe entirely novel extensions that lead to substantial further advantages and applications.

5. Infinite landscape areas

Finite landscape areas — either with periodic, reflective, or absorbing boundaries — can be simulated using either standard coalescence methods or forwards methods. For many ecological applications, however, it is preferable to

eschew boundary conditions altogether by extending the landscape area to infinity (Hubbell, 2001; Rosindell and Cornell, 2007). Infinite landscape areas can only be simulated using the novel method that we now describe and this yields qualitatively different results from finite landscape areas (Rosindell and Cornell, 2007). Furthermore tropical forest data are usually sampled from a large uninterrupted landscape with no clear boundaries, a situation well approximated by infinite area simulations. These are impossible using forwards methods, as this would require us to simulate a landscape area of infinite size. However, it is possible using coalescence methods, which only require us to track the finite number of lineages originating in the survey area.

Even with coalescence, there is the problematic issue that we need to keep track of the current position of lineages in an infinite space in order to identify whether lineages coalesce. The simplest implementation, where every cell in the landscape has an allocated memory space, is clearly impossible in the case of an infinite landscape. A natural solution might record the precise position of every individual in space. At every time step, this requires us to take the position of the parent of the newly born individual and check it sequentially against the positions of all other individuals in the simulation. With simulations involving many lineages, these comparisons are computationally expensive, but we show that they can be substantially reduced by using an indexing system. To index individuals, we store them in a number of slots according to their position in space. Checks for coalescence then only involve individuals in one slot. For example, we allocate a 10×10 grid of slots for indexing, storing individuals in the slot given by the remainder after division of their x and y positions by 10. Thus individuals at grid points $\{56,32\}$ and $\{167,49\}$ are stored in slots $\{6,2\}$ and $\{7,9\}$ respectively. Note that we still retain the exact positional information but have reduced the number of coalescence checks required. If we calculate that a new individual was born from a parent at for example, location $\{17,89\}$, we need only check for coalescence with individuals stored in slot $\{7,9\}$. Using a 10×10 index means that on average 99% fewer coalescence checks are required. A useful feature of such an indexing system is that we can trade off memory against computational efficiency. For example, we could use a 1000×1000 grid of slots, consuming more memory but giving greater efficiency. We quantify the benefits this indexing system in Section 7.

An additional saving in memory use can be gained by replacing the true position of an individual by the quotient from the division. In our 10×10 example above, the individual at $\{167,49\}$ would be stored in slot $s=\{7,9\}$ with position attribute $a=\{16,4\}$. An individual at $\{-43,129\}$ would be stored in slot $s=\{7,9\}$ with attribute $a=\{-5,12\}$. Coalescence occurs only when a parent shares both the same slot and attribute as a pre-existing individual. The true position, if required, can be recovered by calculating $G \times a + s$ for a $G \times G$ grid.

With an infinite area, there is always the possibility that the simulation becomes computationally impossible. For example, if any individual is at distance 2^x from another, where x is greater than the total number of memory bits available, then this distance can not be recorded exactly. Fortunately, if an individual is this far from any other then it is extremely unlikely that its lineage will coalesce before speciation occurs, and its exact location becomes irrelevant.

These methods can equally be applied to efficient simulation of large but finite areas. It is always the survey area that defines the number of lineages under consideration and hence the computational cost of simulation.

6. Speciation spectra

The speciation rate (ν) is an important parameter in neutral models (Hubbell, 2001). The coalescence approach opens the door to a novel, elegant method for investigating different values of ν in a semi-analytical manner. This makes it possible to simulate an entire “spectrum” of speciation rates with negligible extra effort over that required for a single value of ν . We treat speciation in a similar way to that in which mutation is treated in genetic coalescence (Wakeley, 2006). We produce a “coalescence tree” storing the ancestry of all individuals of interest. In standard applications of coalescence with mutations or speciations, these events are scattered stochastically onto the completed tree. Fig. 1b shows this for our illustrated simulation. The original backwards simulation (highlighted in Fig. 1a) can be recovered from this by placing a speciation event on the rightmost lineage at time $t-6$. We can similarly produce a result for any speciation rate ν by placing speciation events onto the bare tree, with probability ν at each birth. The distribution of species follows from the observation that individuals on the tree connected by a speciation-free path are conspecific. For this process to be possible, speciation must not affect the branching process of the tree.

The result of scattering speciation events onto the tree is stochastic. Our semi-analytical method eliminates this stochastic component of the simulation thus substantially reducing varia-

tion in simulation results. We do so by calculating the expected species richness (S^*) of a sample, given a coalescent tree and a speciation rate. To obtain an accurate overall estimate, we simulate multiple trees and take the average expected richness. Throughout the section we encourage the reader to refer to Appendix A where the algorithm is fully explained.

We methodically consider each part of the tree and at each step, refine an interval $[S_{min}, S_{max}]$ within which the expectation S^* must lie. This interval is initialised as $[0, N]$ where N is the total size of the sample. The simplest way to understand the algorithm is to envisage the tree being “pruned”, with both branches (edges) and coalescence events (nodes) being methodically considered and discarded. As they are pruned, S_{min} and S_{max} are updated, and they gradually converge to S^* . In all cases, coalescence involves both a coalescent point and a birth event, but to cleanly separate branches from nodes, we place the birth event on a branch and consider the coalescent point as a node. We start at the bottom of the tree and work upwards, first pruning all “free” branches (those without a node on the bottom). Then we prune any free nodes (those which have had both descendant branches pruned). This produces more free branches, which can themselves be pruned, and so on.

We attach to each branch (i) a probability u_i that the species richness would change given speciation on branch i . For all initially free branches we set this probability to one as speciation always results in a new species. The probability p_i of having any speciation events on the branch i is given by $p_i = 1 - (1 - \nu)^{b_i}$ where b_i is the number of birth events on branch i . Pruning a branch i from the tree also produces a new value $u'_i = u_i(1 - p_i)$ which measures the probability that speciation further up this lineage will change the species richness. Every time we prune a branch,

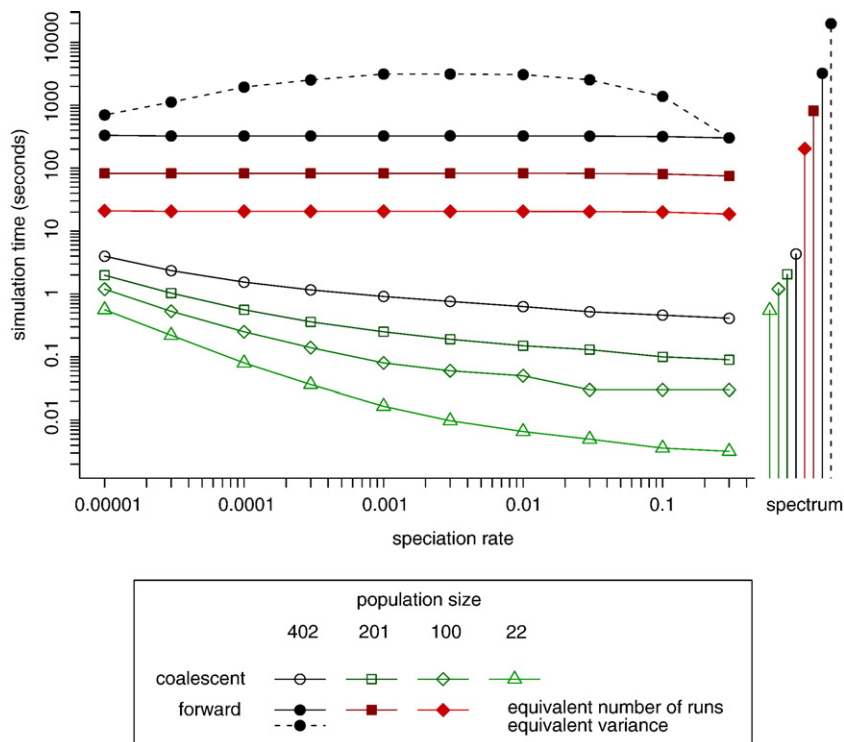


Fig. 2 – Benchmark results (time on a log scale). Simulations were run on a 2 GHz Intel dual processor computer under Linux.

we update S_{\min} with $S_{\min} \leftarrow S_{\min} + p_i u_i$. A node marks the coalescence of two lineages and hence the end of two branches, i and j and the start of a single new combined branch k . To prune a node, we calculate u_k from u_i and u_j with $u_k = 1 - (1 - u_i)(1 - u_j)$ and update S_{\max} with $S_{\max} \leftarrow S_{\max} - u_i u_j$. Once the entire tree is pruned in this manner, S^* is revealed as being in the interval $[S_{\max}, S_{\min}]$.

A drawback to the method thus far described is that constructing the ancestral tree back to a single lineage is often extremely costly, especially when there are large dispersal distances involved. When a small number of lineages remain, they are often far apart in space and a large number of iterations may be required before they reach the same grid cell and coalesce. This cost can be substantially reduced by deciding at the time of tree construction, on a minimum speciation rate that we are interested in and a tolerance for our data. We then calculate $[S_{\min}, S_{\max}]$ in parallel with the simulation of the coalescent tree, keeping track of the error $\varepsilon = S_{\max} - S_{\min}$. Calculating $[S_{\min}, S_{\max}]$ while carrying out the coalescent simulation means that we do not precisely follow the algorithm described above. Instead of pruning all branches followed by all nodes, we keep the tree constantly pruned. That is, every time a birth event occurs, we prune the branch back down to zero, recalculating u_i and thus S_{\min} . Every time a coalescence occurs, we immediately prune the node, and recalculate S_{\max} . We continue to simulate the coalescent tree until ε is sufficiently small. Convergence is guaranteed as detailed in [Appendix A](#) and flow charts for the entire process are shown in [Fig. A2](#) and [Fig. A3](#). If we wish to calculate the effect of a number of speciation rates using the same tree, then ε should be chosen to give the required tolerance for the smallest speciation rate.

7. Benchmarks

Two independent implementations were written, one using our coalescence methods and the other using traditional forward methods. We aim to quantify the benefits of our methods by making comparisons for Hubbell's patch model (2001). The forward method used the same parameters as Hubbell (2001): a grid of 402×402 cells grouped into 201×201 patches each of size 2×2 simulated for 5000 turnovers (generations) of the community, with dispersal parameter of 0.5. We also included smaller simulations of 201×201 and 100×100 cells, to show how size influences the simulation time.

Our benchmarks covered a range of ten different speciation rates, from $\nu = 0.3$ to $\nu = 0.00001$. The results are presented in [Fig. 2](#): lines give simulation time (log scale) as a function of speciation rate. We are also interested in the total time taken to simulate the "spectrum" of 10 different speciation rates. This total should be considerably faster for our coalescent method (where the same simulated trees can be reused) than the forward method (where new simulations are needed for each rate). These results are shown as vertical lines on the right side.

Comparison between results for equivalent sample areas (same point shape) shows that the basic coalescence algorithm is clearly more efficient than the traditional forwards method. The difference is considerably greater when spectra are calculated. It can be seen that the time taken for coalescent simulations (open points) decreases for higher speciation

rates. This reflects the fact that the coalescent simulations terminate once the desired tolerance is reached, which occurs more rapidly for higher speciation rates. The forwards method (filled points), which uses a fixed number of generations (5000 (Hubbell, 2001)), takes approximately the same amount of time regardless of the speciation rate. This difference reinforces two problems with the forwards approach: for low speciation rates steady-state may not have been reached, and for high ones calculation time has been wasted. A further benefit of using the analytical method is a reduced variance. We compared the variances of the forward and coalescence methods, and used the Central Limit Theorem to estimate the number of extra repeat readings required by the forward method to obtain the same standard error. The results of this are plotted as a dotted line on [Fig. 2](#).

Indexing is particularly important for larger sample areas, because there are more lineages to be checked for coalescence at each time step. Simulations with $\nu = 0.00001$ and sample areas of 22×22 , 200×200 , and 402×402 showed negligible, 4-fold, and 33-fold increases in speed respectively due to indexing. All told, the novel analytical and indexing components of our simulation method are a major improvement on even the basic implementations of coalescence, giving a 600-fold increase in speed for a 402×402 survey area and infinite landscape area.

In fact, the advantages of our methods when compared to the widely used forwards methods can be even greater. To minimise boundary effects, Hubbell's implementation sampled a small area (22×22 individuals) from a larger landscape (402×402 individuals). The coalescence algorithm can avoid boundary effects while only tracking the lineages of individuals originating in the survey area. Our results for a sample area of size 22×22 with infinite landscape area can therefore be compared to forwards method results for sampling an area of size 402×402 . Suppose we wish to collect data with a fixed standard error for a sample of 22×22 individuals from an infinite area and a spectrum of ten speciation rates using Hubbell's patch model (2001). The product of all the various benefits we have discussed is an algorithm 36,000 times faster than the forward method (compare the solid line with open triangular points to the dotted line with filled circular points on [Fig. 2](#)). For previous methods one had to take a sabbatical, now one only needs to take a cup of coffee!

8. Example application

As an illustration of the power and efficiency of our method, we wish to obtain the manifold of model parameter values that give rise to the empirically observed species richness for the Barro Colorado Island (BCI) forest census plot data (Condit, 1998; Condit et al., 2005; Hubbell et al., 1999). Our spatially explicit neutral model features normally distributed dispersal kernels (of different widths) and a rectangular survey area of the same shape and average density as the true BCI data, set in an infinite landscape area. For each of a number of kernel widths ranging from 7 m to 70 m, we carried out 120 replicate coalescence simulations using our model. With each replicate, the semi-analytical approach provides us with a fast method to calculate the expected species richness for any speciation rate. We used an interval bisection method with this function

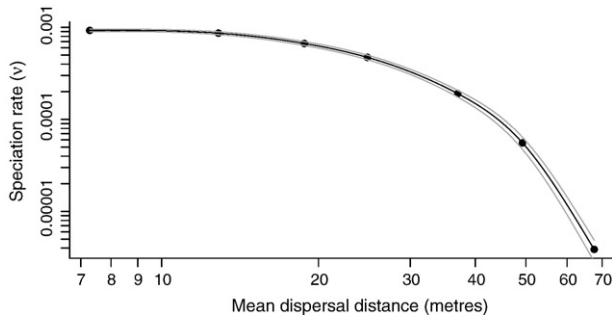


Fig. 3 – A plot of the manifold of parameter values that give rise to the true species richness of trees with dbh > 100 mm on BCI, a commonly studied tropical tree community using the latest 2005 census (Condit, 1998; Condit et al., 2005; Hubbell et al., 1999). This gave a total of 20,852 individuals of 229 distinct species. We simulated Gaussian dispersal kernels with different widths. The mean dispersal distance (in model cells) was computed from the kernel and then converted to meters using the known density of the BCI data. The grey lines indicate \pm one standard deviation on the result.

to obtain the speciation rate for which the expected richness is the same as that observed empirically. This would have been totally infeasible without our novel extensions because each function evaluation would require an entirely new simulation to be undertaken. The result is a set of estimates of model parameter values that produce the richness of the BCI data: these are plotted in Fig. 3.

The same methods were also used in a recently published study of species–area relationships (Rosindell and Cornell, 2007) that would not otherwise have been possible. We expect many further applications to follow shortly involving fat-tailed dispersal kernels, other summary statistics and phylogenetics.

9. Discussion

We have presented a set of novel algorithms, adapted from population genetics techniques, that provide considerable benefits for simulating ecological neutral models. Considering the power of this coalescence method, there have been surprisingly few applications of these techniques in the ecology literature to date, but where they have been used, considerable insight has been gained (Chave and Leigh, 2002; Chave et al., 2002; Etienne and Olff, 2004; Etienne, 2005). As a computing method, few of the benefits have been identified, explained or quantified (Chave and Leigh, 2002; Chave et al., 2002). This has left the existing and powerful methods of basic coalescence under-promoted and not utilised to their full potential. Here we have explained in detail, not only the basic coalescence method but also several novel additions that lead to yet higher efficiency and further functionality such as the ability to deal efficiently with infinite landscape areas. We expect that both the basic coalescence concepts and our novel extensions will yield significant benefits in research, not just for technical neutral modelling work but also for empirical work that includes a modelling component. We wish to make these methods freely

available and understandable to ecology: to this end our C++ code is available as an Appendix.

We have presented benchmark results for our test scenario where we simulated data to within a fixed standard error for a spectrum of 10 speciation rates $v \geq 0.00001$. For a sample of 22×22 our method was approximately 36,000 times faster than the usual forward methods. Furthermore, our advanced coalescence methods are approximately 600 times faster for sample areas of 402×402 than basic implementations of coalescence. The acceleration observed for the basic coalescence method is due to a guaranteed steady-state, eliminating calculations usually wasted in overrun and not considering lineages that later go extinct or remain outside our survey area. By comparison, the forwards method has none of these benefits and must simulate the whole landscape area rather than just the survey area. This is important because, as a means to reduce the bias caused by boundary conditions, there may be a sizable difference between the landscape and survey areas.

Coalescence methods can also be applied when studying times to equilibrium and transient states. A steady-state is reached when the initial state of the system no longer influences the final result. When using coalescence, this occurs when all lineages of interest, apart from one, have either speciated or coalesced. In each simulation step, if there are n lineages remaining in the simulation then $\frac{1}{n}$ generations have passed, we add this true time to a running total providing a far better estimate of time to equilibrium than is achievable with the forwards method. If we wish to observe transient dynamics from a given initial state, then we run the coalescence algorithm for the required number of generations and then identify the species of any remaining lineages from this initial state. Studying transients in this manner still yields all the benefits of coalescence.

The spatially explicit aspect of our work, together with the novel possibility of simulating infinite areas is particularly valuable. An indexing system streamlines memory operations and makes simulation of large areas practical. This method for handling infinite areas has already recently enabled an improved analysis of species–area curves for neutral models (Rosindell and Cornell, 2007). Exciting possibilities also exist in the study of relationships between different species under the neutrality assumption by interpreting the coalescent tree as a phylogeny (Hubbell, 2001). The method for recovering species richness from the tree analytically is novel and provides a robust recipe for collecting the exact result for a key summary statistic for diversity. This facilitates reuse of ancestral data to collect results for a spectrum of speciation rates with only negligible extra cost. The process is analytical so inter-simulation variance is much reduced compared to traditional methods. This can be traded for further acceleration by reducing the number of repeat readings required.

One application of this is parameter estimation. For any given coalescent tree, we can calculate with high precision the speciation rate for which we expect to observe the empirically known species richness. This is only possible to achieve with acceptable speed and accuracy using our semi-analytical methods. These provide an almost instantaneous calculation of the expected species richness from a known coalescent tree so we can use any algorithm to minimise the difference between the required and expected richness. Repeating the process gives

a good estimate of the manifold of parameter values that yield the empirically observed richness. Another advantage is that replicate results for the same parameter values when using coalescence are independent samples: this can be untrue of forward simulations which often sample repeatedly from the same population at intervals in order to ensure that at least some samples are taken from populations at steady-state.

Despite its huge benefits, the coalescence method has some limitations. There are variations of the classic neutral models not suited to coalescence. In general, models are defined by constructing an operation that derives the state of the system at time t from that at time $t-1$. In our examples and many other neutral models, the operation itself does not depend on the state of the system at time $t-1$, the operation being reversible. This is the crucial prerequisite that enables use of coalescence methods. This requirement ensures that at any one time differences between individuals are only dependent on the underlying environment. Despite this constraint, many different models are suitable for coalescence. For example, if we make some areas of a spatially explicit environment more susceptible to death, or have increased dispersal, then the environmental properties are known and the exact identity of any organisms does not influence the update operation. Some forms of habitat heterogeneity can therefore be incorporated into a coalescence simulation but only provided that all individuals remain ecologically equivalent and interchangeable.

Plausible neutral models can be formulated for which coalescence is problematic, such as those with an aging process where age affects fecundity or chance of death. This process may be neutral in that all individuals, regardless of species, age in an identical manner, yet individuals at a particular time can have different probabilities of dying based on age. Models incorporating density dependence in birth or death rates may also pose problems (Alonso et al., 2008). Breaking the neutrality assumption itself and incorporating species-specific traits or allowing dispersal to be species-specific may be desirable (Ozinga et al., 2005), but coalescence methods can no longer be used. This is because the species identity of individuals is not known until the simulation is complete and cannot therefore be allowed to influence the process of simulation. Forwards methods can handle these more complex models, albeit with the problems of computational expense and uncertainty regarding sampling from an equilibrium.

A final, minor limitation is that simulating speciation spectra, the results for different speciation rates will not be independent because we have used the same coalescence tree through out. Any features of this tree that affect speciation will influence the full spectrum of speciation results. Fortunately this problem is much reduced by averaging results from many trees. It also allows us to investigate the effect of the tree topology on the results. This merits further study.

This paper adds to the existing links between the neutral theory of ecology and population genetics by reusing similar theoretical approaches (Chave, 2004; Hu et al., 2006). We have outlined existing efficient simulation techniques that remain strangely elusive in ecology literature and suggested several novel extensions that lead to considerable further benefits. There is potential for more theoretical work on the methods

themselves. We have explained and cited applications that make direct use of our novel extensions and suggested avenues for further research. Our methods show that the greatest benefits accrue when investigating large simulations over broad areas of parameter space. These are exactly the type of problems that will be most interesting ecologically and we hope that our methods will be used to provide a significant insight into neutral models that, for practical reasons, would not have been possible otherwise.

Acknowledgements

We thank Stephen Cornell and Jerome Chave for their feedback and Luis Borda de Agua for his advice. James Rosindell was funded by a University of Leeds Research Scholarship under supervision of Stephen Cornell and Bill Kunin. Rampal S. Etienne thanks the support of the Netherlands Organisation of Scientific Research (NWO). The Forest Dynamics Plot of Barro Colorado Island has been made possible through the generous support of the U.S. National Science Foundation, The John D. and Catherine T. MacArthur Foundation, and the Smithsonian Tropical Research Institute and through the hard work of over 100 people from 10 countries over the past two decades. The BCI Forest Dynamics Plot is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots. We also thank the editor and anonymous referees for their constructive remarks.

Appendix A. An analytical approach to calculating species richness

Here, we present an algorithm to calculate the expected species richness (S^*) of a sample, given a coalescent tree and a speciation rate. At some point in the past, all lineages coalesce to one. At this point, the single remaining individual belongs to what we denote as the “basal species”. Our calculations turn out to be much simplified if, instead of counting the total number of species in the sample (S^*), we count the number of “novel” species (S): those that are different to the basal one. The basic concept is to consider each part of the tree, refining an interval $[S_{\min}, S_{\max}]$ within which the expectation S must lie. Later, we will show that S^* also lies in this interval. Initially, $S_{\min}=0$ (all individuals could be of the basal species), and S_{\max} equals the total number of individuals in the sample (each could be a novel species so all are different from the basal species).

Our method can be thought of as a tree pruning algorithm, with both branches (edges) and coalescence events (nodes) being methodically considered and discarded. As each are pruned, S_{\min} and S_{\max} are updated, and they gradually converge to S . Coalescence involves both a birth event and a coalescent point. However, it is helpful to consider the birth event as occurring on a branch and the coalescent point as a node.

The algorithm can be imagined as starting at the bottom of the tree and working upwards. First we prune all “free” branches (those without a node on the bottom). Then we prune any free nodes (those which have had both descendant branches pruned). This produces more free branches, which

can themselves be pruned, and so on. The principle is illustrated in Fig. A1. We will show that pruning a branch increases S_{\min} but does not change S_{\max} , whilst pruning a node decreases S_{\max} but leaves S_{\min} untouched.

Multiple speciations on a single lineage

A lineage is a novel species if it contains at least one speciation event: the exact number is irrelevant. This makes our calculation of species richness more challenging; we cannot simply sum all speciation events because some do not change the species richness. We introduce a method of keeping track of which speciation events change richness and which are simply repeats. Previously this was unnecessary, as we simply stopped tracing a lineage once speciation occurred. For the analytical method explained here, there are no definite speciation events, only probabilities of speciation. We therefore attach to each branch (i) a probability u_i that the species richness would change given speciation on branch i. This

probability can also be interpreted as the chance that a conventional backwards simulation is still tracing the ancestry of this branch. At the start of a simulation, speciation on any free branch must result in a new species, hence for all initially free branches this probability is equal to 1 (Fig. A1, step 1).

Branch pruning

It is relatively simple to see why pruning a branch increases S_{\min} . This is because a branch may have a speciation event on it, which may increase the species richness. As an initial example, take step 1 from Fig. A1. There is a single birth event on the leftmost branch (branch 1), and hence a single point at which speciation could have occurred. This means that the probability of speciation along the branch is v . After pruning branch 1, the minimum expected number of novel species increases from $S_{\min}=0$ to $S_{\min}=v$.

This first branch had only a single birth event on it. However, in general, a branch i may have any number of birth events b_i on it. When considering whether a branch results in a new species, if one speciation event occurs, any additional speciations on the branch are irrelevant. We thus need to calculate p_i , the probability of having any speciation events on the branch. Since the probability of zero speciations occurring on a branch with b_i births is simply $(1-v)^{b_i}$, we can put

$$p_i = 1 - (1-v)^{b_i} \tag{1}$$

For an initial branch with b_i births the increase in S_{\min} is simply p_i . However this assumes that speciation on branch i

Box A1				
Step	[S_{\min}	, S_{\max}]
1	[0	, 3]
2	[v	, 3]
3	[$3v-v^2$, 3]
4	[$5v-2v^2$, 3]
5	[$5v-2v^2$, $3-(1-v)^3$]
6	[$v^4-2v^3-2v^2+6v$, $3-(1-v)^3$]
7	[$v^4-2v^3-2v^2+6v$, $3-(1-v)^3(2-2v^2+v^3)$]

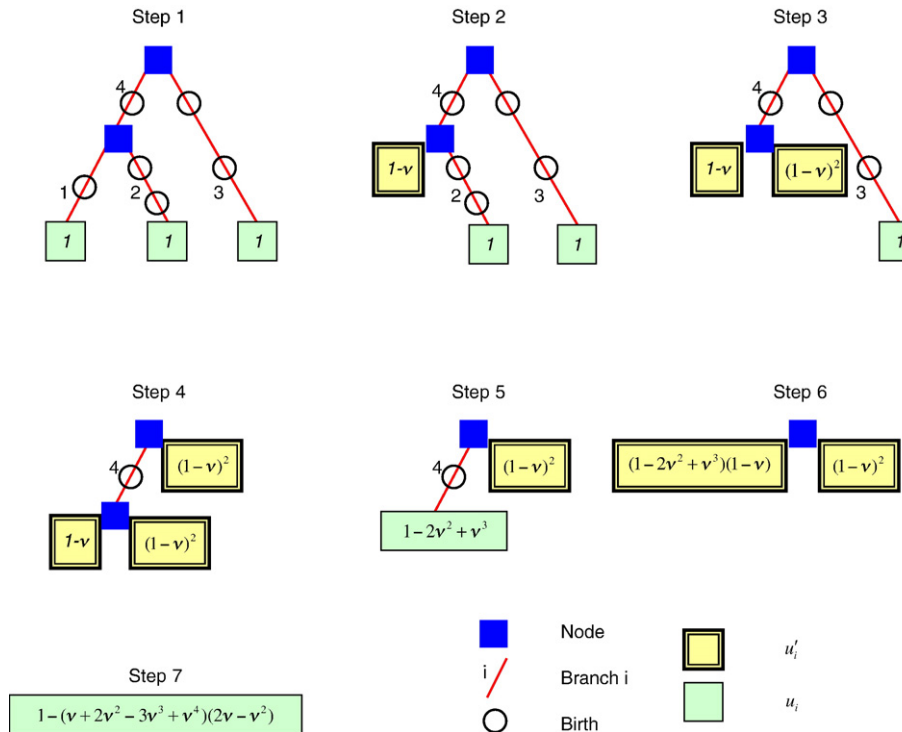


Fig. A1 – Schematic of branch and node pruning. The initial tree (step 1) is identical to that in Fig. 1(b) in the main text and uses the same symbols: blue filled squares for coalescence events, open circles for births and red lines for branches. Box A1 above shows the intervals $[S_{\min}, S_{\max}]$ to correspond with the completion of each step in this figure.

will always result in a new species. This is not the case in general (e.g. when pruning branches higher up in the tree). Previously, we defined u_i as the probability that a speciation event on branch i would result in a new species. Hence, more generally S_{\min} increases by $u_i p_i$ (giving the observed increase of $p_1=v$ in the initial case where $u_1=1$).

Pruning a branch i from the tree also produces a new value u'_i (shown in a double box on Fig. A1) at the top of the branch. As with u_i this measures the probability that speciation further up this lineage will change the species richness. For further speciation to alter the species richness after pruning branch i , two conditions must hold. Firstly, before pruning the branch, we must know that speciation on it will alter the species richness; this has probability u_i . Secondly, there must be no speciation on the branch i itself, otherwise further speciation becomes irrelevant; this has probability $1-p_i$. Since the probabilities of the two are independent, we simply multiply them to obtain

$$u'_i = u_i(1 - p_i) \tag{2}$$

In Fig. A1, step 2 has $u_1=1$ and $p_1=v$, so $u'_1=1(1-p_1)=1-v$.

Recall that pruning branch 1 gave $S_{\min}=u_1 p_1$ at step 2. We must calculate how S_{\min} changes with further pruning. Crucially, within a round of branch pruning, increases in S_{\min} are independent of what happens on other branches. For example, in Fig. A1, note that speciation on branch 2 always increases the number of novel species by 1 (either from 0 to 1 if speciation did not previously occur on the leftmost branch or from 1 to 2 if speciation did previously occur there). This additivity means that pruning branches 1 and 2 results in $S_{\min}=u_1 p_1 + u_2 p_2$. In general, we can continue in this manner until we start node pruning: every time we prune a branch, we update S_{\min} with

$$S_{\min} \leftarrow S_{\min} + p_i u_i \tag{3}$$

Node pruning

A node marks the coalescence of two lineages and hence the end of two branches, i and j and the start of a single new combined branch k . Our first task is to calculate u_k from u'_i and u'_j . As before, u_k is the probability that speciation on branch k increases the species richness. If speciation has not occurred in the lineage down from (and including) branch i , then clearly speciation in branch k or above will increase the richness. We previously defined this probability as u'_i . The same is true for the lineage down from branch j (probability u'_j). Standard probability theory gives $\Pr\{A \text{ or } B\} = 1 - \Pr\{\text{not } A\}\Pr\{\text{not } B\} = 1 - (1 - \Pr\{A\})(1 - \Pr\{B\})$ hence

$$u_k = 1 - (1 - u'_i)(1 - u'_j) \tag{4}$$

In our example, pruning the leftmost node gives $u_4=1 - (1 - u'_1)(1 - u'_2) = 1 - (1 - (1 - v))(1 - (1 - v)^2)$, as shown in step 5. When we prune any node, S_{\min} remains unchanged as, by definition, nodes contain no birth events and therefore no chance of increasing the richness via speciation. However, node pruning does change S_{\max} .

Updating S_{\max}

We will show that, although S_{\max} can be recalculated after pruning either branches or nodes, its value only changes during the latter. At any point during pruning we expect at least S_{\min} novel species, however, there are likely to be more species in the simulation than this, because speciation may have occurred in branches that we have not yet pruned. S_{\max} can therefore be thought of as the sum of S_{\min} and an additional component, w , that we shall now calculate. Consider any partially pruned tree, there will be a number of remaining lineages whose fate (speciation or coalescence) is decided by the unpruned sections of the tree. Although the exact detail of these lineages is unknown, we nevertheless know that the maximum species richness will be obtained if all these remaining lineages represent novel species. However, an added complication exists because if speciation occurs lower down in the tree, some of these novel species may be replaced and hence absent from our final sample.

By assuming all remaining lineages represent novel species, w can be calculated for any partially pruned tree. This requires us to find, for each remaining lineage, the probability that at least one of its descendant lineages is speciation-free. But we have already calculated this probability: it is simply u (for an unpruned branch) or u' (for a pruned branch). Readers may be convinced of this by noting that speciation has an effect on species richness (the definition of u or u') if and only if there is at least one descendant lineage which is speciation-free. So S_{\max} at any stage of pruning is simply S_{\min} , plus w , the sum of u_i for all free branches i , and the sum of u'_j for all just-pruned branches j (in other words, the sum of S_{\min} and all the boxes at any stage of Fig. A1). Pruning a branch i causes w to change by $u'_i - u_i$ which, from Eq. (2), simplifies to $u_i(1 - p_i) - u_i = -p_i u_i$. From Eq. (3), we see that pruning a branch causes an increase in S_{\min} of $p_i u_i$. These cancel out, showing that branch pruning causes no net change in S_{\max} .

Consider the effect that pruning a node has on S_{\max} . Two branches i and j have converged into branch k so u'_i and u'_j are deducted from w while u_k is added. Since S_{\min} is unchanged, this results in a net change of $1 - (1 - u'_i)(1 - u'_j) - u'_i - u'_j = -u'_i u'_j$ and so node pruning results in:

$$S_{\max} \leftarrow S_{\max} - u'_i u'_j \tag{5}$$

Returning to our example in Fig. A1, pruning the one free node (on the left) updates S_{\max} to $3 - (1 - v)(1 - v)^2 = 3 - (1 - v)^3$.

Relation to true species richness S^*

We have presented formulae to calculate the interval $[S_{\min}, S_{\max}]$ for S , the expected number of “novel” species in a sample (that is, those distinct from the basal species). However, we are primarily interested in the total number of species in our sample, a slightly larger value with an expectation of S^* . The two are very similar: $S^* - S$ cannot be greater than one, and is negligible in practice. Note that the maximum number of novel species was found by assuming that all unknown lineages were novel species. In this case, no lineage can represent the basal species, and so $S^*_{\max} = S_{\max}$. Since $S \leq S^*$ we also have

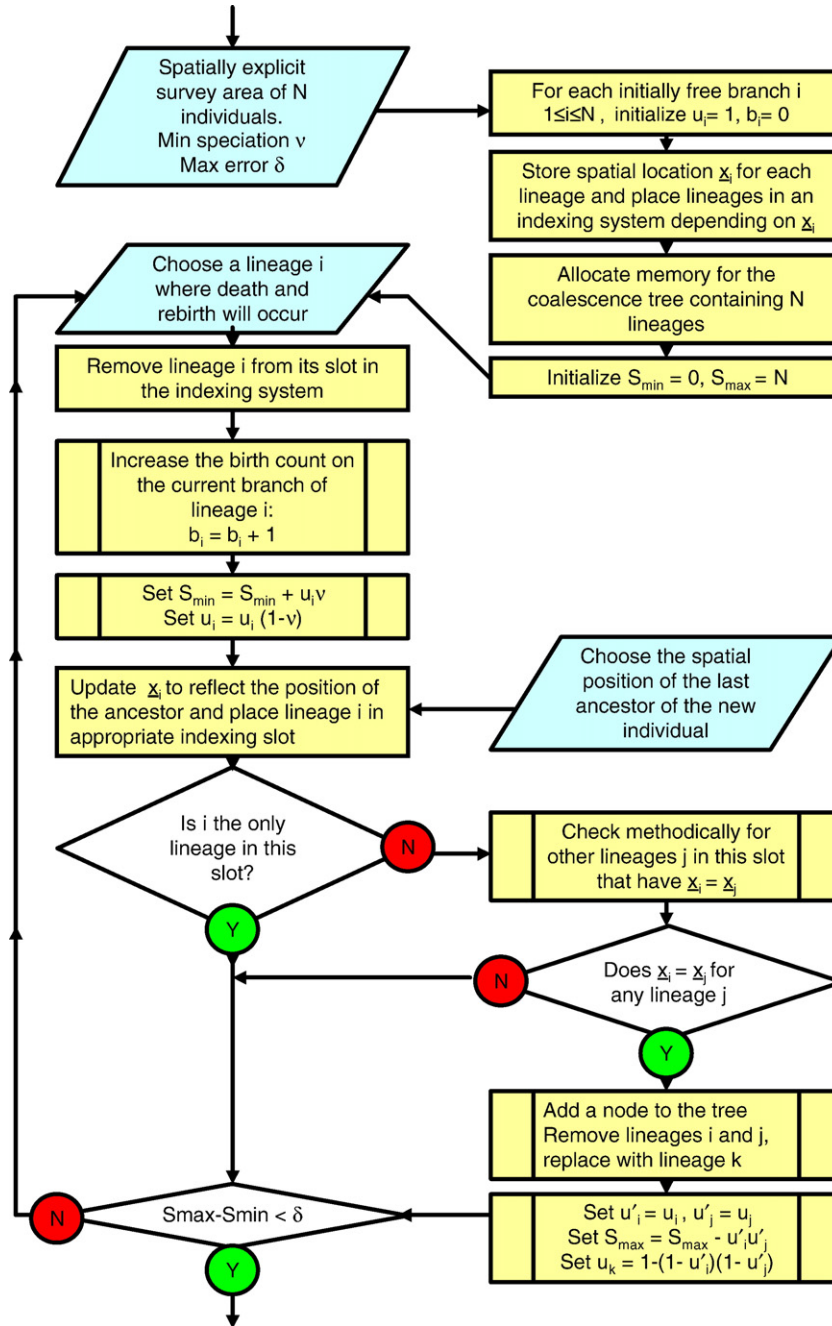


Fig. A2 – Schematic for producing a coalescence tree from a model to be suitable for a predefined spectrum of speciation rates.

$S_{min} \leq S \leq S^*$, hence the true species richness S^* lies in the our interval $[S_{max}, S_{min}]$.

Convergence

We wish to prove that S is uniquely defined, in other words that the width w of the interval $[S_{max}, S_{min}]$ converges to zero for any tree we are pruning. We include in our proof not only cases where lineages converge to form a single tree, but also cases in which some lineages continue indefinitely without coalescing.

Take the general case where we are tracking the ancestry of N lineages. The algorithm initializes $S_{max} = N$ and $S_{min} = 0$ hence

$w = S_{max} - S_{min} = N$. We wish to show that $w \rightarrow 0$ as we prune. We use proof by induction: first we show that $w \rightarrow 0$ if there are no nodes at all ($m=0$) and then we show that if $w \rightarrow 0$ for m nodes, then $w \rightarrow 0$ for $m+1$ nodes as well. Where $m=0$, there is no coalescence at all, so all N lineages continue indefinitely without coalescing. In this case, we have N infinite-length branches and 0 nodes to prune from our tree. When pruning branch i we have $S_{min} \leftarrow S_{min} + p_i u_i$ (by Eq. (3)) and S_{max} unchanged, hence $w \leftarrow w - p_i u_i$. We have $u_i = 1$ by initialisation and since branch lengths are infinite, $p_i = 1 - (1 - v)^\infty$ (by Eq. (1)) hence $p_i = 1 - 0 = 1$ because $0 < v \leq 1$. This gives $w \leftarrow w - 1$ and is repeated N times (once for each branch). Hence as required, $w = N - N = 0$ after all branches have been pruned.

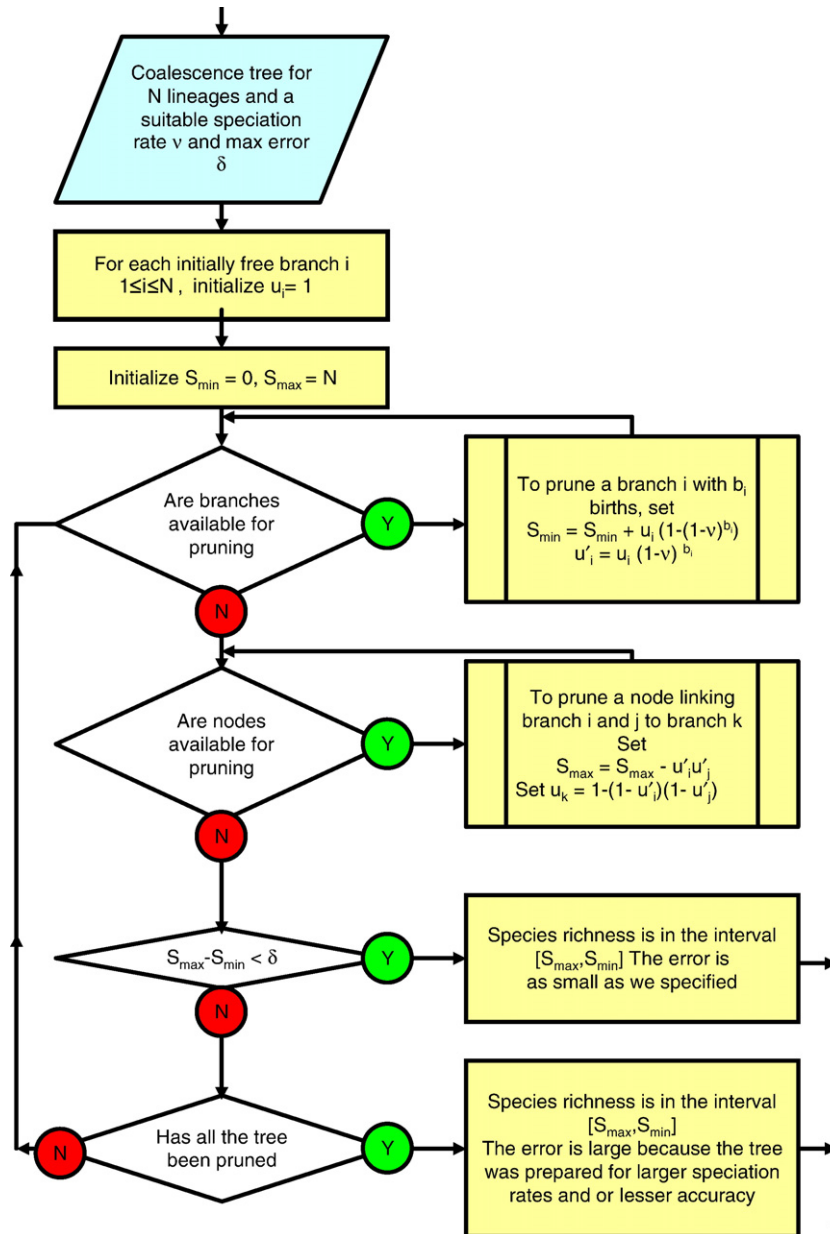


Fig. A3 – Schematic for calculating species richness from a coalescence tree and a speciation rate.

Assume for now that $w \rightarrow 0$ for all trees with m nodes. We shall prove that $w \rightarrow 0$ for an arbitrary tree with $m + 1$ nodes. Take the uppermost node in our tree and remove it: we are now left with two lineages that used to coalesce but now continue indefinitely without ever coalescing. The tree that we now have has only m nodes and so, by our assumption, the property $w \rightarrow 0$ holds. Now re-introduce the uppermost node. This means shortening two branches i and j to lengths b_i and b_j and joining them at a node to a single branch k which continues indefinitely without further coalescence. We simply need to show that this change does not affect w and then we will have $w \rightarrow 0$ as required. Before the change, we have two branches i and j to prune, hence decreasing w by $\Delta w = p_i u_i + p_j u_j$ (by Eq. (3)) where $p_i = p_j = 1 - (1 - v)^{\infty} = 1$ (by Eq. (1)) so $\Delta w = u_i + u_j$. After the change, we have three

branches and a node to prune. Pruning branches i and j still decreases w by $p_i u_i + p_j u_j$ but now $p_i = 1 - (1 - v)^{b_i}$ and $p_j = 1 - (1 - v)^{b_j}$ (by Eq. (1)). The node will also decrease w by $u'_i u'_j$ (by Eq. (5)) and leave a final branch k to be pruned. Pruning branch k decreases w by $p_k u_k$ (by Eq. (3)) $= (1 - (1 - v)^{\infty})(1 - (1 - u'_i)(1 - u'_j))$ (by Eqs. (1) and (4)) $= 1 - (1 - u'_i)(1 - u'_j) = u'_i + u'_j - u'_i u'_j$. Now we total up the decrease to w as a result of pruning branch k and the node, giving $u'_i u'_j + u'_i + u'_j - u'_i u'_j = u'_i + u'_j = u_i(1 - p_i) + u_j(1 - p_j)$ (by Eq. (2)). Finally, include the pruning of the branches i and j we obtain a total decrease in w of $u_i(1 - p_i) + u_j(1 - p_j) + p_i u_i + p_j u_j = u_i + u_j$. This is the same as Δw , the decrease for the tree where we know $w \rightarrow 0$. Hence if convergence is true for m then it is also true for $m + 1$. As we have also shown that $w \rightarrow 0$ for $m = 0$, by induction we have proved convergence for all m .

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.ecoinf.2008.05.001](https://doi.org/10.1016/j.ecoinf.2008.05.001).

REFERENCES

- Alonso, D., Etienne, R.S., McKane, A.J., 2006. The merits of neutral theory. *Trends Ecol. Evol.* (ISSN: 0169-5347) 21 (8), 451–457, Aug.
- Alonso, D., Ostling, A., Etienne, R.S., 2008. The implicit assumption of symmetry and the species abundance distribution. *Ecol. Lett.* (ISSN: 1461-023X) 11.
- Blythe, R.A., McKane, A.J., 2007. Stochastic models of evolution in genetics, ecology and linguistics. *J. Stat. Mech.-Theory Exp.* (ISSN: 1742-5468) 07018.
- Chave, J., 2004. Neutral theory and community ecology. *Ecol. Lett.* (ISSN: 1461-023X) 7 (3), 241–253, Mar.
- Chave, J., Leigh, E.G., 2002. A spatially explicit neutral model of beta-diversity in tropical forests. *Theor. Popul. Biol.* (ISSN: 0040-5809) 62, 152–168.
- Chave, J., Muller-Landau, H.C., Levin, S.A., 2002. Comparing classical community models: Theoretical consequences for patterns of diversity. *Am. Nat.* (ISSN: 0003-0147) 159, 1–23.
- Condit, R., 1998. *Tropical Forest Census Plots*. Springer-Verlag and R. G. Landes company, Berlin, Germany.
- Condit, R., Hubbell, S.P., Foster, R.B., 2005. Barro Colorado forest census plot data. <http://ctfs.si.edu/datasets/bci>.
- Etienne, R.S., 2005. A new sampling formula for neutral biodiversity. *Ecol. Lett.* (ISSN: 1461-023X) 8, 253–260.
- Etienne, R.S., 2007. A neutral sampling formula for multiple samples and an 'exact' test of neutrality. *Ecol. Lett.* (ISSN: 1461-023X) 10 (7), 608–618, Jul.
- Etienne, R.S., Alonso, D., 2007. Neutral community theory: how stochasticity and dispersal-limitation can explain species coexistence. *J. Stat. Phys.* (ISSN: 0022-4715) 128, 485–510.
- Etienne, R.S., Olf, H., 2004. A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.* (ISSN: 1461-023X) 7, 170–175.
- Etienne, R.S., Alonso, D., McKane, A.J., 2007a. The zero-sum assumption in neutral biodiversity theory. *J. Theor. Biol.* (ISSN: 0022-5193) 248, 522–536.
- Etienne, R.S., Apol, M.E.F., Olf, H., Weissing, F.J., 2007b. Modes of speciation and the neutral theory of biodiversity. *Oikos* (ISSN: 0030-1299) 116, 241–258.
- Felsenstein, J., 2004. *Inferring Phylogenies*. Sinauer Associates. (ISBN: 0-87893-177-5).
- Gewin, V., 2006. Beyond neutrality — ecology finds its niche. *PLoS Biol.* (ISSN: 1544-9173) 4 (8), 1306–1310 Aug.
- Holley, R.A., Liggett, T.M., 1975. Ergodic theorems for weakly interacting systems and the voter model. *Ann. Probab.* (ISSN: 0091-1798) 3, 643–663.
- Holyoak, M., Loreau, M., 2006. Reconciling empirical ecology with neutral community models. *Ecology* (ISSN: 0012-9658) 87 (6), 1370–1377, Jun.
- Hu, X.S., He, F.L., Hubbell, S.P., 2006. Neutral theory in macroecology and population genetics. *Oikos* (ISSN: 0030-1299) 113 (3), 548–556, Jun.
- Hubbell, S.P., 2001. *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ. Oxford (ISBN: 0-691-02128-7).
- Hubbell, S.P., 2005. Neutral theory in community ecology and the hypothesis of functional equivalence. *Funct. Ecol.* (ISSN: 0269-8463) 19 (1), 166–172 Feb.
- Hubbell, S.P., Foster, R.B., O'Brien, S.T., Harms, K.E., Condit, R., Wechsler, B., Wright, S.J., de Lao, S.L., 1999. Light-gap disturbances, recruitment limitation, and tree diversity in a neotropical forest. *Science* (ISSN: 0036-8075) 283 (5401), 554–557, Jan.
- Kingman, J.F.C., 1982. On the genealogy of large populations. *J. Appl. Probab* 19A, 27–43.
- Leigh, E.G., 2007. Neutral theory: a historical perspective. *J. Evol. Biol* (ISSN: 1010-061X) 20 (2075–2091).
- Moran, P.A.P., 1962. *Statistical processes of evolutionary theory*. Clarendon press, Oxford.
- Notohara, M., 1990. The coalescent and the genealogical process in geographically structured populations. *J. Math. Biol.* (ISSN: 0303-6812) 29 (1), 59–75.
- Ozinga, W.A., Schaminee, J.H.J., Bekker, R.M., Bonn, S., Poschlod, P., Tackenberg, O., Bakker, J., van Groenendael, J.M., 2005. Predictability of plant species composition from environmental conditions is constrained by dispersal limitation. *Oikos* (ISSN: 0030-1299) 108 (3), 555–561, Mar.
- Rangel, T.F.L.V.B., Diniz, J.A.F., 2005. Neutral community dynamics, the mid-domain effect and spatial patterns in species richness. *Ecol. Lett.* (ISSN: 1461-023X) 8 (8), 783–790, Aug.
- Rosindell, J., Cornell, S.J., 2007. Species–area relationships from a spatially explicit neutral model in an infinite landscape. *Ecol. Lett.* (ISSN: 1461-023X) 10 (7), 586–595, Jul.
- Ulrich, W., Zalewski, M., 2007. Are ground beetles neutral? *Basic Appl. Ecol.* (ISSN: 1439-1791) 8 (5), 411–420.
- Wakeley, J., 2006. *Coalescent theory, an introduction*. Roberts and Company Publishers. (ISBN: 0-9747077-5-9).