

Phylogenetics and speciation

Timothy G. Barraclough and Sean Nee

Species-level phylogenies derived from molecular data provide an indirect record of the speciation events that have led to extant species. This offers enormous potential for investigating the general causes and rates of speciation within clades. To make the most of this potential, we should ideally sample all the species in a higher group, such as a genus, ensure that those species reflect evolutionary entities within the group, and rule out the effects of other processes, such as extinction, as explanations for observed patterns. We discuss recent practical and theoretical advances in this area and outline how future work should benefit from incorporating data from genealogical and phylogeographical scales.

SPECIATION (see Glossary) is the creative process leading to the build up of species diversity; understanding the general patterns and processes of speciation is fundamental to explaining the diversity of life¹. Central questions include: what are the general causes of speciation, and how do rates of speciation vary over time and among TAXONOMIC groups and geographical regions? Although fundamental, such questions are difficult to answer because direct observation is usually impossible and many extant groups do not have adequate fossil records with which to investigate these topics² (but see Benton and Pearson³, this issue).

The expansion of molecular PHYLOGENETICS over the past decade has opened up a powerful new approach to this problem. PHYLOGENETIC TREES, particularly those including all the living species in a higher taxonomic group, provide an indirect record of the speciation events that have led to present-day species⁴. Together with information on the geographical and ecological attributes of species, they can provide information on the causes of speciation within the group^{5–7}. In addition, trees derived from DNA sequences contain information about the relative timing of reconstructed speciation events, and can be used to estimate speciation rates^{8–10}.

This has led to considerable interest in molecular phylogenetics as a tool for solving speciation problems, but success rests on two fundamental issues. First, the approach relies on reconstructing evolutionary species relationships within a CLADE. Few biologists deny that the evolutionary entities referred to as species do exist, but the taxonomically recognized species in a clade might not correspond to those entities¹¹ (see Hey¹², this issue). Second, processes in addition to speciation influence the PHYLOGENY and attributes of present-day species^{1,6}. For example, species are lost by extinction, and subsequent phenotypic evolution might obscure the ecological pattern of speciation events. Hence, a phylogenetic tree cannot simply be read as the history of speciation: analyses must factor out alternative processes as explanations for observed patterns.

Our review discusses recent progress in this area. We discuss methodological issues in reconstructing species histories, and move on to the two broad questions concerning the rates and causes of speciation. Finally, we consider the future possibilities of incorporating fine-scale GENEALOGICAL data into large-scale phylogenetic studies.

Species-level phylogenies

Phylogenetic studies of speciation have focused on reconstructing species-level phylogenies, namely the relationships among species within higher groups such as genera^{13–15}. The INTERNAL NODES of the tree reflect speciation events: it is known 'who split from whom' and, ideally, the timing of those events. Because reconstruction relies on living species, there is no record of speciation events involving species that subsequently went extinct, although extinction can leave a signature on the shape of the tree. Phylogenetic reconstruction from molecular data is an advanced science¹⁶, but two issues have been particularly important when applying phylogenetics to studies of speciation.

The first is the issue of sampling. To obtain an accurate view of speciation in a higher group, nearly all the species from that group should be sampled. Missing species reduce the sample size of reconstructed speciation events available, and can introduce bias, for example by tending to remove the most recent speciation events¹⁷ or those involving rare species⁷. In particular, the ability to consider the effects of other processes, such as extinction, on the observed patterns relies crucially on a very complete sample of species. However, most phylogenetic studies do not sample all the described species within a group, and, for reasons outlined in Box 1, special effort can be required to do so.

The second issue relates to the status of species included in phylogenetic trees. Most phylogenetic studies of higher clades sample one individual from each species named in the checklist for the group. Using the resulting tree to study speciation assumes that taxonomic species generally reflect the evolutionary entities whose origin we wish to explain, but this need not be so. The number and identity of taxonomic species in a group depends upon the judgement of the taxonomist who described them, and opinions can vary (the infamous splitting–lumping continuum). More fundamentally, it might not always be possible to resolve the true history of speciation into a tree of MONOPHYLETIC species, because some modes of speciation lead to PARAPHYLETIC species¹⁸ or even phylogenetic

Timothy G. Barraclough*
Dept of Biology and NERC
Centre for Population
Biology, Imperial College
at Silwood Park, Ascot,
Berkshire, UK SL5 7PY.
*e-mail:
t.barraclough@ic.ac.uk

Sean Nee
Division of Biological
Sciences, University of
Edinburgh, West Mains
Road, Edinburgh,
UK EH9 3JT.

Box 1. Reconstructing species-level phylogenies

Several methodological problems have faced researchers wanting to reconstruct species-level phylogenies for studying speciation. The first is obtaining DNA from all the species of a sufficiently large monophyletic group to detect any patterns. Collection of DNA from rare species might be difficult or impossible and it might be hard to guarantee PCR amplification from every species using universal primers. Statistical methods exist to investigate incomplete samples of species^a, but complete samples will still be needed to detect many patterns. Second, even if samples can be obtained, it might be difficult to obtain a fully resolved tree. Many existing DNA regions were developed for higher level systematics and might evolve too slowly to resolve the species level. Faster evolving regions are available, but can have associated problems, such as the existence of alternative multiple copies or the difficulty of using universal primers in the chosen target group^b. Alternative methods are being explored, such as genetic fingerprinting techniques or the use of transposable elements^{c,d}.

Alternatively, we might have to accept that 100% resolution is not always possible at this level, and develop analytical approaches that take uncertainty into account. For example, Bayesian approaches are being developed that allow estimation of evolutionary parameters, such as speciation and extinction rates, as an integral part of the phylogeny reconstruction process, taking account of uncertainty in the tree^e. Finally, existing studies are usually limited to a single exemplar for each described species in a group, and to one or a few molecular markers. Several processes can complicate reconstruction of species history from the resulting gene trees, including modes of speciation producing paraphyletic species, lineage sorting of ancestral polymorphisms and hybridization^{f,g}. Genealogical methods based on reconstructing multi-gene trees and sampling larger numbers of individuals per named species are available to cope with these issues^{h,i}. These approaches allow accurate reconstruction of species histories, but to date have been too time-consuming or costly to apply to studies of large groups. Aside from difficulties of collection, the above problems are becoming less severe as molecular methods become faster and cheaper, making it easier to obtain comprehensive samples for multiple genes within clades.

References

- a Pybus, O.G. and Harvey, P.H. (2000) Testing macroevolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. London B Biol. Sci.* 267, 2267–2272
- b Hillis, D.M. *et al.*, eds (1996) *Molecular Systematics*, Sinauer Associates
- c Verneau, O. *et al.* (1998) Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. *Proc. Natl. Acad. Sci. U. S. A.* 95, 11284–11289
- d Albertson, R.C. *et al.* (1999) Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5107–5110
- e Yang, Z.H. and Rannala, B. (1997) Bayesian phylogenetic inference using DNA sequences: a Markov Chain Monte Carlo method. *Mol. Biol. Evol.* 14, 717–724
- f Avise, J.C. (2000) *Phylogeography: the History and Formation of Species*, Harvard University Press
- g Brower, A.V.Z. *et al.* (1996) Gene trees, species trees, and systematics: a cladistic perspective. *Annu. Rev. Ecol. Syst.* 27, 423–450
- h Maddison, W.P. (1997) Gene trees in species trees. *Syst. Biol.* 46, 523–536
- i Koufopanou, V. *et al.* (1997) Concordance of gene genealogies reveals reproductive isolation in the pathogenic fungus *Coccidioides immitis*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 5478–5482
- j Kliman, R.M. *et al.* (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics* 156, 1913–1931

NETWORKS¹⁹. The precise description of species histories has been the focus of genealogical and PHYLOGEOGRAPHICAL studies of speciation¹¹ (see Nichols²⁰, this issue), which rely on more intensive sampling within species (Box 1). In some groups,

described species appear to conform well to evolutionary units^{14,21}, but less so in others²². In the absence of detailed information, taxonomically defined species can at best be regarded as hypotheses of evolutionary groups (see Hey¹², this issue).

To date, time and expense have forced studies to reach a compromise between these two issues. Phylogenetic studies have looked at speciation on a broad scale, allowing investigation into general causes and rates, but have usually relied on taxonomic species. Phylogeographical and genealogical studies have delimited evolutionary groups more precisely, but, in most cases, have been restricted to much fewer speciation events by the need for intensive within-species sampling (but see Ref. 11). As sequencing technology becomes faster and cheaper, the need to compromise is lessening. We discuss primarily the phylogenetic approach, but also highlight those areas that should benefit most from integrating phylogenetic and genealogical levels.

A final area of phylogenetics important for studies of speciation is the use of molecular data to date nodes on a phylogeny. Most phylogenetic studies of speciation rely on information of the relative ages of speciation events, whether for estimating speciation rates or identifying recent nodes most likely to reveal possible causes of speciation. Although methods for estimating relative node ages from sequence data are advancing rapidly and no longer rely on the assumption of a strict molecular clock, it can still be difficult to calibrate the tree in real time (Box 2).

Estimating speciation rates

A central question concerning speciation within clades relates to how often speciation occurs. A large literature has developed describing methods for estimating speciation rates from trees containing all the species within a clade¹⁰. These methods use information on the time elapsed between successive branching events, and can be illustrated graphically by plotting the number of LINEAGES through time (Fig. 1). If speciation rates have been constant over time and among lineages, and there has been no extinction, a straight line with slope equal to the average per lineage speciation rate, *b* is expected (Fig. 1).

This approach has been used²³ to estimate the average speciation rate during the radiation of Hawaiian silverswords (family Asteraceae). This plant group has diversified into a wide range of taxa on the archipelago, from cushion plants and vines to trees. Using an inferred maximum age for the divergence between silverswords and their Californian sister clade, a minimum per lineage speciation rate of 0.56 ± 0.17 species my^{-1} was estimated. This suggests that the silverswords have speciated at rates comparable to peak origination rates observed from fossil evidence during continental radiations.

Box 2. Dating phylogenies

DNA sequence data can be used to estimate the relative ages of nodes on a phylogeny. Assuming that nucleotide substitutions between taxa accumulate randomly over time, molecular distances reconstructed onto the phylogeny are expected to be roughly proportional to the time elapsed^a. However, variation in substitution rates among lineages means that we cannot assume a strict molecular clock in most cases. Recently, a variety of methods have been proposed for estimating the relative ages of nodes, with confidence intervals, from sequence data even in the absence of molecular clock. Most are likelihood methods that fit node ages under explicit or heuristic models of how rates change among lineages^{b–d}. The exception is Sanderson's Non-Parametric Rate Smoothing algorithm^e, which converts an unconstrained tree (in which branch lengths reflect rate as well as time) into an ultrametric tree (in which branch lengths only reflect time) by minimizing rate changes across the tree. These methods have not yet been fully evaluated on real data, but represent an important step forward. Additional complexities can arise when considering very recent speciation events, in which case genetic diversity within populations can have a large effect on estimates of divergence times (see Nichols^f, this issue).

Calibration of the tree in real time remains difficult, relying on the availability of fossil dates or biogeographical evidence, which can be lacking in some groups. The traditional approach of using blanket calibrations for rates of molecular evolution, such as the widely used insect mitochondrial DNA clock of 2% pairwise sequence divergence per million years^g, is confounded by rate variation among taxa. However, more sophisticated calibrations should be possible in future, using tests for rate variation between study groups and reference clades with calibrated dates. This will rely on the availability of a library of well-dated phylogenies derived from commonly used genes.

References

- a Hillis, D.M. *et al.*, eds (1996) *Molecular Systematics*, Sinauer Associates
- b Thorne, J.L. *et al.* (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* 15, 1647–1657
- c Huelsenbeck, J.P. *et al.* (2000) A compound Poisson process for relaxing the molecular clock. *Genetics* 154, 1879–1892
- d Yoder, A.D. and Yang, Z.H. (2000) Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* 17, 1081–1090
- e Sanderson, M.J. (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14, 1218–1231
- f Nichols, R. (2001) Gene trees and species trees are not the same. *Trends Ecol. Evol.* 16, 358–364
- g Brower, A.V.Z. (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial-DNA evolution. *Proc. Natl. Acad. Sci. U. S. A.* 91, 6491–6495

The situation is more complicated if the data are not consistent with a constant speciation rate model. For example, extinction occurring randomly and at a roughly constant rate over time is expected to cause an apparent acceleration in speciation rate towards the present (Box 3). In this case, it is possible to estimate speciation and extinction rates separately, rather than just to estimate the net DIVERSIFICATION RATE^{17,24}. However, as described in Box 3, other processes, and sampling and taxonomic artefacts in particular, can affect the shape of the plots. Estimates of speciation rates and their interpretation rely heavily on these issues. Nonetheless, although the outcomes of some processes might be statistically

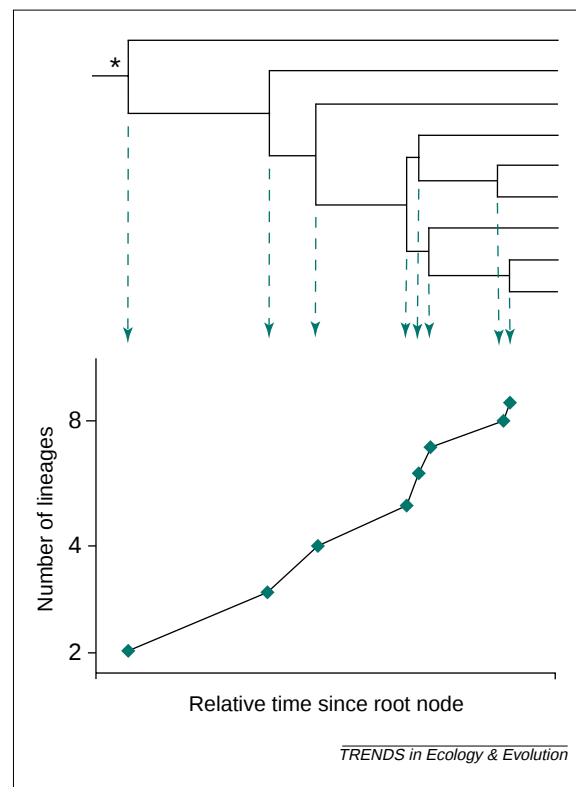


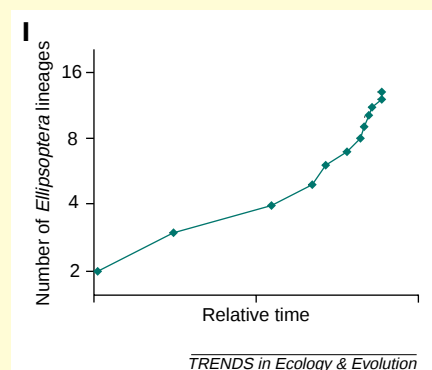
Fig. 1. Lineages-through-time plot for studying speciation rates. The log of the number of lineages is plotted against the relative time of each node since the root node (*) (other graphical representations are discussed in Ref. 10). Under the simplest model (the constant speciation rate model), where the probability of a speciation event occurring in a given time is constant both over time and among species, a straight line with slope equal to the per lineage speciation rate, b is expected. This corresponds to the pure birth stochastic process, an early statistical model that first arose in precisely this context¹⁰. The maximum likelihood estimate of b equals the number of reconstructed speciation events that have occurred since the root node, divided by the total lineage time available for such events to occur. Confidence intervals for the estimate based on the fact that only a finite sample of nodes is available can also be calculated^{10,23}. The probability theory underlying statistical inference from phylogenies is closely related, and often identical to population genetics theory used to make inferences from gene genealogies^{9,17}, but there are some differences. For example, population geneticists largely rely on a coalescence approach, which reverses time and imagines the tree shrinking as its branches coalesce at the nodes. However, this approach does not allow one to theorize about trees that grow according to a birth–death process (i.e. with extinction), for which a forward perspective on time is needed⁸.

indistinguishable, the range of diversification models consistent with the data can be narrowed down, thereby gaining better estimates of speciation rates.

At present, few studies have applied these techniques to estimating speciation rates in real clades, possibly mainly because of the rarity of sufficiently complete trees. As more data become available, particularly those demonstrating the evolutionary status of included species, variations of these methods will be used for broad surveys of speciation rates. In addition, links between genealogical approaches and the methods we describe should allow estimation of a broader range of parameters, such as the relative rates of paraphyletic and monophyletic modes of speciation,

Box 3. Departures from the constant speciation rate model

Several processes can lead to departures from the expectations of the constant speciation rate model outlined in Fig. 1. An apparent increase in speciation rate towards the present, as found in the *Ellipsoptera* group of tiger beetles^a (Fig. I), can be caused by: (1) an actual increase in the speciation rate, or (2) an illusory increase caused by a constant background extinction rate, d , that is high relative to the (constant) speciation rate, b (Ref. b). In the latter case, the slope near the present asymptotically approaches the speciation rate, whereas the slope in the main body of the graph equals speciation minus extinction rate (i.e. the net diversification rate). The two explanations might be distinguishable in some circumstances. The upturn as a result of extinction is expected to start at around $1/(b-d)$ time units before the present^c, whereas, for example, if speciation rates increased during the Pleistocene we expect an upturn within the past 2.5 million years^d. The

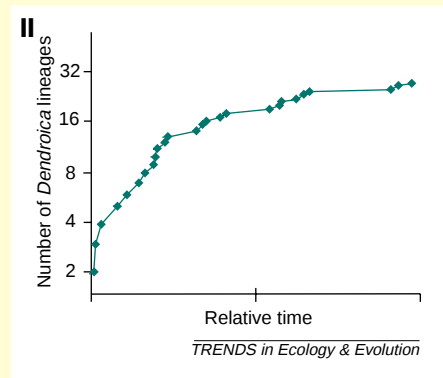


predictions will differ for clades with $b-d < 0.4$ species per million years. In the tiger beetles, the upturn is more recent than expected because of a constant background extinction rate.

A slowdown in the diversification rate towards the present, as found in New World warblers of the genus *Dendroica*^e (Fig. II), can be caused by a decrease in speciation rate or an increase in extinction rate. If background extinction has occurred in the group, we would expect to see a recent upturn above the levelling off stage, but beyond that it is probably not possible to distinguish the two explanations^f.

Sampling artefacts can also influence the patterns we observe. Phylogenies reconstructed from an incomplete sample of species tend to underestimate the number of nodes towards the present, creating an apparent slowdown^g. Methods exist to evaluate the impact of known missing species, assuming those species comprise a random sample^g. However, over-splitting or lumping by taxonomists could also lead to artefactual upturns or slowdowns respectively, if taxonomic species are used as terminals.

Finally, the analyses are sensitive to any biases affecting the methods used to reconstruct node ages. For example, multiple substitutions at the loci used to build the phylogeny could lead to saturation of genetic distances, producing shorter branch lengths deeper in the tree. Maximum likelihood and some distance methods can be used to avoid this problem^h, but there has been little general work on how methods for reconstructing



branch lengths and node ages might bias analysis of diversification rates.

References

- a Barraclough, T.G. *et al.* (1999) Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). *Proc. R. Soc. London B Biol. Sci.* 266, 1061–1067
- b Nee, S. *et al.* (1994) Extinction rates can be estimated from molecular phylogenies. *Philos. Trans. R. Soc. London Ser. B* 344, 77–82
- c Harvey, P.H. *et al.* (1994) Phylogenies without fossils. *Evolution* 48, 523–529
- d Klicka, J. and Zink, R.M. (1999) Pleistocene effects on North American songbird evolution. *Proc. R. Soc. London B Biol. Sci.* 266, 695–700
- e Lovette, I.J. and Bermingham, E. (1999) Explosive speciation in the New World *Dendroica* warblers. *Proc. R. Soc. London B Biol. Sci.* 266, 1629–1636
- f Kubo, T. and Iwasa, Y. (1995) Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* 49, 694–704
- g Pybus, O.G. and Harvey, P.H. (2000) Testing macroevolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. London B Biol. Sci.* 267, 2267–2272
- h Hillis, D.M. *et al.*, eds (1996) *Molecular Systematics*, Sinauer Associates

and the frequency of hybrid speciation. Even without these extensions, the approach offers a means to investigate otherwise unanswerable questions in groups lacking fossil records.

Changes in speciation rate over time

Speciation rates change over time for a variety of reasons and the methods described can be extended to test for such changes. For example, one widespread hypothesis is that glacial cycles during the Pleistocene (2.5 to 0.01 million years ago) increased speciation rates in Northern Hemisphere groups^{25,26}. The main evidence for this is that many species and populations display genetic divergences from their nearest relatives consistent with separation during the late Pleistocene. However, even if speciation rates remained constant, it is expected that many

divergences would be clustered towards the present. It is therefore vital to test for a significant increase in speciation rate during the Pleistocene compared to the null model of constant speciation rate. Such a test was performed using mitochondrial DNA phylogenies of 11 lineages of passerine birds²⁷. It was found that net diversification rates decreased towards the present rather than increased, contrary to the predictions of the Late Pleistocene origins hypothesis.

Many studies have used similar patterns to argue that diversification rates typically decline during the radiation of clades^{28,29}. This might occur because the opportunity for speciation decreases as ecological or geographical space becomes filled³⁰, or as a consequence of decreased range sizes following successive subdivisions of ancestral ranges³¹. However, several artefacts can also create the illusion

of early bursts of 'explosive' speciation. These include incomplete sampling, taxonomic underestimates of the number of evolutionary species, and possible biases in estimates of node ages (Box 3). Even if none of these problems apply, it might be impossible to distinguish falling speciation rates from rising extinction rates³².

Variation in speciation rate among regions

Species richness varies widely among regions, but what causes hotspots of diversity? Phylogenetic data can provide insights into this problem. For example, studies in South American and African tropical birds have used phylogenetic data to infer precisely where in the tropics the most recent speciation occurred^{33,34}. In both cases, they concluded that the phylogenetically youngest species do not tend to be found in the central areas of lowland rain forest, but rather in the surrounding topographically complex, montane areas. This appears to rule out lowland areas as hotbeds of recent speciation.

With complete species-level phylogenies for study groups, it should be possible to estimate how speciation and extinction rates vary among regions. One example is the recent investigation of the evolutionary basis of a species–area relationship, by comparing speciation rates of *Anolis* lizards among Caribbean islands of different sizes³⁵. Future studies could apply similar approaches in continental regions and at a global scale. For example, does high speciation and/or low extinction explain the occurrence of floristic hotspots in Mediterranean climate regions³⁶, and, what are the relative roles of speciation and extinction in causing latitudinal gradients in species richness³⁷?

Variation in speciation rates among taxa

Speciation rates of taxa might depend on their biological characteristics, such as body size or the degree of sexual selection. To date, most tests of such hypotheses have made only limited use of phylogenetic information, simply comparing the numbers of species in sister taxa that differ in the trait of interest⁶. As a result, it cannot be ruled out that extinction, rather than speciation, might have caused the observed patterns. As more species-level phylogenies become available, it will be possible to estimate speciation and extinction rates separately for multiple sister taxa pairs. Similarly, it might be possible to test for the effects of rapidly changing traits on speciation rates, even in the absence of sister taxa differing unambiguously in their expression of the trait³⁸. For hypotheses where the effect of the trait is uncertain, such as body size, these tests will be crucial in narrowing down possible mechanisms.

Causes of speciation

Perhaps the most fundamental question in the study of speciation is what causes a single ancestral species to split into two (or more) daughter species? The

traditional evidence for answering this has been to look at recently split species to infer what processes might have been important. For example, evidence that sexual selection can play a key role in speciation initially came from observations of large differences in SECONDARY SEXUAL TRAITS among closely related species³⁹. Molecular phylogenetics allows more precise tests: species relationships can be more accurately known and information on the timing of speciation allows us to focus on the most recent events. However, a key limitation with this approach is that changes can occur since speciation: any patterns observed for even closely related species could be the incidental outcome of the independent evolutionary histories of those species, rather than indicative of the forces promoting speciation^{1,6}. Also, extinction can affect the characteristics displayed by surviving species. Hence, some way is needed to factor out the effects of incidental change and extinction on the patterns observed. We discuss two areas now exploring this approach.

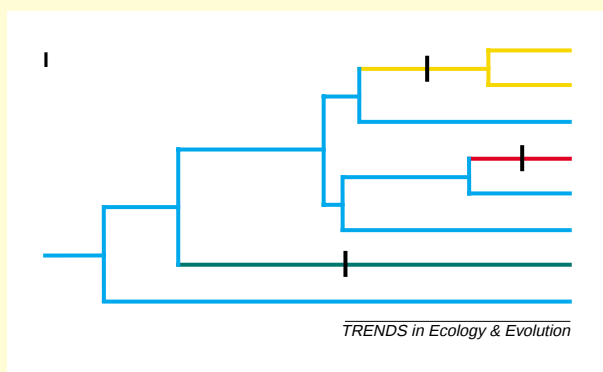
Geographical isolation

Several authors have used species-level phylogenies to assess the relative frequencies of different geographical modes of speciation, based on the geographical distributions of recently formed sister species^{5,7,40}. For example, in a recent study of several bird, insect and fish groups⁴¹, all of the most recent splits had no range overlap, suggesting that ALLOPATRIC SPECIATION is the norm (with some possible exceptions). Also, range size differences between recently split species suggest that speciation often involves the isolation of small populations, so-called peripatric speciation. Other measures can be used to explore other modes, such as the frequency of PARAPATRIC SPECIATION, or the role of external boundaries, such as rivers and mountain ranges⁴².

The Achilles heel of these studies is that ranges of species can move, even over very short timescales. If range movements are common, present-day ranges might not preserve a record of the mode of speciation. In the above study, several clades showed the same qualitative pattern of range overlaps as would be expected if species ranges had moved to random locations within the area occupied by the whole clade⁴¹. Recent work has tried to deal with this problem in several ways: (1) by looking at situations where range movements are unlikely to cloud the results, such as speciation on small oceanic islands⁴³; (2) by checking the success of biogeographical methods in cases known to have a particular mode of speciation, for example presumed SYMPATRIC SPECIATION of polyploid plants⁷; or (3) by proposing tests for historical signals, for example randomizing ranges of species among the tips of the phylogeny and comparing observed overlaps of sister species to those obtained between random pairs of species⁴¹. Although choosing a realistic null model can be difficult⁴⁴, the third approach could offer a general way forward.

Box 4. Character shifts and speciation

The locations of shifts in ecological characters can be reconstructed onto a tree using maximum parsimony or likelihood methods^a. Figure I shows a hypothetical example of a clade with three reconstructed changes (marked by vertical bars) in the state of a discrete character. If ecological shifts occur incidentally over time with no effects on speciation or extinction, we expect shifts to be distributed across the phylogeny roughly in proportion to the time available for them to occur, that is, the branch lengths of the tree^{b,c}. If shifts are associated with speciation, we expect to see: (1) more ecological differences



between recently split species than expected, given average rates across the tree and the short divergence time of those splits^d; and (2) that the amount of change within lineages is more directly proportional to the number of nodes than to branch lengths^c. If ecological differences are associated instead with the long-term persistence and subsequent radiation of lineages, we might expect that most shifts would be observed between more distantly related lineages within the clade^d.

Any test of these predictions must take into account that the location of shifts are reconstructed and are therefore prone to uncertainties or biases of the reconstruction method^a. Other discrete or continuous species traits thought to be involved in speciation, such as reproductive morphology or genetic characters, could be treated in the same way.

References

- a Cunningham, C.W. and Omland, K.E. (1998) Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol. Evol.* 13, 361–366
- b Pagel, M. (1997) Inferring evolutionary processes from phylogenies. *Zool. Sci.* 26, 331–348
- c Mooers, A.O. *et al.* (1999) Using phylogenies to test macroevolutionary hypotheses of trait evolution in Cranes (Gruinae). *Am. Nat.* 154, 249–259
- d Barraclough, T.G. *et al.* (1999) Testing whether ecological factors promote cladogenesis in a group of tiger beetles (Coleoptera: Cicindelidae). *Proc. R. Soc. London B Biol. Sci.* 266, 1061–1067

Also, tests for particular modes can look at additional predictions, for example that peripatric speciation should lead to imbalanced trees⁴⁵ or that sympatric speciation should involve host shifts⁶.

Extinction can also affect geographical patterns, but its effects have been little explored. Random extinction occurring at a constant rate over time tends to prune older lineages¹⁷, leaving many recent speciation events 'untouched', and therefore could have a fairly minor effect. However, selective extinction, for example for species with smaller ranges or those that interact with large numbers of close relatives, could have a much larger effect, perhaps even biasing evidence for geographical modes. Another source of bias might come from taxonomic artefacts, for example if taxonomists were more likely to recognize taxa as species when they are sympatric with their siblings.

Despite the problems, phylogenetics provides a framework for considering the effects of other processes and has led to refinements of existing tests. Further refinements should come when phylogeographical data are available for all the species in a higher clade. This should make it possible to determine, for example, what proportion of populations go on to form reproductively isolated and morphologically recognizable species, and what distinguishes the winners from the losers.

Ecological shifts

Recent work has renewed interest in the role of ecological shifts in speciation, and phylogenetics provides a means to test the generality of these

findings at scales where detailed experiments are not possible (see Schluter⁴⁶, this issue). The basic approach is to map ecological characters onto a tree (Box 4), with the prediction that recently split species will tend to occupy different niches or habitats. For example, a large body of work on herbivorous insect groups has shown that closely related species are often found on different host-plant species^{13,47}. Similarly, phylogenetic evidence supports a possible role of habitat shifts in several Hawaiian groups, including silversword plants⁴⁸, *Drosophila*, and spiders from the genus *Tetragnatha*⁴⁹. However, except in the most extreme cases, there is again the problem that ecological differences between even closely related species could have evolved since speciation, as a by-product of independent evolutionary history.

One solution is to compare observed ecological differences among species to those expected under a null model in which ecological traits evolve randomly through time with no effect on speciation (Box 4). This approach is stricter in the criteria needed to demonstrate that ecological shifts are linked to speciation, but there remain limitations. Characters that change predominantly at speciation events need not play a causal role in the speciation process: the evolution of some characters might be speeded up by conditions during speciation, such as small population size⁵⁰. Also, even using null models it might be difficult to distinguish rapidly evolving characters from those involved in speciation. Hence, theoretical and experimental evidence is vital for supporting the likelihood that a given trait is involved

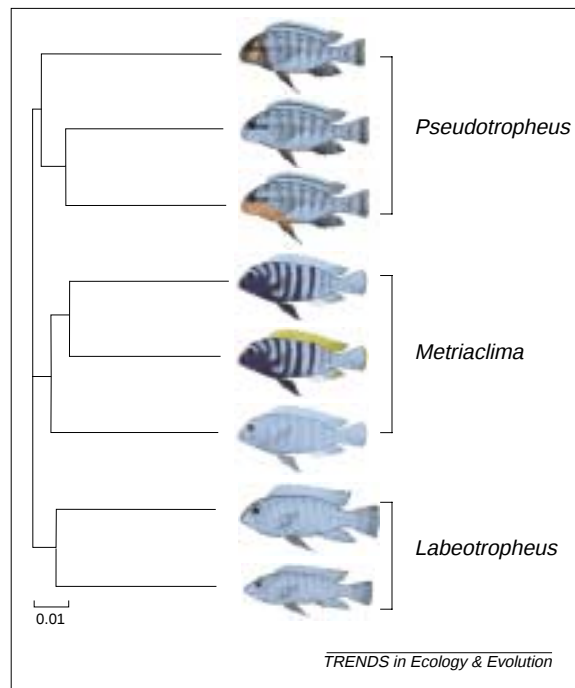


Fig. 2. Speciation of rock-dwelling cichlid fish (mbuna) in Lake Malawi. Ecological diversification has been thought to be important in the speciation of cichlids in the African Great Lakes, but among a group of mbuna from Lake Malawi, ecological differences in jaw morphology and associated feeding style are observed primarily between distantly related species (i.e. between genera²⁸). For example, *Pseudotropheus* spp. use their downward-pointing mouths to pinch off mouthfuls of algae, whereas *Metriacrima* spp. use their terminal mouths to comb food from attached algae and *Labeotropheus* spp. use their robust sub-terminal jaws to scrape algae from rocks. Field studies confirm a frequent lack of resource partitioning within genera and that closely related species use broadly overlapping resources⁵¹. Hence, although ecological divergence might have played a role in the early radiation or long-term persistence of lineages, it appears to have been less important for recent speciation events that led to the majority of diversity in the lake. Instead, closely related species often differ in male nuptial colouration, suggesting that female choice for male colouration might be a more important factor in recent speciation. A similar pattern has been found in Nicaraguan crater-lake cichlids¹⁵. Adapted, with permission, from Ref. 28. The tree is based on amplified fragment length polymorphisms (AFLP), a DNA fingerprinting technique. Scale bar = 1% difference in the AFLP profile.

in speciation. More work is needed to establish a strong statistical framework for these kinds of tests, but, in some cases, interpretation can be more straightforward. For example, ecological shifts appear to have played little role in recent speciation events of Lake Malawi rock-dwelling cichlids, because ecological differences are primarily observed between distantly related taxa²⁸ (Fig. 2).

These examples show the potential and some pitfalls of using phylogenies to investigate causes of speciation. Studies that deal critically with the problems have the potential to answer major questions concerning the generality of mechanisms outlined by theory or experiments. For example, what are the relative roles of geographical isolation and DIVERGENT SELECTION between environments in promoting speciation? What are the relative roles of species interactions⁵² and ecological shifts⁵³ in

promoting REPRODUCTIVE ISOLATION and diversity in sexually selected characters? What role do genetic processes, such as changes in chromosome number, gene rearrangements and duplications, play in speciation⁵⁴? Considering the evolution of genetic characters with respect to phylogeny will allow estimates of the relative frequencies of different types of change and tests for their intimate association with speciation. Work on these questions has already started, but major advances should be seen over the next few years, as more complete reconstructions of evolutionary relationships in clades become available.

Linking genealogical and phylogenetic scales

The previous discussion has shown how broad-scale phylogenetic studies can help to identify general trends in speciation, but a potential weakness to date has been that most studies at this level rely on taxonomic species as surrogates for evolutionary species. Current technology for DNA sequencing means it is becoming feasible to sample genealogies of individuals within all the species of a higher taxonomic group, such as a genus. These data will allow tests of the validity of using taxonomic species in speciation studies, but more importantly will allow a broader range of questions to be answered. By identifying boundaries between evolutionary species, it will be possible to ask what features are associated with species boundaries, whether the multifarious aspects of species tend to evolve in concert, and what role the external environment plays in producing genetic and phenotypic clusters⁵⁵. Two areas would seem particularly hard to address without integrating genealogical and phylogenetic approaches: the role of hybridization in speciation, and speciation in asexually reproducing lineages.

Hybridization and speciation

Hybrid speciation occurs when hybridization between two species leads to the formation of a new, third species. It has been long considered important in plants, with 11% of plant species richness attributed to this mode by recent authors¹⁹, but it might also play a role in animals⁵⁶. The main tool of current research is detailed genetic analysis, but phylogenetics could play a key role in the future resolution of the general prevalence of hybrid speciation. Current work suggests two ways in which such tests might proceed. First, extensions of traditional methods of phylogeny reconstruction could be used to reconstruct detailed histories of hybridization and CLADOGENESIS in terms of networks^{57,58}. As yet, the practicalities of this approach for large data sets are uncertain: allowing for lateral connections among taxa increases the number of possible solutions among which to search for the optimum. The second approach is to estimate the frequency of hybridization without reconstructing

an explicit history of those events. Current methods have been developed to quantify levels of recombination among groups of bacterial and fungal sequences^{59,60}, but similar tests could be applied to hybridization between species. A problem common to both approaches is that it might be hard to distinguish hybrid formation of a new species from gene flow between two existing species that does not lead to formation of a third species.

Asexual speciation?

Sexual reproduction might explain the existence of the discrete units recognized as species: interbreeding maintains coherence within populations, whereas reproductive isolation leads to genetic and phenotypic discontinuities among isolated populations (see Turelli *et al.*⁶¹, this issue). However, if similar genetic and morphological clusters were found in asexual taxa, external ecological factors might be more important than is sexual reproduction in explaining why species exist. One way to test this idea would be to compare genealogical histories between closely related sexual and asexual taxa. The data could be used to: (1) identify whether discrete genetic and morphological clusters exist in asexual taxa⁶²; (2) test whether the degree of clustering is more extreme than expected simply from a stochastic birth–death model of asexual division⁶³; (3) compare the rate of origin of observed clusters between sexual and asexual taxa; and (4) test what role divergent selection from environmental variables might play in the process.

Conclusions

Our aim has been to show the enormous potential of molecular phylogenetics for answering long-standing

questions about speciation. Current studies have been caught between the competing demands of sampling enough species to detect any patterns and establishing the evolutionary status of the included species. It can still be difficult to reconstruct an accurate species phylogeny for higher groups, but ongoing practical and theoretical advances are making it feasible to meet both demands in a single study. Even so, there is still the need for studies that sacrifice detail for scale, and vice versa.

Assuming a suitable tree is available, theoretical questions remain about how best to extract information on speciation and in particular how to rule out the effects of other processes such as extinction. The basic theory on rates of speciation is well established, but extensions are needed to answer questions at finer and broader scales, for example, what proportion of populations go on to form fully fledged species, and how does the probability of speciation vary geographically? The theory for investigating the causes of speciation is much more diffuse, and a stronger framework is needed before the generality of mechanisms outlined in this special issue can be firmly assessed. Crucially, ways are needed to identify when the signal of speciation has been lost entirely from present species.

Finally, we have outlined how major advances over the next few years will probably result from bridging the gap between genealogical and phylogenetic scales. Genealogies allow accurate reconstruction of species histories and insights into population processes, but a phylogenetic perspective is needed to identify general trends and to consider the full array of processes leading to species diversity within clades.

Acknowledgements

We thank C. Albertson for use of his line drawings of cichlids, and C. Albertson, N. Barton, V. Savolainen, A. Vogler and three referees for comments on this article. TGB is funded by a Royal Society University Research Fellowship.

References

- Mayr, E. (1963) *Animal Species and Evolution*, Harvard University Press
- Panchen, A.L. (1992) *Classification, Evolution and the Nature of Biology*, Cambridge University Press
- Benton, M.J. and Pearson, P.N. (2001) Speciation in the fossil record. *Trends Ecol. Evol.* 16, 405–411
- Hennig, W. (1966) *Phylogenetic Systematics*, University of Illinois Press
- Brooks, D.R. and McLennan, D.A. (1991) *Phylogeny, Ecology, and Behaviour: a Research Program in Comparative Biology*, Chicago University Press
- Barraclough, T.G. *et al.* (1998) Revealing the factors that promote speciation. *Philos. Trans. R. Soc. London Ser. B* 353, 241–249
- Berlacher, S.H. (1998) Can sympatric speciation be proven from phylogenetic and biogeographic evidence? In *Endless Forms: Species and Speciation* (Howard, D.J. and Berlacher, S.H., eds), pp. 99–113, Oxford University Press
- Hey, J. (1992) Using phylogenetic trees to study speciation and extinction. *Evolution* 46, 627–640
- Purvis, A. (1996) Using interspecies phylogenies to test macroevolutionary hypotheses. In *New Uses for New Phylogenies* (Harvey, P.H. *et al.*, eds), pp. 153–168, Oxford University Press
- Nee, S. On inferring speciation rates from phylogenies. *Evolution* (in press)
- Avise, J.C. (2000) *Phylogeography: the History and Formation of Species*, Harvard University Press
- Hey, J. (2001) The mind of the species problem. *Trends Ecol. Evol.* 16, 326–329
- Sato, A. *et al.* (1999) Phylogeny of Darwin's finches as revealed by mtDNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5101–5106
- Scheffer, S.J. and Wiegmann, B.M. (2000) Molecular phylogenetics of the holly leaf miners (Diptera: Agromyzidae: *Phytomyza*): species limits, speciation, and dietary specialization. *Mol. Phylog. Evol.* 17, 244–255
- Wilson, A.B. *et al.* (2000) Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. *Proc. R. Soc. London B Biol. Sci.* 267, 2133–2141
- Hillis, D.M. *et al.*, eds (1996) *Molecular Systematics*, Sinauer Associates
- Nee, S. *et al.* (1994) Extinction rates can be estimated from molecular phylogenies. *Philos. Trans. R. Soc. London Ser. B* 344, 77–82
- Omland, K.E. *et al.* (2000) Cryptic genetic variation and paraphyly in ravens. *Proc. R. Soc. London B Biol. Sci.* 267, 2475–2482
- Rieseberg, L.H. (1997) Hybrid origins of plant species. *Annu. Rev. Ecol. Syst.* 28, 359–389
- Nichols, R.A. (2001) Gene trees and species trees are not the same. *Trends Ecol. Evol.* 16, 358–364
- Avise, J.C. and Walker, D.E. (1999) Species realities and numbers in sexual vertebrates: Perspectives from an asexually transmitted genome. *Proc. Natl. Acad. Sci. U. S. A.* 96, 992–995
- Trewick, S.A. (2000) Mitochondrial DNA sequences support allozyme evidence for cryptic radiation of New Zealand *Peripatoides* (Onychophora). *Mol. Ecol.* 9, 269–281
- Baldwin, B.G. and Sanderson, M.J. (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. U. S. A.* 95, 9402–9406
- Purvis, A. *et al.* (1995) Macroevolutionary inferences from primate phylogeny. *Proc. R. Soc. London B Biol. Sci.* 260, 329–333
- Avise, J.C. and Walker, D. (1998) Pleistocene phylogeographic effects on avian populations and the speciation process. *Proc. R. Soc. London B Biol. Sci.* 265, 457–463
- Klicka, J. and Zink, R.M. (1999) Pleistocene effects on North American songbird evolution. *Proc. R. Soc. London B Biol. Sci.* 266, 695–700
- Zink, R.M. and Slowinski, J.B. (1995) Evidence from molecular systematics for decreased avian

- diversification in the Pleistocene epoch. *Proc. Natl. Acad. Sci. U. S. A.* 92, 5832–5835
- 28 Albertson, R.C. *et al.* (1999) Phylogeny of a rapidly evolving clade: the cichlid fishes of Lake Malawi, East Africa. *Proc. Natl. Acad. Sci. U. S. A.* 96, 5107–5110
 - 29 Lovette, I.J. and Bermingham, E. (1999) Explosive speciation in the New World *Dendroica* warblers. *Proc. R. Soc. London B Biol. Sci.* 266, 1629–1636
 - 30 Schluter, D. (2000) *The Ecology of Adaptive Radiation*, Oxford University Press
 - 31 Rosenzweig, M.L. (1996) *Species Diversity in Space and Time*, Cambridge University Press
 - 32 Kubo, T. and Iwasa, Y. (1995) Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* 49, 694–704
 - 33 Fjeldsa, J. (1994) Geographical patterns for relict and young species of birds in Africa and South America and implications for conservation priorities. *Biodiv. Conserv.* 3, 207–226
 - 34 Roy, M.S. (1997) Recent diversification in African greenbul (Pycnonotidae: *Andropadus*) supports a montane speciation model. *Proc. R. Soc. London B Biol. Sci.* 264, 1337–1344
 - 35 Losos, J.B. and Schluter, D. (2000) Analysis of an evolutionary species-area relationship. *Nature* 408, 847–850
 - 36 Cowling, R.M. *et al.* (1996) Plant diversity in Mediterranean-climate regions. *Trends Ecol. Evol.* 11, 362–366
 - 37 Chown, S.L. and Gaston, K.J. (2000) Areas, cradles and museums: the latitudinal gradient in species richness. *Trends Ecol. Evol.* 15, 311–315
 - 38 Pagel, M. (1997) Inferring evolutionary processes from phylogenies. *Zool. Sci.* 26, 331–348
 - 39 Darwin, C. (1871) *The Descent of Man and Selection in Relation to Sex*, John Murray
 - 40 Chesser, R.T. and Zink, R.M. (1994) Modes of speciation in birds – a test of Lynch's method. *Evolution* 48, 490–497
 - 41 Barraclough, T.G. and Vogler, A.P. (2000) Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.* 155, 419–434
 - 42 Collins, A.C. and Dubach, J.M. (2000) Biogeographic and ecological forces responsible for speciation in *Ateles*. *Int. J. Primatol.* 21, 421–444
 - 43 Coyne, J.A. and Price, T.D. (2000) Little evidence for sympatric speciation in island birds. *Evolution* 54, 2166–2171
 - 44 Gotelli, N.J. and Graves, G.R. (1996) *Null Models in Ecology*, Smithsonian Institution
 - 45 Chan, K.M.A. and Moore, B.R. (1999) Accounting for mode of speciation increases power and realism of tests of phylogenetic asymmetry. *Am. Nat.* 153, 332–346
 - 46 Schluter, D. (2001) Ecology and the origin of species. *Trends Ecol. Evol.* 16, 372–380
 - 47 Bush, G.L. and Smith, J.J. (1998) The genetics and ecology of sympatric speciation: a case study. *Res. Popul. Ecol.* 40, 175–187
 - 48 Baldwin, B.G. (1997) Adaptive radiation of the Hawaiian silversword alliance: congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations. In *Molecular Evolution and Adaptive Radiation* (Givnish, T.J. and Sytsma, K.J., eds), pp. 103–128, Cambridge University Press
 - 49 Roderick, G.K. and Gillespie, R.G. (1998) Speciation and phylogeography of Hawaiian terrestrial arthropods. *Mol. Ecol.* 7, 519–531
 - 50 Cheverud, J.M. *et al.* (1999) Epistasis and the evolution of additive genetic variance in populations that pass through a bottleneck. *Evolution* 53, 1009–1018
 - 51 Genner, M.J. *et al.* (1999) Foraging of rocky habitat cichlid fishes in Lake Malawi: coexistence through niche partitioning? *Oecologia* 121, 283–292
 - 52 Coyne, J.A. and Orr, H.A. (1997) 'Patterns of speciation in *Drosophila*' revisited. *Evolution* 51, 295–303
 - 53 Podos, J. (2001) Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature* 409, 185–188
 - 54 Mascheretti, S. *et al.* (2000) Phylogenetics of the speciose and chromosomally variable rodent genus *Ctenomys* (Ctenomyidae, Octodontidae), based on mitochondrial cytochrome *b* sequences. *Biol. J. Linn. Soc.* 70, 361–376
 - 55 Giessler, S. *et al.* (1999) Morphological evolution and genetic differentiation in *Daphnia* species complexes. *J. Evol. Biol.* 12, 710–723
 - 56 Dowling, T.E. and Secor, C.L. (1997) The role of hybridization and introgression in the diversification of animals. *Annu. Rev. Ecol. Syst.* 28, 593–619
 - 57 Strimmer, K. and Moulton, V. (2000) Likelihood analysis of phylogenetic networks using directed graphical models. *Mol. Biol. Evol.* 17, 875–881
 - 58 Xu, S.Z. (2000) Phylogenetic analysis under reticulate evolution. *Mol. Biol. Evol.* 17, 897–907
 - 59 Smith, J.M. (1999) The detection and measurement of recombination from sequence data. *Genetics* 153, 1021–1027
 - 60 Goddard, M.R. and Burt, A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. U. S. A.* 96, 13880–13885
 - 61 Turelli, M. (2001) Theory and speciation. *Trends Ecol. Evol.* 16, 330–343
 - 62 Taylor, D.J. and Foighil, D.O. (2000) Transglobal comparisons of nuclear and mitochondrial genetic structure in a marine polyploid clam (*Lasaea*, Lasaeidae). *Heredity* 84, 321–330
 - 63 Higgs, P.G. and Derrida, B. (1992) Genetic distance and species formation in evolving populations. *J. Mol. Evol.* 35, 454–465

TREE online – making the most of your personal subscription

High quality printouts (from PDF files)

Links to other articles, other journals and cited software and databases

To claim your FREE online access to *Trends in Ecology and Evolution*:

Go to <http://www.bmn.com/general/subkey> and select *Trends in Ecology and Evolution* from the list

Enter your own BioMedNet login details when prompted (if you are not yet a member, joining takes minutes and is FREE)

Follow the instructions on the *Trends in Ecology and Evolution* page under 'Personal Subscriber Access'

You only need to register once.

For subsequent visits, bookmark: <http://journals.bmn.com>

Tip: If you do not use a shared terminal, you can tick the 'save password' box when you first log on to BioMedNet so that you only need to register once

If you have any questions e-mail: info@current-trends.com

