Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation

Mark Pagel[1] and François Lutzoni[2]

[1] School of Animal and Microbial Sciences, University of Reading,
Reading RG6 6AJ UK
[2] Department of Biology, Duke University,
Box 90338 Durham, NC 27708 USA

Abstract. We describe the application of Markov Chain Monte Carlo (MCMC) methods to two fundamental problems in evolutionary biology. Evolutionary biologists frequently wish to investigate the evolution of traits across a range of species. This is known as a comparative study. Comparative studies require constructing a phylogeny of the species and then investigating the evolutionary transitions in the trait on that phylogeny. A difficulty with this approach is that phylogenies themselves are seldom known with certainty and different phylogenies can give different answers to the comparative hypotheses. MCMC methods make it possible to avoid both of these problems by constructing a random sample of phylogenies from the universe of possible phylogenetic trees for a given data set. Once this sample is obtained the comparative hypotheses can be investigated separately in each tree in the MCMC sample. Given the statistical properties of the sample of trees – trees are sampled in proportion to the probability under a model of evolution – the combined results across trees can be interpreted as being independent of the underlying phylogeny. Thus, investigators can test comparative hypotheses without the real concern that results are valid only for the particular tree used in the investigation. We illustrate these ideas with an example from the evolution of lichen formation in fungi.

1 Introduction

Phylogenetic trees describe the pattern of descent amongst a group of species. With the rapidly accumulating quantities of DNA sequence data, more and more phylogenies are being constructed based upon sequence comparisons (Fig. 1). The combination of phylogenies and statistical models for the analysis of trait evolution provides investigators with a means to reconstruct the probable ancestral states and trajectories of traits as they evolved in the past, and to test hypotheses about correlations among pairs of traits (Pagel, 1997, 1999a). Comparative methods comprise one of biology's most enduring set of techniques for investigating evolution and adaptation (Harvey and Pagel, 1991). They are widely used in evolutionary biology, molecular evolution, animal behaviour, ecology and conservation.

Recent applications of statistical comparative models of trait evolution to phylogenies include reconstructing the nucleotide content of the common ancestor to life on Earth (Galtier, Tourasse, and Gouy, 1999), predicting ancestral

In *Biological Evolution and Statistical Physics* (2002) (eds M. Lässig and A. Valleriani),
Springer, pp. 148-161.

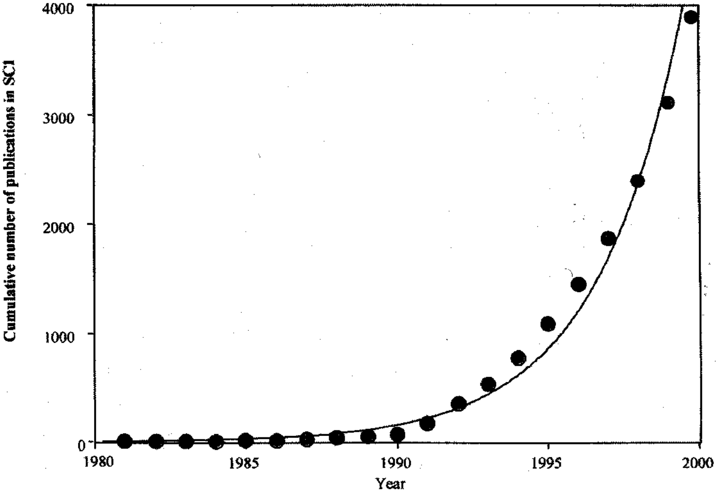

Fig. 1. Plot of the total number of articles reporting the key words "molecular phylogeny" in their title, keywords, or abstract in the years 1981 to 2000. Data from the Science Citation Index. The data are well approximated by exponential curve

ribonuclease enzyme sequences in artiodactyls (Jermann et al. 1995; Schluter, 1995), investigations of parallel molecular evolution in the opsin genes of the visual system (Chang and Donoghue, 2000), detection and reconstruction of ancestral 'signature sequences' that identify common ancestry amongst a group of organisms (van Dijk, et al., 2001), estimation of timings of key events in the history of evolution (Bromham and Rambaut, 1998; Cooper and Penny, 1997; Hedges, et al., 1996), detection of correlated evolution among different sites of the same gene (Krakauer, et al., 1996; Pollock, Taylor, and Goldman, 1999), and of shifts in rates of nucleotide substitution as a consequence of transitions to mutualism in fungi (Lutzoni and Pagel 1997).

Comparative analyses of this kind presume that the phylogeny is known without error. This assumption has long plagued the field because phylogenies are inferences from data, they are subject to error and uncertainty, and different estimates of the phylogenetic tree can return different answers to the comparative question. As a result, all conclusions derived from comparative analyses performed on a single phylogenetic tree are conditional upon that phylogeny.

Attempts to resolve the problem of phylogenetic uncertainty frequently involve trying out the statistical inference on a sample of 'best' or 'favoured' trees, or on all equally most parsimonious trees (e.g., Hibbett, Gilbert, and Donoghue, 2000). Another approach is to construct a consensus tree from some subset of suitably chosen trees. These approaches are all limited by lacking any clear probabilistic basis. The consensus tree is by definition not the most probable tree (there may be some exceptions), and there is no particular reason to believe that the best or most probable tree according to some goodness of fit criterion (such as maximum likelihood or parsimony) is necessarily the true tree. The vast number of possible phylogenetic trees even for small numbers of species or lineages (over 34 million for a tree with only ten tips, over 1076 for a tree of 50 lineages) makes it impractical, save for the smallest cases, to enumerate and analyse all possible trees.

2 Markov-chain Monte Carlo methods

Markov-Chain Monte Carlo Methods Markov-chain Monte Carlo (MCMC) methods offer a formal statistical procedure for taking phylogenetic uncertainty into account in comparative studies. The MCMC approach is to construct a Markov-chain which, if allowed to run long enough, produces states in direct proportion to the equilibrium distribution of states in the model (Gilks, et al. 1996). Applied to phylogenetic trees the goal is to construct a Markov-chain whose states are the possible phylogenetic trees in the universe of all possible phylogenetic trees. If this chain is allowed to reach equilibrium, then successive states of the chain will sample trees in proportion to their 'equilibrium' probabilities, that is, in proportion to their probabilities under some model of evolution.

Given a random sample of phylogenetic trees, the comparative parameters of interest (rates of evolution, ancestral states, trends, correlations, and so on) can then be estimated over the sample, yielding their density distributions. Inferences about trait evolution based upon these distributions will be independent of any particular phylogenetic hypothesis. Owing to characteristics of the MCMC sampling, the statistical distribution of the comparative parameters, if weighted by an appropriate prior probability, can be interpreted as the Bayesian posterior probability distribution.

3 Bayesian statistics and the MCMC approach to phylogenies

MCMC methods implemented in a phylogenetic context, attempt to produce a Markov-chain that if allowed to run long enough will randomly visit sites in the universe of phylogenetic trees in proportion to their probabilities under the model of evolution. If all of these possible sites (trees) are thought *a priori* to be equally likely then the distribution of tree probabilities in the MCMC sample directly estimates the true distribution of tree-probabilities, that is, the probability density distribution of trees.

However, some trees may be thought more or less likely on *a priori* grounds. The MCMC sample then provides a way to update those prior beliefs. Bayesian statistical logic provides the formal framework within which one uses a realised set of outcomes to update a set of prior beliefs, the result being a posterior set of beliefs.

Let S represent a set of aligned gene-sequence data on a set of species, and let ω be a phylogenetic tree, specified by the parameters of a model of gene-sequence evolution and its branch lengths. We may wish to say something about the distribution of ω (that is, about the probability distribution of phylogenetic trees) as a function of S. Formally, Bayes' Rule states that,

$$P(\omega|S) = \frac{P(\omega)P(S|\omega)}{P(S)}, \tag{1}$$

where $P(\omega|S)$ is the posterior probability of ω given S, $P(\omega)$ is the prior probability of ω (in the absence of any knowledge about S), $P(S|\omega)$ is the probability of the sequence data S given ω and $P(S)$ is the probability of the sequence data. $P(S)$ is calculated over all possible trees. In the absence of any prior beliefs is typically set to $1/N$ for all trees where N is the total number of trees.

To accomplish the sampling needed to estimate the frequency distribution or probability density of $p(\omega|S)$, a Markov-chain is constructed whose states are the possible phylogenetic trees. At each step in the chain a new tree is proposed by a tree-proposal algorithm that alters characteristics of the current tree. New trees are accepted or rejected with probabilities determined by the Metropolis-Hastings (M-H) algorithm (Metropolis et al., 1953; Hastings, 1970):

$$\frac{p(S|\omega^*)p(\omega^*)}{p(S|\omega)p(\omega)}\frac{q(\omega^*,\omega)}{q(\omega,\omega^*)}, \tag{2}$$

where asterisks denote 'new' trees (sets of model parameters and branch lengths) and $q(\omega,\omega^*)$ is the probability of moving in the parameter-space from ω to ω^*.

If the M-H ratio is greater than 1, the new tree is accepted with probability 1. If the ratio is less than 1, the new tree is accepted with probability equal to the ratio. If the new tree is not accepted the chain remains in the 'old' state. The M-H algorithm ensures that the Markov-chain, if allowed to run long enough, visits successive states in the universe in proportion to their likelihoods. The chain is run until a large number (say 500,000) of trees is generated, from which a smaller number (say 5000) is sampled to ensure independence among successive trees in the chain. Then the distribution of ω is estimated from the smaller sample.

The emphasis on estimating the distribution of ω rather than on finding the single best (e.g., highest likelihood, most parsimonious, shortest distance) tree distinguishes MCMC approaches from conventional 'single-tree' studies. MCMC-phylogenetic methods do not seek the best tree, rather they seek to sample in an unbiased way from the probability distribution of trees. The logic underlying this approach is that there is no particular reason to believe that the best tree under some model of evolution corresponds to the true phylogenetic tree. For example, the 'best' tree under a model of parsimony is the one that yields the fewest evolutionary transitions. However, if events of parallel or convergent evolution occur in a number of independent lineages, then seeking the shortest tree may not return the best estimate of the true phylogeny.

The MCMC approach can be contrasted with existing approaches for sampling the universe of trees. The best known of these procedures is the nonparametric bootstrap (Felsenstein, 1985). The bootstrap procedure derives a sample of phylogenetic trees by resampling repeatedly from S. If the data matrix is of length n sites, a single bootstrap sample is created by sampling n

sites at random with replacement from S. If this is repeated, say, 100 times, and only one most optimal tree was recovered from each resampled data set, then a bootstrap sample of 100 trees can be constructed. There has been much debate about precisely what the formal statistical properties of the bootstrap are (Felsenstein and Kishino, 1993; Newton, 1996). Some authors suggest that it approximates a MCMC sample with a uniform prior (i.e., all trees equally likely *a priori*). Whatever its correct interpretation, the computational effort to create a bootstrap sample of trees using maximum likelihood procedures is prohibitive (Larget and Simon, 1999), and the MCMC procedure is not restricted to uniform priors.

MCMC methods have begun to be used to infer aspects of phylogenies (e.g., Yang and Rannala, 1997; Larget and Simon, 1999; see Lewis, 2001 for a recent review) and to estimate population genetic parameters on genealogies (e.g., Wilson and Balding, 1998). How to combine MCMC methods for phylogenetic inference with comparative methods for investigating evolutionary processes has received very little attention. Our own work on the evolution of lichen-formation and loss of lichenization in fungi (Lutzoni, Pagel, and Reeb 2001) and a demonstration of estimating gains and losses of horned soldiers in aphids (Huelsenbeck et al., 2000), are to our knowledge the only attempts to combine comparative methods and MCMC sampling of phylogenies.

4 MCMC and phylogenetic-comparative methods

The MCMC approach to comparative methods must somehow include the comparative data and associated evolutionary model parameters in the Markov-chain along with the sequence data and model of evolution used to construct the phylogenetic trees. Let D be the set of comparative data and μ be the parameters of the model of evolution used to analyse the comparative data (μ might for example contain variances and covariances, the parameters of correlations and regression coefficients). The comparative data might be a set of quantitative traits such as body size and life history variables, or traits that adopt a finite number of states, such as mating system or diet. The goal is to estimate some feature of the comparative data - for example, the ancestral state of some trait at a specified interior node of the phylogeny - simultaneously accounting for the uncertainty in the phylogeny.

The most obvious way to combine the comparative data analysis with estimation of the phylogeny is to build a Markov-chain that simultaneously samples the comparative parameters and the phylogenies (Huelsenbeck et al., 2000). That is, the parameters in μ and ω are estimated simultaneously on S and D. The Markov-chain is made to traverse simultaneously the space of phylogenetic trees and comparative outcomes (e.g., ancestral states, rates of evolution, etc). This is accomplished by adding a comparative-parameter-proposal mechanism to the usual tree-proposal mechanism. The Metropolis-Hastings algorithm is then applied to accept or reject new combinations of trees and comparative relationships. If left to run long enough, this approach will visit trees and their comparative results in proportion to their joint probabilities in the universe. The posterior distribution of the comparative parameters, such as a correlation or regression, or an ancestral state, can then be calculated directly from the sampled Markov-chain.

Although this is the formal MCMC approach, it may have drawbacks. There is no particular reason to believe that the comparative data will in general provide good information about the phylogeny. Traits are often selected for investigation in comparative studies because they evolve independently a number of times on the tree, a phenomenon known as homoplasy. The more homoplasy a character shows the less information it has about phylogenetic history. Most phylogenetic tree reconstruction algorithms will try to minimise the amount of homoplasy. This will influence the way the Markov-chain traverses the tree-space because combinations of trees and comparative results that return the highest likelihood (least homoplasy) will be preferred.

Another way to describe this problem is that the MCMC algorithm will traverse a different sample of phylogenetic trees depending upon the comparative data used in combination with the gene-sequence data. Different hypothesis tests may be based upon different kinds of samples. The argument in favour of including the comparative data is that any information that is available should be included when searching the tree space.

A second approach to comparative-phylogenetic MCMC samples the phylogenetic trees independently from the comparative outcomes. In this approach a sample of phylogenetic trees is produced by MCMC from a set of aligned gene-sequence data (S). Then the comparative data D are analysed on each tree to derive the posterior distribution of the parameters in μ. The values of mu cannot influence the sampling of phylogenies, but the approach retains the desirable feature that the comparative outcomes are automatically weighted by the probability of a given tree type in the posterior distribution of trees. If the parameters of the comparative model are independent of the phylogenetic tree topology, this method will approximate very closely the Bayesian posterior, and return results similar to those of procedure (i). We have employed this approach in the Lutzoni et al. (2001) investigations and is the technique we shall report below.

5 Application to the evolution of lichen formation

The lichen symbiosis consists of an obligate mutualistic association of a fungus species with an alga, a cyanobacterium, or with species of both photobiont types. We wished to investigate the evolution of lichen symbioses in the Ascomycota fungi. This phylum contains approximately 98% of all known lichens. One of us (FL) obtained sequence data on the small and large subunit nuclear ribosomal DNA (SSU and LSU nrDNA) of 54 fungi species. Fifty two of these species are in the Ascomycota phylum and two Basidiomycota were used as outgroups. In addition, we recorded for each species whether it was lichen-forming or not.

5.1 MCMC phylogenetic tree sampling

We used Markov-chain Monte Carlo (MCMC) methods (Larget and Simon, 1999) to approximate the posterior probability density of phylogenetic trees. Given a set of aligned gene sequences S, and following Bayes' theorem, the posterior probability of the ith tree sampled ω_i is

$$P(\omega_i|S) = \frac{L(S|\omega_i)\,P(\omega_i)}{\sum_{j=1}^{N} L(S|\omega_j)\,P(\omega_j)}, \tag{3}$$

where $L(S|\omega_i)$ is the likelihood of the sequence data given the tree, and $P(\omega_i)$ is the prior probability of the tree (here assumed to be uniform at $1/N$). The summation in the denominator is over the N possible trees for the set of species. The likelihood of the sequence data given the tree is integrated over all possible combinations of branch lengths and parameters in the model of sequence evolution. The posterior probability, $P(\omega_i|S)$, is the probability that tree i would arise given the model of sequence evolution.

For a tree of 54 taxa the summation in the denominator of $P(\omega_i|S)$ is vast. The MCMC procedure is used to approximate $P(\omega_i|S)$ by drawing a random sample of trees. We used the general time reversible model of gene-sequence evolution combined with gamma rate heterogeneity to estimate the likelihood of each tree (Hillis et al. 1996). Following convergence of the Markov-chain we generated 200,000 trees. We excluded information on the state (lichen-forming/non-lichen-forming) of each species from the MCMC sampling procedure to ensure that the distribution of trees was not influenced by this trait. A series of runs using the BAMBE (Larget and Simon, 1999) 'global' and 'local' options was conducted to ensure that the Markov-chain converged to the same region in the universe of trees.

5.2 Reconstruction of gains and losses, and ancestral states

We employed a continuous time Markov-model of trait evolution, as implemented in the computer program Discrete (Pagel, 1994), to investigate the evolutionary rate of gains and losses of lichenization. The Markov- model approach in Discrete permits independent gains (q_{01}) and losses (q_{10}) in each branch of the phylogenetic tree. We calculated these separately for each tree sampled in the MCMC procedure. Because trees are represented in the sample in proportion to their probability, investigating the rates over all trees automatically weights our results by the probability of a particular tree.

The gain and loss parameters q_{01} and q_{10} contain the information required to reconstruct ancestral states. The calculation of the posterior probability of the ancestral state at a node follows Pagel (1999b) using the 'local' method and was found from

$$P(n_i = s_j|D,t) = \frac{L(D|n_i = s_j, t)\,P(n_i = s_j)}{\sum_{j=0}^{1} L(D|n_i = s_j, t)\,P(n_i = s_j)}, \tag{4}$$

where $P(n_i = s_j|D, t)$ is the probability that node i takes state j, given the data (lichen-forming/non-lichen-forming) and phylogenetic tree, t, $L(D|n_i = s_j, t)$ is the likelihood of the data given that node i takes state j on tree t, and $P(n_i = s_j)$ is the prior probability that node i takes state j, here assumed to be 1/2. The summation in the denominator is over the two possible states.

5.3 Convergence of the Markov-chain

Fig. 2 shows a typical run of the Markov-chain of trees, evaluated by their log-likelihoods. The chain quickly ascends from a region of trees that produce very poor fits of the data to the model of evolution, into a region in which the chain reaches an asymptote. The enlarged region of this part of the chain (inset) shows how the converged chain then 'wanders' around tree space, sampling trees as directed by the Metropolis-Hastings algorithm. This serves to underscore that the MCMC procedure is not designed to find the best tree, but rather visits trees in proportion to their probabilities under the model of evolution. We sampled 20,000 trees from 200,000 trees generated from the converged chain. We then removed the first 100 to ensure that no trees were included prior to convergence of the chain. Fig. 3 shows the frequency distribution of log-likelihoods for the remaining 19,900 trees.

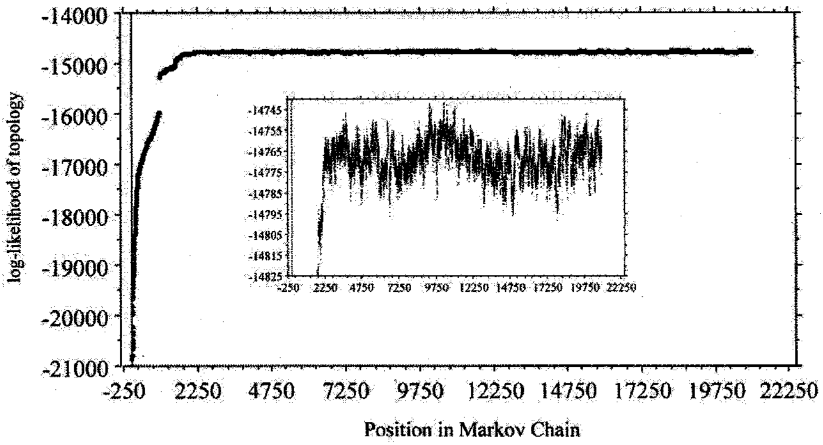

Fig. 2. The convergence of a Markov-chain. The y-axis plots the log-likelihood of successive phylogenetic trees of 52 species of Ascomycota fungi (plus two basidiomycete outgroup species) in the Markov-chain. Likelihoods were calculated from a model of gene-sequence evolution allowing unequal rates of transitions and transversions and allowing for unequal rates of evolution at different sites (Hillis, et al., 1996 [14]). Data were small and large subunit nuclear ribosomal DNA. The inset shows how the converged Markov-chain 'wanders' the tree-space in the converged region.

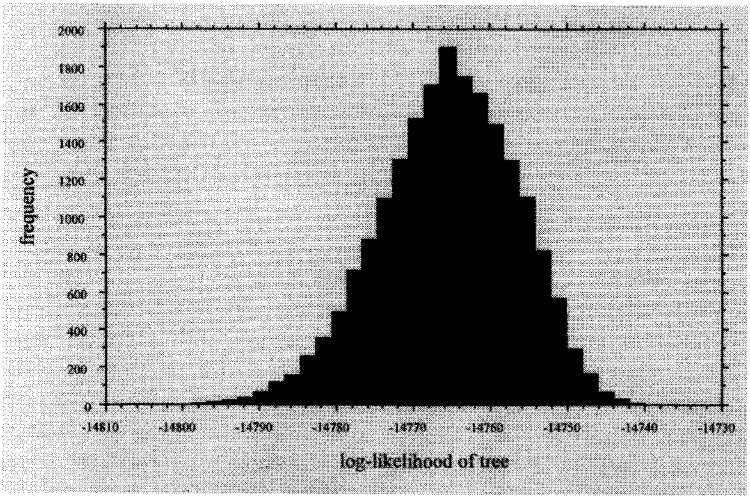

Fig. 3. The frequency histogram of 19,900 trees sampled from the converged Markov-chain.

5.4 Rates of gains and losses of lichenization

Fig. 4 shows that in 18,029 of 19,900 (90.6 %) trees sampled from the Markov-chain the estimated rate of loss exceeds the rate of gain (i.e., the loss/gain ratio is greater than one, and therefore is above the diagonal line in Fig. 4. The average ratio of the rate of loss to the rate of gain is 1.56 ± 0.53 and is positively skewed towards higher ratios (range = 0.56 – 3.24). The highest rates of loss are associated with the lowest rates of gain (Fig. 4); $r = 0.40$) when examined across trees. The ratio of losses to gains is, however, independent of the phylogenetic tree topology (correlation between ratio of rate of losses to gains and log-likelihood of trees = −0.024).

These results indicate that approximately 1.5 times as many losses of the lichen symbiotic state as gains are expected to have occurred during the evolution of the Ascomycota. This conclusion can be drawn without reference to any given phylogenetic tree. Previous work on the evolution of the lichen symbiosis based upon single phylogenetic trees and non-stochastic models of trait evolution based upon parsimony (Gargas, et al., 1995) suggested that lichens evolved independently on many separate occasions, with few losses.

For purposes of comparison with the conventional 'single-tree' approach to comparative studies, we estimated the same rates of gain and loss of lichenization on the consensus tree of the 19,900 trees. The Table shows the estimates and 95% confidence intervals for the single consensus tree and for the data reported in Fig. 2. The rate of loss of lichenization exceeds that of gains on the consensus tree, but the 95% confidence intervals are wide and overlap. Estimates derived from the MCMC procedure are more similar to each other but confidence intervals are narrower and do not overlap the value of the opposite parameter. Thus, the single-tree approach would not allow one to reject the hypothesis of equal rates

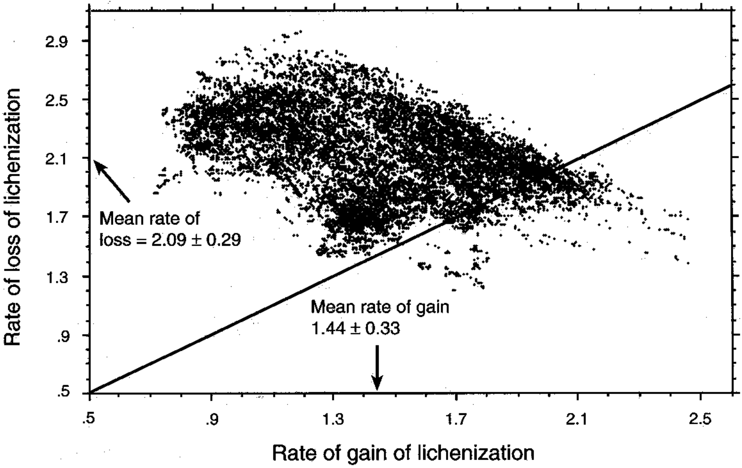

Fig. 4. The rate of loss of lichenization exceeds the rate of gains of lichenization, independently of tree topology. Data are for 19,900 MCMC trees. Solid line is the 1:1 relationship.

Table 1. Rates of gain and loss of lichenization estimated on consensus tree and MCMC sample

Transition Rate	Consensus tree of 19,000 MCMC trees	Separate estimates on each MCMC tree
gain	1.04 (0.05-4.5)	1.44 (0.85-2.05)
loss	2.41 (0.7-5.6)	2.09 (1.55-2.64)

of gains and losses, whereas strong evidence against this hypothesis emerges from Fig. 4.

5.5 Probable phylogenetic position and number of gains and losses

Fig. 1 (right panel) shows the majority-rule consensus phylogenetic tree as derived from our MCMC sample. We show this tree not to propose a particular phylogeny, but to provide a vehicle for identifying events of probable gains and losses of lichenization. The numbers above each internal branch correspond to the node to which a branch points. These numbers represent the proportion of trees in our MCMC sample in which that node was observed. Nodes with values of 100 define a collection of species all of which and only those of which appeared in every one of the trees sampled from the Markov-chain. Other nodes were less certain.

We reconstructed the most probable ancestral states (Pagel 1999b) of fourteen nodes. These nodes identify groups of lichen-forming and non-lichen-forming species in such a way as to make it possible to put reasonable upper and lower limits on the number of independent gains and losses of lichenization. Based upon these reconstructions, we can infer that green areas of the tree are regions of lichen symbiosis, red areas are regions in which the ancestral state is uncertain, and the remaining (uncoloured) branches correspond to non-lichenized regions of the tree.

The left panel of Fig. 5 plots the ancestral state for each of these labeled nodes separately across the 19,900 sampled trees (average probability across all trees labeled on right y-axis of each plot). The reconstructed ancestral state is independent of the phylogeny for some nodes, but for others, the particular tree topology can exert a substantial effect.

The pattern of reconstructed ancestral states implies that a minimum of one and a maximum of three gains of lichenization occurred during the evolution of the Ascomycota. If lichen formation evolved immediately after node 65 (labeled # on phylogeny), then one gain of lichenization is implied for the Ascomycota. Two origins of lichen symbiosis are implied (labeled * on phylogeny) if lichenization evolved independently at node 89 and again at the base of the clade that includes the Lichinales (LI), Arthoniales (AR) and Pyrenomycetes-Dothideales (PD). Three independent gains are implied if the closely related AR and LI groups independently ėvolved lichenization (labelled † on phylogeny).

By comparison, a minimum of three and possibly four losses of the lichen symbiosis have occurred in the Ascomycota. Nodes 93, 97, and 101 high posterior probabilities of being lichenized (left panel of Fig. 5), and each is followed by an unambiguous loss of the lichen symbiosis. For these three nodes the MCMC procedure leaves little doubt that a loss of lichenization quickly followed in the descendant species. A fourth loss of lichenization is implied at the base of the PD group if lichen-formation indeed originated at the points labeled '#' or '*' on the tree.

6 Conclusions and discussion

We have shown how to combine estimation of the phylogenetic tree with a statistical model of trait evolution to account for phylogenetic uncertainty when investigating historical events of evolution. MCMC sampling makes it possible to derive the posterior probability distribution of parameters that are relevant to testing hypotheses of evolution and adaptation. This can add statistical power to inferences and allows one to distinguish between relatively certain and less certain results about the evolution of a given trait.

Our results for the evolution of the lichen symbiosis overturn the conventional wisdom that lichens evolved independently on many separate occasions. Rather, our results suggest that lichens evolved earlier than previously believed and that some of the major fungal lineages that are strictly composed of non-lichenized extant species are derived from lichen ancestors. The minimum of three losses of the lichen symbiosis that we have identified indicate that entire orders of non-lichen-forming fungi are in fact derived from lichen-forming ancestors. This result serves to emphasize that an important distinction must be drawn between ancestrally non-lichen forming and secondarily derived non-lichen-forming fungi. Intriguingly, many of the non-lichen forming Ascomycota fungi that have impor-

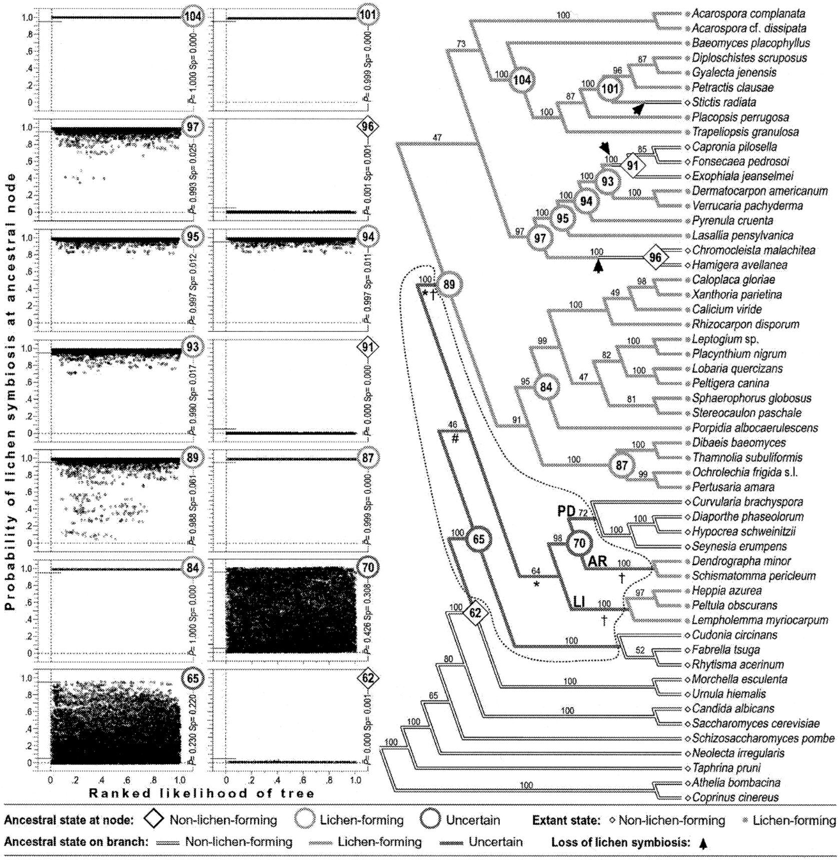

Fig. 5. Bayesian posterior probabilities for reconstructed evolution of the lichen symbiosis and for each node of the Ascomycota phylogeny. Numbers within each of the node symbols (e.g., 62, 65, ... 104) refer to specific nodes and connect nodes in the tree with their respective graphs. Left panel (set of 14 graphs), reconstructed probability that ancestral state was lichen-forming at specified node calculated on each of 19,900 trees generated by MCMC sampling. The average probability and standard deviation are provided on the y-axis to the right of each graph. Right panel (phylogeny), Ascomycota majority-rule consensus of 19,900 MCMC sampled trees based on SSU and LSU nrDNA sequences. Numbers above each internal branch correspond to the posterior probability (%) of the node to which it points. The region of the tree for which the ancestral states of branches could not be extrapolated, because of uncertainty associated with specific nodes, is delimited by a dotted line. The pattern of ancestral states indicates that lichenization has been gained and then lost in the same tree. See text for description of symbols (#, *, +) associated with various evolutionary scenarios for gains of the lichen symbiosis. Pagel and Lutzoni Phylogenetic Uncertainty in Comparative Studies page 15.

tant medical or health benefits to humans are from this group of secondarily derived non-lichen forming species (Lutzoni, Pagel and Reeb, 2001).

MCMC methods are relatively new to biology and in particular to phylogenetics and comparative methods. Owing to the vast number of possible phylogenetic trees for samples of even moderate numbers of species, MCMC methods cannot always be counted on to converge to the optimal region of the universe (e.g., Larget and Simon, 1999). New developments in MCMC sampling, notably Metropolis-coupled MCMC (or MCMCMC) may improve convergence especially in large samples (Gilks and Roberts, 1996).

Acknowledgements

This work was supported by a grant from the Leverhulme Trust, the Natural Environment Research Council and the Biology and Biotechnology Research Council of the UK (MP). F.L. is supported by the US National Science Foundation (DEB-9615542). The molecular work was performed at The Field Museum's Pritzker Laboratory of Molecular Systematics and Evolution, operated with support from an endowment from the Pritzker Foundation.

References

1. Bromham, L., Rambaut, A., Fortey, R., Cooper, A. & Penny, D. Testing the Cambrian explosion hypothesis by using a molecular dating technique. Proc. Natl. Acad. Sci. USA **95**, 12386- 12389, 1998.
2. Chang BSW and Donoghue MJ. Recreating ancestral proteins. Trends in Ecology and Evolution **15**, 109-114, 2000
3. Cooper, A. & Penny, D. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science **275**, 1109-1113, 1997.
4. Felsenstein, J. Confidence limits on phylogenies - an approach using the bootstrap. Evolution **39**, 783-791, 1985.
5. Felsenstein, J. and Kishino, H. Is there something wrong with the bootstrap? A reply to Hillis and Bull. Syst. Biol. **42**, 193-200, 1993.
6. Galtier, N., Tourasse, N., & Gouy, M. A nonhyperthermophilic common ancestor to extant life forms. Science **283**, 220-221, 1999.
7. Gargas, A., DePriest, P.T., Grube, M. &Tehler, A. Multiple origins of lichen symbioses in fungi suggested by SSU rDNA phylogeny. Science **268**, 1492-1495, 1995.
8. Gilks, W.R. and Roberts, G.O. Strategies for improving MCMC. In, Markov Chain Monte Carlo in Practice (Gilks, W.R., Richardson, S., Spiegelhalter, D.J. eds). Chapman and Hall, 1996.
9. Gilks, W.R., Richardson, S., Spiegelhalter, D.J. Introducing Markov chain Monte Carlo. In, Markov Chain Monte Carlo in Practice (Gilks, W.R., Richardson, S., Spiegelhalter, D.J. eds). Chapman and Hall, 1996.
10. Harvey, P.H. & Pagel, M. The comparative method in evolutionary biology. Oxford: Oxford University Press, 1991.
11. Hastings, W. Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**, 97-109, 1970.
12. Hedges, S.B., Parker,P.H., Sibley, C.G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. Nature **381**, 226-229, 1996.

13. Hibbett DS, Gilbert LB, and Donoghue MJ. Evolutionary instability of ectomycorrhizal symbioses in basidiomycetes. Nature **407**, 506-508, 2000
14. Hillis, D.M., Moritz, C. & Mable, B.K. Molecular Systematics, 2nd edition. Sinauer: Sunderland, Ma. , 1996.
15. Huelsenbeck, J., Rannala, B., and Masly, J.P. Accmodating phylogenetic uncertainty in evolutionary studies. Science **288**, 2349-2350, 2000.
16. Jermann, T.M., Opitz, J.G., Stackhouse, J. & Benner, S.A. Reconstructing the evolutionary history of the artiodactyl ribonulcease superfamily. Nature **374**, 57-59, 1995.
17. Krakauer D.C., Pagel M., Southwood T.R.E., et al. Phylogenesis of prion protein. Nature **380**, 675-675, 1996
18. Larget, B. & Simon, D.L. Markov chain monte carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. **16**, 750-759 , 1999
19. Lewis, P. O. Phylogenetic systematics turns over a new leaf. Trends in Ecology and Evolution **16**, 30-37, 2001.
20. Lutzoni, F. and Pagel, M. Accelerated molecular evolution as a consequence of transitions to mutualism. Proc. Natl. Acad. Sci. USA **94**, 11422-11427, 1997.
21. Lutzoni, F. ,Pagel, M., and Reeb, V. 2001. Major fungal lineages derived from lichen-symbiotic ancestors. Nature, **411**, 937-940.
22. van Dijk, M. A.M., Madsen, O., Catzeflis, F., Stanhope, M.J., de Jong, W.W. and Pagel, M. Protein sequence signatures support the 'African clade' of mammals. Proceedings of the National Academy of Sciences **98**, 188-193, 2001.
23. Metropolis, N., et al. Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087-1092, 1953.
24. Newton, M. Bootstrapping phylogenies: large deviations and dispersion effects. Biometrika **83**, 315-328, 1996.
25. Pagel M. Inferring the historical patterns of biological evolution. Nature **401**, 877-884, 1999a.
26. Pagel, M. Detectinq correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society (B) **255**, 37-45, 1994.
27. Pagel, M. Inferring evolutionary processes from phylogenies. Zoologica Scripta **26**, 331-348, 1997.
28. Pagel, M. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Syst. Biol. **48**, 612-622, 1999b.
29. Pollock DD, Taylor WR, and Goldman N. Coevolving protein residues: Maximum likelihood identification and relationship to structure. J. Mol. Biol. **287**, 187-198, 1999.
30. Yang, Z. and Rannala, B. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol. Biol. Evol. **14**, 717-724, 1997.
31. Schluter, D. Uncertainty in ancient phylogenies. Nature **377**, 108-109,1995.
32. Wilson, I, and Balding, D. Genealogical inference from microsatellite data. Genetics **150**, 499- 510, 1998.