most of the growing season but, following certain environmental cues, amictic females produce mictic eggs. These mictic females produce haploid eggs meiotically that develop into either males or, if fertilized, resting eggs. Mictic females can only be fertilized within a few hours of birth.

Therefore, in Vollrath’s example a Brachionus rubens female does not fit Fisher’s theory of sex allocation in a simple way because she does not invest half of her reproductive resources into males and half into females. A rotifer female produces either male or female offspring, but not combinations of both. An amictic female produces only daughters, whereas a mictic female produces either males or resting eggs. As a result, the ratio of the investment in males and amictic females has nothing to do with sex allocation theory. Sex allocation should be judged on the basis of the ratio of male-producing mictic females to resting-egg-producing mictic females. An example of the appropriate application of sex-ratio theory to the sexual phase of the rotifer life cycle is presented by Aparicio et al. They showed that sex ratio theory predicts that half of the mictic females are male producing and half resting-egg producing. The equal sex-allocation principle implies, in this case, that half of the mictic females have a female sexual role, receiving sperm and producing resting eggs, and half have a male role of producing sperm. Consequently, Fisher’s sex-ratio principle is applicable to rotifers, but its proper application requires a thorough understanding of the life cycle.

So the question remains: why are mictic rotifiers dwarf? At present we cannot provide a definitive answer, but there are several features of rotifers relevant to explaining this phenomenon. Rotifer males are not only dwarf, they are also haploid and have reduced morphology (e.g., they do not feed). Their development and maturation are faster, they swim quickly and they have a short life span. Furthermore, egg size is around 30% of the adult rotifer size, which suggests that the amount of resources a mother allocates to each egg is a substantial investment. Dwarf rotifer males might be, therefore, a consequence of selection on mothers to produce rapidly as many males as possible.

References
Despite the wide interest in ancestral state reconstruction, there has been a growing number of workers calling for caution\textsuperscript{1,2,6–12}. To quote Swoford and Maddison\textsuperscript{1}: ‘character state reconstructions can provide a powerful mechanism for studying many facets of the evolutionary process. However, the zeal with which these techniques are sometimes advocated belies the complexity of the problem’. Sources of error and sensitivity analyses in parsimony reconstructions Maximum parsimony is by far the most widely used method for reconstructing ancestral character states (Box 1). Because parsimony attempts to minimize the number of evolutionary events, there are at least two conditions under which parsimony can be misleading: when rates of evolution are rapid\textsuperscript{11,12}, and when the probabilities of gains and losses are not equal\textsuperscript{19}.

Figure 1 shows a case in which a combination of rapid evolution and unequal probabilities of gains and losses lead to dramatic errors in ancestral state reconstruction. In this experimental viral phylogeny, the true ancestors are known, allowing us to evaluate the accuracy of parsimony reconstructions. Each of the four bifurcating lineages experienced identical deletion events within the first 70 lytic cycles (the period between viral inoculation and the complete lysis of the host bacteria in liquid culture), whereas thousands of base pairs were eliminated from the genome. The nature of these deletions makes them irreversible, so that the probabilities of gains and losses are not equal. Although the character arose six times independently, parsimony incorrectly reconstructs the deletion in all ancestral nodes\textsuperscript{51}(Fig. 1). In this case, parsimony was unable to distinguish accurately between homology and convergence.

The problem of rapid evolution cannot be overcome when using parsimony, but the problem of unequal gain-loss probabilities can be investigated by applying a range of weights\textsuperscript{16,17}. Step matrices can be used to reconstruct ancestral states over a range of transformation probabilities, from matrices favoring the repeated loss of a character state to matrices favoring parallel gains. Such sensitivity analyses are important because, for most characters, we would rarely expect the rate of gains and losses to be exactly equal. Consider the evolution of plumage coloration in dabbling ducks (Fig. 2). It has long been thought that dichromatism – elaborately colored males and cryptic females – is the ancestral state for duck clades, with monochromatism having evolved independently many times.
When Omland applied parsimony with equal probability of gains and losses (equally weighted and unordered), he found that the inferred ancestral state for the group was monochromatism, implying that dichromatism had evolved convergently many times. In a sensitivity analysis, Omland found that only when losses were five times more likely than gains was the traditional hypothesis upheld that dichromatism is not a convergent character (Fig. 2). Omland went on to argue that a complex character, such as dichromatism, is more likely to be lost than to be gained. Furthermore, he noted several cases where monochromatic species were nested within paraphyletic dichromatic species, which was not accounted for by the species level phylogeny. For these and other reasons, Omland claimed that unequal weighting represents a more realistic model for the evolution of dichromatism. At the very least, this sensitivity analysis shows that the reconstruction of ancestral character states in dabbling ducks is not consistent across different weighting schemes.

Sensitivity analyses like this one raise a number of important questions. How can we interpret the observation that equally weighted parsimony reconstructs a monochromatic ancestor? Does this reconstruction falsify the hypothesis that the ancestor was dichromatic, or does the reconstruction say more about the assumptions of equally weighted parsimony than about the characteristics of the ancestor? Even if we accept the reconstruction favored by equally weighted parsimony, can we estimate the degree of confidence associated with this reconstruction? Although Maddison has developed a model for investigating the general accuracy of parsimony reconstructions, this model has not yet been applied to the problem of estimating the degree of confidence at individual nodes for specific character state distributions.

Estimating probabilities of maximum-likelihood reconstructions

Over the past few years, interest has grown in developing maximum likelihood (ML) approaches for reconstructing ancestral character states of discrete characters. These models include several important features:

1. The probability of change at a point in time along any branch of the phylogenetic tree depends only on the character state at the time, not on prior character states (a consequence of the Markovian process).

2. Transitions along each branch are independent of changes elsewhere on the tree (thereby allowing unparsimonious changes).

3. Rates of change are constant throughout time and along all branches. The rates of change between any two character states can be assumed to be equal (0-in1-1=0) or unequal (0-in1-1≠0), and
these rates are estimated by maximizing the likelihood with respect to the distribution of observed characters on the tree. When rates are unequal, the number of parameters to be estimated increases rapidly with number of character states. Because of the difficulty of accurately estimating multiple parameters, Schluter et al. recommend using the equal change model.

Schluter et al. compared the behavior of their ML models to equally weighted parsimony (Fig. 3). Using an allelic-based phylogeny of Darwin’s finches (Geospiza), parsimony and ML differ in their reconstruction of states at one of the deeper nodes (node Y, Fig. 3a). Although parsimony finds no support for granivory outside of the Geospiza clade (Fig. 3a), ML re-constructs the ancestor immediately below the Geospiza clade as granivorous, albeit with only 70% confidence (node Y in Fig. 3b, likelihood ratio of 2.3:1). As Schluter et al. point out, the ML reconstruction at node Y is a consequence of relative branch lengths and the assumption of a constant rate of evolution. Because the branch between node X and Y is relatively short, it is unlikely that a character evolving so slowly would have changed over the small time period it represents.

These results demonstrate that even at nodes with unambiguous parsimony reconstructions ML can prefer less parsimonious reconstructions (e.g. node Y, Fig. 3a,b). Even when parsimony and ML prefer the same reconstructions, the error associated with these reconstructions can be considerable (e.g. node Z, Fig. 3b). This error must be taken seriously because feeding mode in finches appears to evolve at a slow rate – the ideal circumstance for applying parsimony.

Whereas the limitations of the parsimony approach are well known, this ML model also has some obvious limitations. As Schluter et al. readily acknowledge, the most significant limitation is the assumption that every character evolves at a constant rate across the entire tree. As demonstrated here, this assumption of rate constancy is the reason why ML tends not to infer changes on relatively short branches (e.g. branch XY, Fig. 3b).

Because rate constancy has been shown to be a questionable assumption in phylogenetic reconstruction, even when averaged across many characters, its application to single-character reconstructions should be approached with caution. Although ML can theoretically incorporate variable rates, this can be problematic because more parameters would have to be estimated from a very small amount of data.

Another factor that can be considered either a strength or a weakness of Schluter et al.’s ML approach is its reliance on relative branch lengths to infer ancestral reconstructions. If branch lengths are accurate, they add an important temporal component lacking in discrete parsimony approaches. Unfortunately, estimates of branch lengths from any source of data are prone to error. Of special concern are branch lengths based on genes that have experienced substitutional saturation.

**Box 2. What is the role of ancestral states in correlative comparative methods?**

Correlative comparative methods test evolutionary hypotheses based on correlations between two character states found in the same organism, or between an organismal character and an environmental variable (reviewed in Ref. 27). For example, in mammals there is an imperfect correlation between living in the ocean and having fins. These methods depend on the assumption that each point in the correlation is independent. Because character correlations can also be created by shared phylogenetic history, a simple regression or contingency table will incorrectly treat all species as independent points.

In the example illustrated below, a simple contingency table reveals a significant correlation between states in the two characters. When ancestral states are reconstructed, however, there is only one observed transformation in each character. This means that the character states ‘a’ and ‘c’ appear together in many species simply because no transformation has occurred in either character. Ridley’s method of correcting for phylogenetic non-independence counts only the character associations in nodes that have experienced a transformation from the node below (described in Ref. 25). When this is done, the sample size is too low to detect any statistically significant association, even though both characters changed on the same branch.

In correlative comparative methods that use explicit reconstructions, ancestral states play a secondary role by either confirming or weakening the strength of the observed correlation. By contrast, in the homology-based hypothesis testing approaches that are the main focus of this review (e.g., Figs. 1–3), the ancestral reconstructions themselves play a primary role in hypothesis testing by distinguishing between homology and convergence.

---

**Figure adapted, with permission, from Ref. 1.**
Box 3. The generalized linear model approach to reconstructing discrete and continuous characters

Although much of this article focuses on discrete (saying characters), continuous characters are also important. Two parsimony methods exist for reconstructing continuous ancestral characters. Linear parsimony minimizes the absolute change on a tree, whereas squared change parsimony minimizes the sum of squared changes and is equivalent to assuming a Brownian motion model. Recently, a general method has been proposed that uses expected variances among taxa26. How these variances are determined depends on the evolutionary model and therefore allows flexibility in model choice (e.g. punctuational and selective). Other advantages are the ability to include intra-specific variation, the ability to calculate variances at ancestral nodes, and options for discrete or continuous character data. A simple continuous character example is illustrated.

1. Determine expected variances and covariances among terminal taxa, \( \text{var}(Y) \), and between ancestors and terminal taxa, \( \text{var}(A,Y) \), from a phylogeny and model of evolution.

\[
\text{var}(Y) = \begin{pmatrix}
5 & 2 & 6 & 8 & 4 \\
11 & 7 & 0 & 0 & 0 \\
4 & 10 & 0 & 0 & 0 \\
0 & 0 & 13 & 11 & 6 \\
0 & 11 & 12 & 12 & 12 \\
0 & 0 & 6 & 6 & 14
\end{pmatrix}
\]

2. Calculate the grand mean, an average of terminal taxa weighted by expected variances and covariances.

\[
A = \text{var}(A,Y) (\text{var}(Y)^{-1} (Y - M) + M)
\]

\( Y \) is a vector \( \{5,2,6,8,4\} \) of observed character data, \( A \) is a vector of ancestral states and \( M \) is the grand mean. In this example, \( A = \{3.54,7.11,5.46,8.61\} \), on the tree:

\[
\text{var}(A,Y) = \begin{pmatrix}
2 & 1 & 0 & 0 & 0 \\
1 & 2 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 2 & 2 \\
0 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 1
\end{pmatrix}
\]

These results are equivalent to weighted squared change parsimony (with a rooted tree). Assuming different models in (1) would yield different results.

With such genes, the lengths of internal branches may be systematically underestimated13. Because Schluter et al.'s ML method tends not to infer changes on small branches, errors in branch length estimation for saturated genes might introduce systematic biases towards inferring changes on terminal branches.

Schluter et al.'s method makes no assumptions about the magnitude of evolutionary rates. From a Bayesian perspective, this approach is equivalent to assuming a flat (equi-probable) prior-probability distribution21. This default approach might sometimes be necessary for those morphological and behavioral characters that have little basis for estimating rates. The main difficulty with this approach is that maximum-likelihood parameters are calculated from the small amount of data available for a single character22. The next version of Page's computer program, "Discrete"23, will provide likelihoods of ancestral states. (This program is available from the author by email at mark.page@oxford.ac.uk)

When inferring ancestral nucleotide and amino acid substitutions, Bayesian reconstruction methods can use models of molecular evolution to estimate the distribution of prior probabilities. These models can incorporate information about the sequence as a whole and about \( a \) priori information, such as empirically derived frequencies of transitions between amino acids and even protein structural information24. As with Schluter et al.'s method, these Bayesian approaches estimate the probabilities of ancestral reconstructions. Preliminary simulation studies suggest that, for molecular data, Bayesian methods generally outperform equally weighted parsimony22.

Testing evolutionary hypotheses with ancestral reconstructions

The recent work described here raises many concerns about the accuracy of ancestral reconstructions. In the past, any unambiguous parsimony reconstruction was generally considered sufficiently robust for cladistic hypothesis testing25. If the reconstructed state for the ancestral node of a group contradicted a long-held hypothesis, that hypothesis was considered falsified. We have seen that invoking unequal weighting (Fig. 2) or using ML to infer the ancestral state (Fig. 3) can change unambiguous equally weighted parsimony reconstructions. Furthermore, the results of the ML analyses show that there could be considerable uncertainty in ancestral reconstructions, especially for the deeper nodes in the tree26. These results suggest that caution should be used when testing evolutionary hypotheses using ancestral state reconstructions (but see discussion of correlative comparative methods in Box 2 and Ref. 27).

On the one hand, equally weighted parsimony might be appropriate when attempting to reject a null hypothesis of equal rates of gains and losses28-30. If a reconstruction based on equally weighted parsimony suggests that a character of interest evolved in parallel significantly more often than it was lost, this would falsify a null hypothesis of equal probability of gains and losses.

On the other hand, testing an evolutionary hypothesis with equally weighted parsimony is problematic when the null hypothesis is directional21,22. For example,
it is widely believed that complex characters, such as dichromatism, eyes, feeding larvae and wings, have a low probability of origin but can be lost or reduced by the action of a few genes\(^1\). Once lost, complex characters are often believed to be difficult to regain, especially after a long period of time\(^2\). As seen in the case of plumage, these characters can be lost through the degree of homoplasy necessary to support the original, directional hypotheses\(^3,4\). Finally, when testing evolutionary hypotheses we must consider the accuracy of the phylogenies themselves. Although this important issue is beyond the scope of the phylogenies themselves. Although this important issue is beyond the scope of this article, careful attention must be placed on the details of phylogenetic reconstruction. In particular, it is important to test explicitly whether trees supporting alternative hypotheses of character evolution represent a significantly worse fit to the data than minimum length trees\(^5\).

Developing better models for ancestral state reconstructions

In many cases, ancestral character state reconstructions based on simple models will not resolve issues such as the homology and directionality of evolutionary transformations. Although complex models are available for molecular evolution, it is not always obvious how to develop and apply better models for reconstructing the evolution of morphological and behavioral characters (but see Roh 30 and 32). From the standpoint of implementation, Martins and Hansen\(^5\) have made a significant theoretical advance with their general method for applying sophisticated evolutionary models to reconstructing ancestral character states for discrete and continuous characters (Box 3). Ultimately, however, determining the parameters for more realistic models of character evolution will have to depend on information from genetics, developmental biology, functional morphology and the fossil record. This information could also be useful for making more informed homology statements and for confirming the accuracy and plausibility of ancestral state reconstructions. Tree-based approaches to reconstructing ancestral character states form an important tool in generating and testing hypotheses of character evolution, but these methods should not be considered in isolation.

Acknowledgements

C.W.C. is supported by an NSF systematic grant DEB-9615461; K.E.O. is supported by an NSF postdoctoral fellowship; and T.H.O. is supported by a NASA GRIP fellowship. We thank the many individuals who commented on earlier drafts of this article, including Mitch Leslie, Emilia Martins, Paul Manos, Fred Nijhout, Steve Novacek, Rick Rice, Dolph Schluter, Steve Vogel, Will Wilson, Bob Zink, and two anonymous reviewers.

References

7. Brooks, D.R. and McLennan, D.A. (1991) Phylogenies and the fossil record. This information could also be useful for making more informed homology statements and for confirming the accuracy and plausibility of ancestral state reconstructions. Tree-based approaches to reconstructing ancestral character states form an important tool in generating and testing hypotheses of character evolution, but these methods should not be considered in isolation.

Acknowledgements

C.W.C. is supported by NSF systematic grant DEB-9615461; K.E.O. is supported by an NSF postdoctoral fellowship; and T.H.O. is supported by a NASA GRIP fellowship. We thank the many individuals who commented on earlier drafts of this article, including Mitch Leslie, Emilia Martins, Paul Manos, Fred Nijhout, Steve Novacek, Rick Rice, Dolph Schluter, Steve Vogel, Will Wilson, Bob Zink, and two anonymous reviewers.

References

7. Brooks, D.R. and McLennan, D.A. (1991) Phylogenies and the fossil record. This information could also be useful for making more informed homology statements and for confirming the accuracy and plausibility of ancestral state reconstructions. Tree-based approaches to reconstructing ancestral character states form an important tool in generating and testing hypotheses of character evolution, but these methods should not be considered in isolation.

Acknowledgements

C.W.C. is supported by an NSF systematic grant DEB-9615461; K.E.O. is supported by an NSF postdoctoral fellowship; and T.H.O. is supported by a NASA GRIP fellowship. We thank the many individuals who commented on earlier drafts of this article, including Mitch Leslie, Emilia Martins, Paul Manos, Fred Nijhout, Steve Novacek, Rick Rice, Dolph Schluter, Steve Vogel, Will Wilson, Bob Zink, and two anonymous reviewers.