Evolution, 54(2), 2000, pp. 397–405

INDEPENDENT CONTRASTS SUCCEED WHERE ANCESTOR RECONSTRUCTION FAILS IN A KNOWN BACTERIOPHAGE PHYLOGENY

TODD H. OAKLEY[1] AND CLIFFORD W. CUNNINGHAM[2]
Evolution, Ecology and Organismal Biology Group, Zoology Department, Duke University, Box 90325, Durham, North Carolina 27708-0325
[1]*E-mail: tho@duke.edu*
[2]*E-mail: cliff@duke.edu*

Abstract.—Methods of ancestor reconstruction are important tools for evolutionary inference that are difficult to test empirically because ancestral states are rarely known with certainty. We evaluated reconstruction methods for continuous phenotypic characters using taxa from an experimentally generated bacteriophage phylogeny. Except for one slowly evolving character, the estimated ancestral states of continuous phenotypic characters were highly inaccurate and biased, even when including a known ancestor at the root. This error was caused by a directional trend in character evolution and by rapid rates of character evolution. Computer simulations confirmed that such factors affect reconstruction of continuous characters in general. We also used phenotypic viral characters to evaluate two methods that attempt to estimate the correlation between characters during evolution. Whereas a nonphylogenetic regression was relatively inaccurate and biased, independent contrasts accurately estimated the correlation between characters with little bias.

Key words.—Ancestral reconstruction, comparative methods, computer simulation, independent contrasts, known phylogeny, squared-change parsimony.

Received January 20, 1999. Accepted September 21, 1999.

Methods of ancestor reconstruction play a central role in the study of character evolution. Ancestor reconstruction is not only necessary for homology-based, or character-mapping, studies (Coddington 1988; Donoghue 1989; Brooks and McLennan 1991), but also for some methods that estimate the correlation of characters during evolution (Huey and Bennett 1987; Harvey and Pagel 1991). Although reconstructing ancestors is undeniably important, it is difficult to compare methods empirically because actual ancestors must be known. One potential source of known ancestors is the fossil record. Although many ancestral fossils may be preserved (Foote 1996), demonstrating ancestor-descendant relationships is difficult (Nelson 1989; Smith 1994, pp. 125–128). Furthermore, many characters are not preserved in the fossil record. To compare methods of ancestral state reconstruction, we must rely on other systems.

Experimentally generated phylogenies, which have a degree of realism not possible in computer simulations, provide a rare opportunity where the actual ancestors are known (Hillis et al. 1992). Such a phylogeny has been created by evolving viruses in the laboratory and has already been used to examine reconstruction of discrete genetic characters (Hillis et al. 1992). In this study, we use the same phylogeny to examine evolution of multiple phenotypic characters with continuous variation. This allowed us to perform the first empirical test of methods that reconstruct continuous character evolution.

''Homology-based'' hypotheses require the investigator to distinguish between homology and convergence to make explicit estimates of character states at every node of a phylogeny (Greene 1986; Coddington 1988; Brooks and McLennan 1991; Frumhoff and Reeve 1994). The ancestral estimations themselves are then used to test the hypothesis at hand. Whereas the homology-based approach is commonly used in studies of discrete characters, continuous characters—the focus of this study—can be used in a similar fashion. For example, knowledge of ancestral character states was required to test whether body size changed multiple times in *Anolis* lizards that inhabit islands in sympatry with related species (Losos 1990). The lack of convergence in size in these lizards was taken as evidence for nonrandom colonization of islands (Losos 1990).

Although homology-based methods rely on the accuracy of ancestor reconstruction, this is not necessarily true for methods that seek to quantify the correlation between characters during evolution, which are referred to here as ''correlative methods.'' Some such methods do not even reconstruct ancestors per se (Felsenstein 1985; Grafen 1989; Pagel 1998). Other correlational methods do estimate ancestors as a preliminary step in analyses designed to determine the manner or extent that multiple characters (or a character and environmental variable) have covaried over the course of evolution (e.g., Huey and Bennett 1987). In all these correlative methods, phylogeny is used to correct for statistical nonindependence caused by shared history (Felsenstein 1985). For example, we might use correlational analyses to test whether brain size and body size is correlated (Pagel and Harvey 1988, 1989).

We began this study by measuring phenotypic characters—namely, different measures of growth rate—of ancestral and terminal taxa from a known virus phylogeny. We then used character values of terminal taxa to reconstruct ancestral character values. Next, we compared these reconstructed values to the actual ancestral values, which revealed that explicit reconstruction was highly inaccurate, regardless of the specific method employed. We then used computer simulation of continuous characters to show that error increases with faster character evolution and increased directional bias in character evolution. Finally, we used the same bacteriophage data to examine the accuracy of two correlative comparative methods, a nonphylogenetic approach and independent contrasts (Felsenstein 1985). Whereas the nonphylogenetic ap-

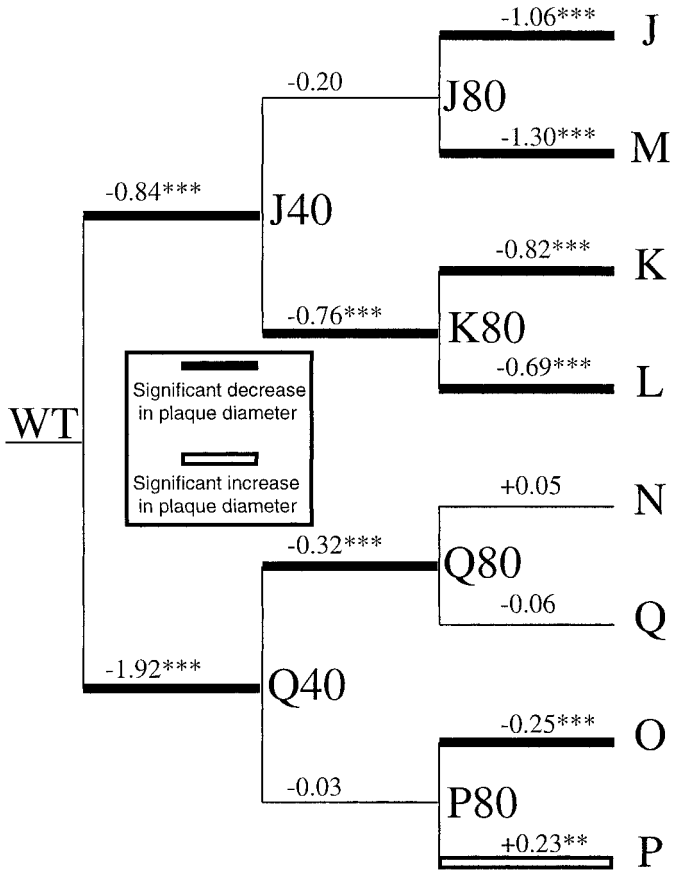

FIG. 1. Change in plaque diameter along experimental virus phylogeny. Numbers represent change along each branch. Significance of this value was calculated by comparing the mean value of the descendant to the mean value of ancestor plaques with a *t*-test. Fourteen of 16 branches underwent a decrease in plaque diameter, which is significant based on a binomial test ($P = 0.002$).

proach was prone to error, independent contrasts (Felsenstein 1985) was able to accurately estimate the correlation between characters.

MATERIALS AND METHODS

Generating the Virus Phylogeny

Bacteriophage T7, a virus that infects *Escherichia coli,* was used to create the known phylogeny used in the current study (White et al. 1991; Hillis et al. 1992). Details have been described elsewhere. Briefly, T7 was grown in liquid culture and exposed to a mutagen to increase the rate of evolution. At prescribed intervals, phage cultures were reduced to a single virus by standard plating techniques, thereby imposing a stringent bottlenecking regime resulting in large amounts of genetic drift (Bull et al. 1993). Evolving phage lineages were split in a bifurcating manner to create a known, symmetric phylogeny with eight terminal taxa (see Fig. 1 for topology). All branch lengths were equal in terms of number of generations. Viable viruses sampled at each bifurcation were stored, allowing for measurement of characteristics of both terminal taxa and actual ancestors. Viruses were dormant during storage, so no additional evolution occurred.

Four Measures of Viral Phenotype

Plaque diameter and shape were measured after growing plaques under controlled conditions. Plaques are circular areas of viral growth formed when bacteriophage are grown on a petri plate covered with a ''lawn'' of host bacteria. We grew plaques on single-layer plates to minimize environmental variability caused by heterogeneity in nutrient agar (Yin 1991). We measured plaque diameter and shape for all taxa simultaneously, using aliquots of the same host culture for all petri plates because host quantity and quality both affect viral growth (Wang et al. 1996). Host *E. coli* strain W3110 was grown overnight at 37°C. We diluted virus cultures so plaques would grow uncrowded on plates and mixed these cultures with 0.6 ml of overnight host culture and 6 ml of molten LB top agar (Sambrook et al. 1989) at 50°C. We then poured mixtures into one or two petri plates per taxon (plaques of the same taxon grown on multiple plates did not show significant variation between plates; results not shown). Once solid, plates were incubated at 37°C for 12 h.

To measure plaques, images of each plate were digitally captured with a charged-coupled device camera. After digitally imaged plaques were identified manually, the computer measured diameter and shape automatically using the computer program Image-1 (Universal Imaging, Brandywine, PA). Shape varies between zero and one and measures the circularity of the plaque; a value of one is a perfect circle (Image-1 manual). Between 58 and 101 plaques were measured for each taxon. The mean values of diameter and shape were used in subsequent analyses. To test for a significant change in diameter from ancestors to descendants, we performed t-tests by comparing the mean of each ancestor to the mean of each of its descendants.

Another phenotype measured was lysis efficiency. A culture of host bacteria, *E. coli* strain W3110, was grown overnight and diluted to an optical density (OD) of 0.8. Next, 3 ml of the host culture were added to a dilution of each virus culture corresponding to 50 plaque forming units (pfu). After 24 h at 37°C, the OD of each virus/host mixture was measured. Bacterial lysis resulted in lowered OD values. Some viral taxa did not lyse bacteria in liquid culture, in which case OD increased above 0.8 after 24 h.

Finally, growth rate in liquid culture was used as a phenotype (Bull et al. 1997). Again, host W3110 was grown overnight and diluted to OD = 0.8. Between 10 and 600 plaque forming units (pfu) of each virus were added to 1 ml aliquots of the bacterial culture. This mixture was incubated for 1.5 h and virus concentration counted by standard plating before and after incubation. The number of times virus concentration doubled during the growth period was used as a measure of growth rate in liquid culture (Bull et al. 1997).

For the remainder of this paper, we refer to the plaque diameter phenotype as ''diameter,'' plaque shape as ''shape,'' lysis efficiency as ''lysis,'' and growth rate in liquid culture as ''liquid fitness.''

Contamination Check

Contamination could cause mixing of taxa and was therefore a concern. Great care was taken not to contaminate viral cultures and nucleotide sequencing was also used as a means

of identifying viruses used in each test. The nucleotide sequence corresponding to bases 34624 to 35070 of the published T7 genome (Dunn and Studier 1983) was determined from a plaque drawn at random from each plate of the plaque diameter/shape test and from each taxon used in the liquid fitness test. The sequenced region has unique mutations for each viral strain, including all ancestors (C. W. Cunningham, unpubl. data). No contamination was detected.

Explicit Ancestor Reconstruction

We used several different methods to reconstruct ancestral states. For all methods, the known tree topology was input as a rooted tree with equal branch lengths because an equal number of generations separated all ancestors and immediate descendants (White et al. 1991; Hillis et al. 1992). The percent error of the estimates for each node was calculated by subtracting the actual ancestral value from the estimated value and dividing by the actual value. Mean percent errors across all nodes were calculated for each method to gauge the bias of the estimates. The mean absolute value of all nodal estimates computed by each method was also calculated to gauge overall accuracy.

We first employed the rooted squared-change parsimony algorithm (Huey and Bennett 1987; Maddison 1991) of MacClade 3.06 (Maddison and Maddison 1992) to reconstruct ancestral states of all four characters. We then used two other methods to reconstruct ancestors. First, we used maximum-likelihood implemented with a Brownian motion model in ancml (Schluter et al. 1997). This method gives results identical to squared-change parsimony (if branch lengths are incorporated), but also allows standard errors of reconstructions to be estimated. Second, we used linear parsimony, which assigns ancestral states that minimize the absolute character change on a phylogeny. This method often results in multiple, equally parsimonious reconstructions represented by a range of character values at a node. Other methods that assume Browninan motion (e.g., Garland et al. 1999) were not used because results are identical to the maximum-likelihood method.

In addition to the above analyses, we investigated the effects of adding an outgroup or a fossil. The wild-type ancestor was added on a zero-length branch attached to the root, to simulate knowledge of an ancestral fossil. Also, we investigated the effect of adding a hypothetical outgroup. To do so, we erected a hypothetical branch that was the same length as the length between the root and the terminal taxa. To this branch, we added wild-type virus as an outgroup taxon (wild type was not remeasured when included as an outgroup, so character values were identical to the root node characters).

Computer Simulation

Computer simulation was used to test the hypothesis that a directional bias and an increased rate of change in character evolution adversely affect the accuracy of reconstructed ancestral characters. Variations in the simulation, such as including an outgroup or ancestral fossil or reconstructing ancestors with multiple methods was considered beyond the scope of the current paper.

The simulation used the same tree topology as the virus phylogeny with all branches of equal length. Continuously varying characters evolved along the tree with three models of evolution. First, we implemented a Brownian motion model by adding a random number for each branch to an arbitrary initial character value. The computer program drew the random numbers from a normal distribution with mean equal to zero and a variance that was systematically changed between simulations. Lower variance in this context simulates smaller amounts of character change (Diaz-Uriarte and Garland 1996). The other two models incorporated the random variation of Brownian motion as well as a directional trend in character evolution. For the two models with a trend, the computer program drew random numbers from distributions with a different mean for each model—both greater than zero—thus imposing the directional trends (Garland et al. 1993). Again variances were systematically changed between simulations to imitate different rates of character change.

After evolving characters along the tree 1000 times for each parameter value, the algorithm for squared-change parsimony described by Maddison (1991) was used to reconstruct ancestral states for each replicate using the known tree and terminal character states. Maddison's algorithm gives identical ancestral values as ''ancml'' in cases (such as this virus phylogeny) when all branches are of equal length (Schluter et al. 1997). Actual ancestral values from the simulation were then compared to the parsimony reconstruction and differences expressed as percent errors. The average percent error at each node was calculated by subtracting the actual ancestral value from the reconstructed ancestral value and dividing by the actual value.

Correlational Analyses

We compared actual and reconstructed correlations for all six possible pairwise combinations of four phenotypic characters. The actual correlation between characters was determined by first calculating change along each branch for each character. These values were in turn used to determine the correlation between the characters. This measure is an appropriate standard to gauge correlative comparative methods because comparisons are statistically independent; there is no correlation introduced by phylogenetic history. Martins and Garland (1991) referred to this measure as the ''realized evolutionary correlation.'' Percent errors were calculated by comparing the realized evolutionary correlation to estimates of correlation from two different methods.

Independent contrasts was one correlational method we evaluated (Felsenstein 1985). We first checked whether branch lengths were appropriately standardized (Garland et al. 1992). The highest correlation between branch length standard deviations and absolute contrast value for any character (liquid fitness; $r = 0.612$, $P = 0.144$) was not significant. It was therefore assumed that branch lengths were adequately standardized, and raw data were not transformed in any way. We then calculated contrast values using CONTRAST of the PHYLIP package (Felsenstein 1995). We do not present results of the directional method of Huey and Bennett (1987) because it arrives at correlation values identical to independent contrasts when branch lengths are equal (see also Pagel 1993). In addition to independent contrasts, we performed a

TABLE 1. Virus phenotypes. Note that an increase in lysis efficiency reflects a decrease in fitness.

Taxon	Plaque diameter[1]		Plaque shape[2]		Lysis efficiency[3]		Liquid fitness[4]
	Mean	SD	Mean	SD	Mean	SD	
WT	3.02	0.94	0.83	0.07	0.22	0.04	8.79
J40	2.18	0.89	0.76	0.09	0.26	0.05	7.26
Q40	1.10	0.45	0.66	0.08	0.30	0.07	4.01
J80	1.98	0.73	0.75	0.08	0.38	0.12	6.88
K80	1.41	0.52	0.69	0.07	0.27	0.04	4.62
P80	1.07	0.32	0.66	0.06	0.44	0.09	3.48
Q80	0.78	0.22	0.61	0.06	0.39	0.02	2.43
J	0.92	0.28	0.63	0.07	0.78	0.34	3.93
K	0.59	0.16	0.54	0.07	0.45	0.06	0.58
L	0.72	0.14	0.58	0.06	0.52	0.04	0.37
M	0.68	0.22	0.57	0.08	1.29	0.26	1.98
N	0.83	0.30	0.60	0.09	0.58	0.12	2.14
O	0.82	0.23	0.61	0.07	1.24	0.22	1.47
P	1.30	0.54	0.68	0.08	0.36	0.04	3.52
Q	0.72	0.17	0.59	0.05	1.16	0.39	1.06

[1] $N = 58–101$ plaques on one or two plates.
[2] $N = 58–101$ plaques on one or two plates.
[3] $N = 4$.
[4] $N = 1–2$.

nonphylogenetic correlation. This was simply a conventional Pearson correlation analysis that directly compared two character values of terminal taxa. Correlation (r) and corresponding P-values were calculated for the correlational analyses using the computer program StatView 4.1 (Abacus Concepts, Berkeley, CA).

RESULTS

Measures of Viral Phenotype

All actual measures of viral phenotype showed a tendency toward decreased fitness or shape value (Table 1). Because all characters were similar in this respect, only one—diameter—is presented in detail. Diameter decreased along a significant majority of branches (14 of 16; Fig. 1; binomial $P = 0.002$). In addition, nine of these branches showed a significant decrease in plaque diameter based on t-tests that compared mean diameter of each taxon to its immediate ancestor (Fig. 1). The branch leading to taxon P showed the only significant increase in diameter based on t-tests.

Explicit Ancestor Reconstruction

When using squared-change parsimony, reconstructed ancestors were highly inaccurate for all characters except shape (Table 2). Plaque shape, the least variable character, showed the lowest error at 14.4% (Table 2). Estimates of most ancestral character values for shape, diameter, and liquid fitness were underestimated, which was a result of the decreasing trends in character values. All ancestral character values were overestimated for lysis, a character that showed an increasing trend during evolution. The mean absolute error for all four characters and all nodes was 67.7%. The root node showed the highest error (absolute mean = 110.8%) compared to other nodes when averaging all four characters.

For linear parsimony analysis and analyses including an outgroup or fossil, we show results only for plaque diameter (Table 3), because other characters yielded very similar results. All methods resulted in poor estimates of ancestral diameter, even when including an outgroup or fossil (Fig. 2; Table 3). Although including a fossil or outgroup did little to change absolute reconstruction accuracy, there were three other effects. First, including a fossil largely eliminated the bias toward underestimating ancestors (mean percent error = 6%). Second, standard errors of nodal estimates increased in both fossil and outgroup analyses, a result of increased variation in data values. This caused all nodes in the fossil analysis to fall within 95% confidence intervals. Third, including a fossil or wild-type outgroup allowed inference of

TABLE 2. Percent error of ancestral character estimates for four characters at seven ancestral nodes. Note that a positive percent error for liquid fitness represents an underestimation of fitness because higher character values represent lower fitness.

Node	Plaque diameter	Plaque shape	Lysis efficiency	Liquid fitness	Absolute mean (nodal)
WT	−72.8%	−27.4%	264.3%	−78.6%	110.8%
J40	−64.7%	−22.4%	202.0%	−75.4%	91.1%
Q40	−20.2%	−7.1%	174.1%	−50.7%	63.0%
J80	−60.2%	−21.1%	151.5%	−62.7%	73.9%
K80	−50.9%	−17.8%	114.2%	−80.2%	65.8%
P80	−6.1%	−4.2%	82.0%	−33.2%	31.4%
Q80	3.3%	−1.0%	117.7%	−28.8%	37.7%
Absolute mean	39.7%	14.4%	158.0%	58.5%	67.7%

TABLE 3. Ancestral reconstruction of plaque diameter using various methods. Standard errors of maximum-likelihood (ML) estimates are in parentheses. Ranges for linear parsimony represent the range of equally most parsimonious results.

Node	Actual value	Squared-change parsimony/ML			Linear parsimony	
		Ingroup only	With "fossil"	With outgroup	Ingroup only	With outgroup
WT	3.02	0.822 (0.160) −72.8%	fixed	1.319 (0.350) −56.3%	0.718–0.819 −76% to −73%	0.819–3.02 −73% to 0%
J40	2.18	0.768 (0.131) −64.7%	1.710 (0.554) −21.4%	0.981 (0.317) −54.9%	0.718–0.819 −67% to −62%	0.718–0.825 −67% to −62%
Q40	1.10	0.877 (0.131) −20.2%	1.818 (0.554) 65.5%	1.089 (0.317) −0.8%	0.720–0.819 −35% to −26%	0.819–0.825 −25%
J80	1.98	0.789 (0.108) −60.2%	1.103 (0.522) −44.4%	0.860 (0.268) −56.6%	0.718–0.819 −64% to −59%	0.718–0.825 −64% to −58%
K80	1.41	0.693 (0.108) −50.9%	1.007 (0.522) −28.6%	0.764 (0.268) −45.8%	0.718 −49%	0.718 −49%
P80	1.07	1.000 (0.108) −6.1%	1.314 (0.522) 23.4%	1.071 (0.268) 0.5%	0.819 −23%	0.819–0.825 −23%
Q80	0.78	0.807 (0.108) 3.3%	1.121 (0.522) 43.4%	0.878 (0.268) 12.3%	0.720–0.819 −8% to +5%	0.819–0.825 +5% to +6%
Mean % error		−38.8%	+6.3%	−28.8%	−43.4%	−35.5%
Mean absolute % error		39.7%	32.4%	32.5%	45.5%	36.9%

the actual trend toward smaller diameter in all cases, except when using linear parsimony. The analysis using linear parsimony and an outgroup resulted in two possible values for the root node, one of which would not allow inference of the actual trend toward smaller diameter.

Computer Simulation

Figure 3 shows changes in absolute error of reconstruction at particular nodes. An increase in absolute error was often associated with increased directional bias and with increased rates of character change. These results were expected based on earlier studies of discrete characters (Holmquist 1979; Saitou 1989; Tateno 1990; Frumhoff and Reeve 1994; Maddison 1995; Schultz et al. 1996), but have not been demonstrated for continuous characters. In addition, the root node was the most sensitive to changes in character rate/bias and the nodes closest to the terminal taxa were less sensitive. This has also been shown for discrete characters (Maddison 1995).

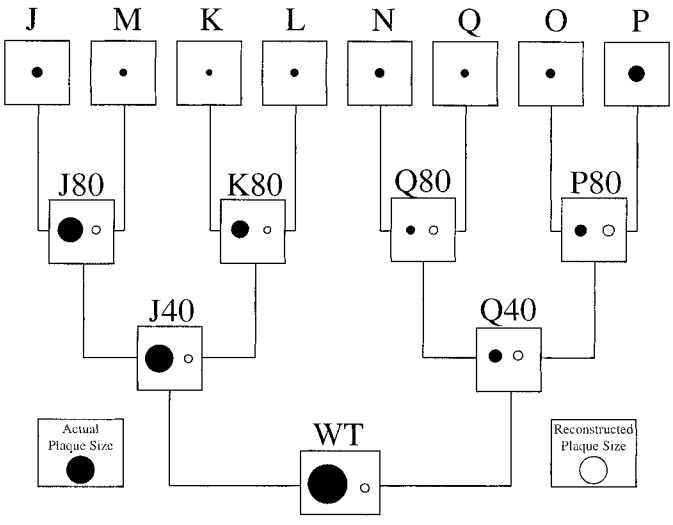

FIG. 2. Comparison of actual (mean) and reconstructed plaque diameters without an outgroup or fossil. Viral plaques from an experimentally generated phylogeny were measured and data for the terminal taxa were used to reconstruct ancestral states using squared-change parsimony (Maddison 1991). Actual values are proportional to black circles and reconstructed values are proportional to white circles.

Two apparent peculiarities are present in Figure 3. First, when there was a large bias in character evolution, increased rates of evolution did not drastically affect reconstruction. This is because changes in character value caused by the trend are large compared to changes caused by increasing rates of character evolution: Therefore, character change is mainly influenced by the trends and not by the random processes introduced by increased variance. This is supported by the fact that at extremely high rates of evolution—at parameter values not shown on the graphs in Figure 3—percent error increased.

The second apparent anomaly was that at high rates of character evolution, the percent error of reconstruction became erratic. This was probably caused by stochasticity: At high rates of evolution, the terminal taxa are more variable, resulting in an increased variance of reconstruction accuracy. Therefore, much of the change in reconstruction accuracy from one parameter value to the next may be the result of the overall variability in accuracy. This is supported by the fact that root node reconstruction (Fig. 3A), which depends equally on eight terminal taxa, is less erratic than subterminal node reconstruction (Fig. 3B), which mostly depends on only two terminal taxa.

Correlational Analyses

"Cross-sectional" correlative methods compare character states of terminal taxa and do not necessarily reconstruct ancestors (Pagel 1998). The independent contrasts method is

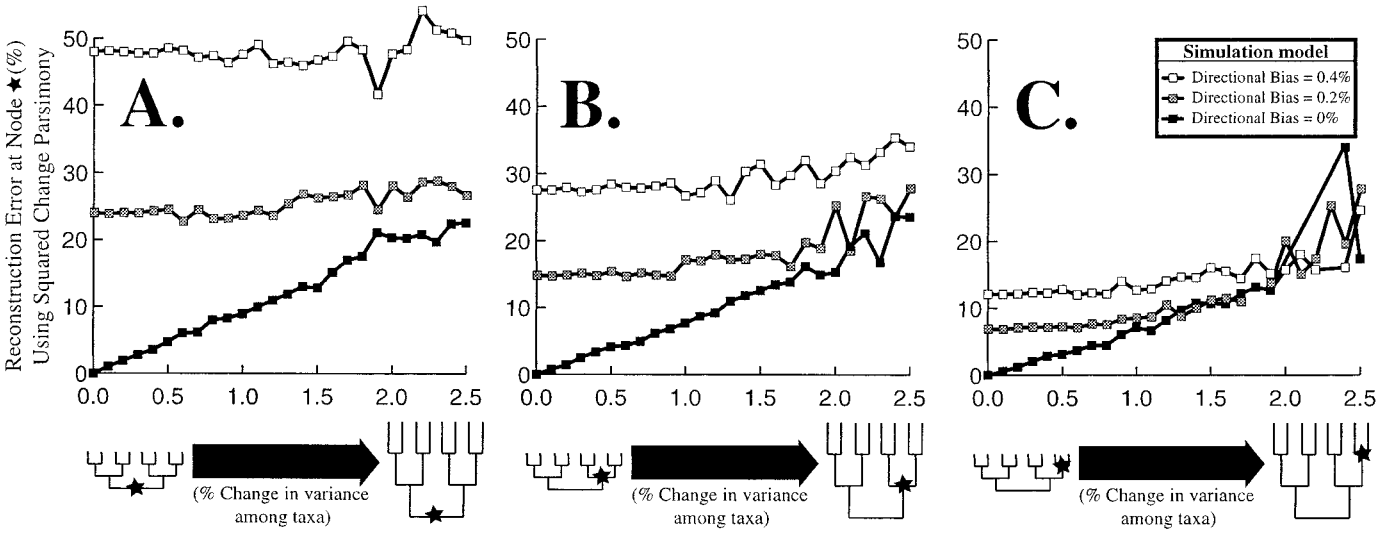

FIG. 3. Results of a computer simulation in which characters were simulated along a symmetric phylogeny with eight terminal taxa. Three models of character evolution were used, one with no directional bias in character evolution, the others with different amounts of bias. The x-axis represents increasing amounts of character change along the phylogeny and the y-axis represents mean absolute value of the percent error of reconstruction of 1000 replications using squared-change parsimony. Three different nodes of the phylogeny are illustrated in A, B, and C.

a well-known cross-sectional method in which statistically independent comparisons are made between taxa to estimate the correlation between characters (Felsenstein 1985). This method fared the best of any examined here. The average absolute error of correlation estimates was 15.0% (Table 4), which is far lower than the error of explicit ancestor reconstruction with the same characters (67.7%). However, *P*-values estimated by independent contrasts were usually higher than those estimated by calculating the actual change along branches (Table 4). This may be attributed in part to a difference in statistical power between change along branches and independent contrasts (calculating change along branches includes actual ancestors and therefore has more datapoints). This can be confirmed by comparing *P*-values estimated by independent contrasts with *P*-values estimated by contrasts using actual ancestors. These two analyses use similar numbers of datapoints and indeed arrive at more similar *P*-values (Table 4).

Nonphylogenetic regression in general produced low correlation values compared to other measures, with an average absolute error of 36.7% (Table 4). One comparison, liquid versus lysis, was highly nonsignificant ($P = 0.7990$). This may seem counterintuitive because, in computer simulation, nonphylogenetic methods showed high amounts of Type I error, the incorrect rejection of null hypotheses (Felsenstein 1985; Martins and Garland 1991; Diaz-Uriarte and Garland 1996). However, inflated Type I error in nonphylogenetic comparisons is not caused by a directional bias toward lower

TABLE 4. Comparison of correlational methods. Values are absolute value of ''R'' and associated *P*-values are in parentheses. Percentages are percent error of R assuming change along branches (using actual ancestors; CAB) is the actual correlation. IC-AA, independent contrasts calculated with actual ancestors; IC, independent contrasts; NP, nonphylogenetic regression.

	Actual correlation		Correlative method	
Characters compared	CAB	IC-AA	IC	NP
Diameter vs. liquid	0.931 (< 0.0001)	0.866 (0.0055)	0.834 (0.0100) −10.4%	0.762 (0.0280) −18.2%
Diameter vs. shape	0.954 (< 0.0001)	0.956 (0.0002)	0.959 (0.0002) 0.5%	0.971 (< 0.0001) 1.8%
Liquid vs. lysis	0.673 (0.0059)	0.749 (0.0324)	0.546 (0.1618) −18.9%	0.108 (0.7990) −84.0%
Diameter vs. lysis	0.517 (0.0485)	0.801 (0.0169)	0.685 (0.0607) 32.5%	0.363 (0.3772) −29.8%
Shape vs. lysis	0.633 (0.0112)	0.681 (0.0630)	0.552 (0.1557) −12.8%	0.214 (0.6113) −66.2%
Shape vs. liquid	0.980 (< 0.0001)	0.849 (0.0076)	0.834 (0.0100) −14.9%	0.783 (0.0216) −20.1%
Mean error			−4.0%	−36.1%
Mean absolute error			15.0%	−36.7%

P-values, but rather by an increased variance of the estimated *P*-value (Martins and Garland 1991; Ackerly 1998; D. D. Ackerly, unpubl.). Indeed, our data are consistent with this idea, because the *P*-values estimated by nonphylogenetic methods are more variable than *P*-values calculated by other methods for the same characters (Table 4).

Discussion

Examination of viral phenotypes from a known phylogeny revealed a dynamically evolving biological system. Although a decrease in diameter during evolution was the norm, one terminal branch showed a significant increase, thus illustrating a potential for the viruses to recover from lowered fitness. In addition, two clades showed very different character trajectories. One clade (the left half of the phylogeny in Fig. 2) showed a gradual decrease in diameter over time. The other half of the phylogeny showed a rapid initial change followed by relative stasis. Such a dynamic system was appropriate for analyzing reconstruction of ancestral phenotypes.

A Failure to Reconstruct Known Ancestors

All reconstruction methods inaccurately estimated ancestral viral phenotypes despite several factors in this experimental system favoring accuracy. First, unlike most empirical studies, we knew the correct tree topology and branch lengths. We also used an ancestral fossil in some analyses, which was known to be the true ancestor at the base of the tree. Finally, we used an outgroup in some analyses that was known not to have experienced the directional trend present in the ingroup taxa. If we had chosen to use the outgroup taxon ''R'' (Hillis et al. 1992), which exhibited the directional trend (data not shown), our results would have been even worse. Despite all these factors favoring accuracy, we were not able to estimate continuous ancestral characters accurately. Estimated ancestors were in some cases wrong by more than an order of magnitude.

The inability to reconstruct continuous ancestral characters accurately seems to be in direct conflict with a previous study of the viral system. Hillis et al. (1992) accurately reconstructed discrete genetic characters using the same experimental phylogeny described here. The conflict may be explained by the difference in mode of evolution between character types. Although nucleotide and restriction site changes did occur, there was far less homoplasy in the genetic characters. This is in contrast to extreme homoplasy in the phenotypes we measured. These results suggest the possibility that homoplastic phenotypes have evolved by nonhomoplastic genetic changes, not an unlikely scenario because many different genetic mutations can affect viral fitness (e.g., Escarmis et al. 1999). This hypothesis could be tested in an experimentally tractable system like T7.

Homoplasy and the resulting errors in reconstruction of phenotypic characters were caused by two factors. The first factor was rapid character evolution. The slowest evolving character, shape, had the lowest percent error in estimation. In addition, the computer simulation of continuous characters generally revealed increased error with faster rates of evolution (Fig. 3), a result that has been previously demonstrated theoretically for discrete characters (Frumhoff and Reeve 1994; Schultz et al. 1996). This result is intuitive: If no change has happened in a character, we can estimate ancestors with absolute certainty. More change in a continuous character means a greater variance in character state and therefore more uncertainty.

The second factor causing errors in ancestral reconstruction was a strong directional bias during character evolution. Ancestral estimates always fall within the range of terminal taxa; therefore, large-scale trends in character evolution may be impossible to recover without a fossil or an outgroup that did not undergo the trend (Martins and Hansen 1997). In the virus phylogeny, a bias was caused by a Muller's ratchet-like phenomenon: a progressive decrease in fitness caused by an accumulation of deleterious mutations that is especially enhanced by small effective population sizes and high mutation rate (e.g. Charlesworth et al. 1993; Duarte et al. 1993). Both of these factors that accelerate Muller's ratchet were present in the viral system (Bull et al. 1993), thus causing a trend toward lowered fitness as deleterious mutations accumulated.

Also because of the directional trend, standard errors were positively misleading. Standard errors are important for measuring reliability of ancestral estimates. If accurate, large errors could indicate fast character evolution and designate cases in which characters are unsuitable for reconstruction (Schluter et al. 1997). Standard errors were quite small in the virus phylogeny because little variation was present among terminal taxa, even though much change happened during evolution. Much of the actual character change was masked because terminal taxa evolved in parallel toward similar character values. Directional biases in evolving characters cause the insidious combination of increased certainty and decreased accuracy when ancestral values of such characters are estimated. It is not difficult to imagine factors in nature, such as directional selection, that could produce similar results.

Can We Detect Evolutionary Trends?

If we could detect directional biases or evolutionary trends in characters, we might anticipate their confounding effects on ancestor reconstruction. Including an outgroup or fossil did allow detection of a decreasing trend in diameter—but unfortunately only in special circumstances. A fossil that is a direct ancestor can allow unambiguous interpretation of a trend. For example, when an ancestor is large and multiple descendants are small, one does not need to reconstruct any ancestors to infer a trend. However, this situation does require showing an ancestor-descendant relationship, which may not be possible because doing so relies on negative evidence—a lack of autapomorphies in the fossil taxon (Nelson 1989; Smith 1994, pp. 125–128).

Including an outgroup may also indicate a trend, but again only in certain situations. To infer a trend, the outgroup must not have experienced the trend. Even when using outgroups unaffected by a directional trend, inferring a trend in the ingroup depends on the assumptions we make about the mode of character evolution. For example, using the wild-type phage as an outgroup did suggest a trend, but only when using methods that assumed Brownian motion. With linear

parsimony, multiple equally parsimonious reconstructions made inferring the trend impossible. Correctly inferring the trend thus depended on the model employed, but selecting appropriate models for character evolution remains a challenge (Omland 1997; Cunningham et al. 1998). Evolutionary trends and directional biases are therefore quite problematic for ancestral state reconstruction. They not only decrease the accuracy of reconstructions with misleading certainty, but they also may often be undetectable (Felsenstein 1988; but see Hansen 1997; Garland et al. 1999).

Can We Improve the Accuracy of Ancestral Estimates?

An appropriate, but highly specific model could easily be employed with a generalized linear models approach (Martins and Hansen 1997) that would allow for accurate reconstruction. Such a model might have a rapid decrease in diameter followed by stasis on one part of the tree and gradually decreasing diameter on another part of the tree. However, it seems doubtful that such detailed models could often be correctly designed a priori without knowledge of the very ancestors being estimated.

The Success of Independent Contrasts

We used two methods to quantify the inferred correlation between characters during evolution. The nonphylogenetic approach was highly inaccurate. This is not surprising because the unfavorable consequences of ignoring phylogeny have been well illustrated elsewhere (e.g., Felsenstein 1985; Harvey and Pagel 1991; Martins and Garland 1991; Diaz-Uriarte and Garland 1996). Yet, despite inaccuracies with nonphylogenetic methods and explicit ancestor reconstruction, independent contrasts fared well with the lowest mean absolute error of any method tested, even though using the same data.

Although independent contrasts performed well in this study, the method is not perfect. For example, in three of six cases independent contrasts failed to detect a significant correlation between characters (assuming $\alpha = 0.05$), but this conservatism may be related to the small number of terminal taxa. Furthermore, computer simulations have illustrated conditions in which the method performs poorly; namely, when assumed models are severely violated (Martins and Garland 1991; Diaz-Uriarte and Garland 1996). The virus data presented a somewhat limited test of correlative methods because the different phenotypes, mainly measures of growth rate, all evolved in a similar fashion. Unfortunately, we were not able to measure any phenotypic characters not correlated with growth rate. Nevertheless, even if these results do not generalize to all character types and evolutionary scenarios, they do call attention to important differences in how ancestral states are used in different methods.

Conclusions

Although based largely on a single data set, this paper suggests that the independent contrasts method succeeds under conditions where explicit ancestor reconstruction fails. If this hypothesis is true in general, researchers should consider rethinking hypotheses of character evolution to allow the use of independent contrasts in addition to or instead of the more error-prone ancestral estimation procedures. Our results concur with theoretical assertions that independent contrasts do not rely on the accuracy of reconstructed ancestors (e.g., Martins and Hansen 1996). Therefore, correlative methods that compare only terminal taxa to eliminate a perceived reliance on ancestral state reconstruction (e.g., Møller and Birkhead 1992; Mitani et al. 1996) may be unnecessary and reduce the power of comparative analyses.

Nonetheless, some hypotheses absolutely require the explicit estimation of ancestral states (e.g., Losos 1990, 1992; Ryan and Rand 1995, 1999; Schluter et al. 1997; Parker et al. 1998). In such studies, phylogenetic evidence alone may not be sufficient to unambiguously interpret character evolution (Strathmann and Eernisse 1994; Cunningham et al. 1998). Researchers must consider additional evidence—besides phylogenetic evidence—to test the hypothesis at hand. Detailed morphological, genetic, paleontological or biogeographical studies should be used to test alternatives in character reconstruction. From this perspective, use of phylogeny for understanding character evolution becomes a way of generating rather than testing hypotheses (see also Losos 1996). This seems a productive, yet conservative, use of phylogeny in the study of character evolution.

Acknowledgments

We thank M. Badgett for sending virus samples and M. White, M. Badgett, D. Hillis, J. Bull, and I. Molineaux for generating the known phylogeny used in this study. Thanks are due to A. Beard for assisting with collection of virus phenotype data while supported by a Howard Hughes undergraduate summer fellowship and to Y. Huang who collected sequence data. D. Ackerly, T. Garland, T. Hansen, J. Losos, C. Nunn, K. Omland, and an anonymous reviewer provided helpful comments and discussion. THO was supported by a GSRP fellowship from the National Aeronautics and Space Administration. This work was also supported by National Science Foundation systematics grant DEB-9615461 to CWC.

Literature Cited

Ackerly, D. D. 1998. Leaf size, sapling allometry, and Corner's rules: a phylogenetic study of correlated evolution in maples (*Acer*). Am. Nat. 152:767–791.

Brooks, D. R., and D. A. McLennan. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. Univ. of Chicago Press, Chicago.

Bull, J. J., C. W. Cunningham, I. J. Molineux, M. R. Badgett, and D. M. Hillis. 1993. Experimental molecular evolution of bacteriophage T7. Syst. Biol. 47:993–1007.

Bull, J. J., M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux. 1997. Exceptional convergent evolution in a virus. Genetics 147:1497–1507.

Charlesworth, D., M. T. Morgan, and B. Charlesworth. 1993. Mutation accumulation in finite outbreeding and inbreeding populations. Genet. Res. 61:39–56.

Coddington, J. A. 1988. Cladistic tests of adaptational hypotheses. Cladistics 4:3–22.

Cunningham, C. W., K. O. Omland, and T. H. Oakley. 1998. Reconstructing ancestral character states: a critical reappraisal. Trends Ecol. Evol. 13:361–366.

Diaz-Uriarte, R., and T. Garland. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts:

sensitivity to deviations from Brownian motion. Syst. Biol. 45: 27–47.
Donoghue, M. J. 1989. Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. Evolution 43: 1137–1156.
Duarte, E. A., D. K. Clarke, A. Moya, S. F. Elena, E. Domingo, and J. Holland. 1993. Many-trillion-fold amplification of single RNA virus particles fails to overcome the Muller's ratchet effect. J. Virol. 67:3620–3623.
Dunn, J. J., and F. W. Studier. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. 166:477–535.
Escarmis, C., M. Davila, and E. Domingo. 1999. Multiple molecular pathways for fitness recovery of an RNA virus debilitated by operation of Muller's ratchet. J Mol Biol 285:495–505.
Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.
———. 1988. Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19:445–471.
———. 1995. PHYLIP: phylogeny inference package. Vers. 3.5c. Distributed by the author, University of Washington, Seattle.
Foote, M. 1996. On the probability of ancestors in the fossil record. Paleobiology 22:141–151.
Frumhoff, P. C., and H. K. Reeve. 1994. Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. Evolution 48:172–180.
Garland, T., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically indpendent contrasts. Syst. Biol. 41:18–32.
Garland, T., Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by compter simulation. Syst. Biol. 42:265–292.
Garland, T., Jr., P. E. Midford, and A. R. Ives. 1999. An introduction to phylogenetically based statistical methods, with a new method for confidence intervals on ancestral values. Am. Zool. 39: 347–388.
Grafen, A. 1989. The phylogenetic regression. Phil. Trans. R. Soc. Lond. 326:119–157.
Greene, H. W. 1986. Diet and arboreality in the emerald monitor, *Varanus prasinus*, with comments on the study of adaptation. Fieldiana Zool. N.S. 31:1–12.
Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.
Harvey, P. H., and M. D. Pagel. 1991. The comparative Method in evolutionary biology. Oxford Univ. Press, Oxford.
Hillis, D. M., J. J. Bull, M. E. White, M. R. Badgett, and I. J. Molineux. 1992. Experimental phylogenetics: generation of a known phylogeny. Science 255:589–592.
Holmquist, R. 1979. The method of parsimony: an experimental test and theoretical analysis of the adequacy of molecular restoration studies. J. Mol. Biol. 135:939–958.
Huey, R. B., and A. F. Bennett. 1987. Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. Evolution 40:1098–1115.
Losos, J. B. 1990. Ecomorphology, performance capability, and scaling of West Indian Anolis lizards: An evolutionary analysis. Ecol. Monogr. 60:369–388.
———. 1992. The evolution of convergent structure in Caribbean *Anolis* communities. Syst. Biol. 41:403–420.
———. 1996. Phylogenies and comparative biology, stage II: testing causal hypotheses derived from phylogenies with data from extant taxa. Syst. Biol. 45:259–260.
Maddison, D. R. 1995. Calculating the probability distributions of ancestral states reconstructed by parsimony on phylogenetic trees. Syst. Biol. 44:474–481.
Maddison, W. P. 1991. Squared-change parsimony reconstructions of ancestral states for continuous valued characters on a phylogenetic tree. Syst. Zool. 40:304–314.
Maddison, W. P., and D. R. Maddison. 1992. MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, MA.
Martins, E. P., and T. Garland, Jr. 1991. Phylogenetic analysis of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557.
Martins, E. P., and T. F. Hansen. 1996. Phylogenetic comparative methods: the statistical analysis of interspecific data. Pp. 22–75 *in* E. P. Martins, ed. Phylogenies and the comparative method in animal behavior. Oxford Univ. Press, Oxford.
———. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am. Nat. 149:646–667.
Mitani, J. C., J. Gros-Louis, and J. H. Manson. 1996. Number of males in primate groups: comparative tests of competing hypotheses. Am. J. Primatol. 38:315–332.
Møller, A. P., and T. R. Birkhead. 1992. A pairwise comparative method as illustrated by copulation frequency in birds. Am. Nat. 139:644–656.
Nelson, G. 1989. Species and taxa: systematics and evolution. Pp. 60–81 *in* D. Otto and J. A. Endler, eds. Speciation and its consequences. Sinauer, Sunderland, MA.
Omland, K. E. 1997. Examining two standard assumptions of ancestral reconstructions: repeated loss of dichromatism in dabbling ducks (Anatini). Evolution 51:1636–1646.
Pagel, M. D. 1993. Seeking the evolutionary regression coefficient: an analysis of what comparative methods measure. J. Theor. Biol 164:191–205.
———. 1998. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.
Pagel, M. D., and P. H. Harvey. 1988. The taxon-level problem in the evolution of mammalian brain size: facts and artifacts. Am. Nat. 132:344–359.
———. 1989. Taxonomic differences in the sealing of brain on body weight among mammals. Science 244(4912):1589–1593.
Parker, A. R., D. R. McKenzie, and S. T. Ahyong. 1998. A unique form of light reflector and the evolution of signalling in *Ovalipes* (Crustacea: Decapoda: Portunidae). Proc. R. Soc. Lond. B Biol. Sci. 265:861–867.
Ryan, M. J., and A. S. Rand. 1995. Female responses to ancestral advertisement calls in Tungara frogs. Science 269:390–392.
———. 1999. Phylogenetic influence on mating call preferences in female tungara frogs, *Physalaemus pustulosus*. Anim. Behav. 57: 945–956.
Saitou, N. 1989. A theoretical study of the underestimation of branch lengths by the maximum parsimony principle. Syst. Zool. 38:1–6.
Sambrook, E., F. Fritsch, and T. Maniatis. 1989. Molecular Cloning. Cold Spring Harbor Press, Cold Spring Harbor, NY.
Schluter, D., T. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51: 1699–1711.
Schultz, T. R., R. B. Cocroft, and G. A. Churchill. 1996. The reconstruction of ancestral character states. Evolution 50:504–511.
Smith, A. B. 1994. Systematics and the fossil record. Blackwell Scientific Publications, London.
Strathmann, R. R., and D. J. Eernisse. 1994. What molecular phylogenies tell us about the evolution of larval forms. Am. Zool. 34:502–512.
Tateno, Y. 1990. A method for molecular phylogeny construction by direct use of nucleotide sequence data. J. Mol. Evol. 30: 85–93.
Wang, I. N., D. E. Dykhuizen, and L. B. Slobodkin. 1996. The evolution of phage lysis timing. Evol. Ecol. 10:545–558.
White, M. E., J. J. Bull, I. J. Molineaux, and D. M. Hillis. 1991. Experimental phylogenies from T7 bacteriophage. Dioscorides Press, Portland, OR.
Yin, J. 1991. A quantifiable phenotype of viral propagation. Biochem. Biophys. Res. Comm. 174:1009–1014.

Corresponding Editor: J. Losos