Phylogenetic Targeting of Research Effort in Evolutionary Biology

Phylogenetic Targeting of Research Effort in Evolutionary Biology

Christian Arnold[1,2*] and Charles L. Nunn[1]

[1] *Department of Human Evolutionary Biology, Harvard University, 11 Divinity Avenue, Cambridge MA 02138*

[2] *Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*

Christian Arnold: carnold@fas.harvard.edu

Charles L. Nunn: cnunn@oeb.harvard.edu

Submitted as an "Article"

Keywords: comparative method, phylogeny, correlated evolution, taxon sampling, pairwise comparison

[*] Corresponding author. Email: carnold@fas.harvard.edu

Abstract

Many questions in comparative biology require that new data be collected, either to build a comparative database for the first time or to augment existing data. Given resource limitations in collecting data, which species should be studied to increase the size of comparative datasets? By taking the hypotheses, existing data relevant to the hypotheses, and a phylogeny, we show that a method of "phylogenetic targeting" can systematically guide data collection while taking potentially confounding variables and competing hypotheses into account. Phylogenetic targeting selects potential candidates for future data collection using a flexible scoring system based on differences in pairwise comparisons. We used simulations to assess the performance of phylogenetic targeting, as compared to a less systematic approach of randomly selecting species (as might occur when data have been collected without regard to phylogeny and variation in the traits of interest). The simulations revealed that phylogenetic targeting increased the statistical power to detect correlations and that power increased with the number of species in the tree, even when the number of species studied was held constant. We also developed a web-based computer program called *PhyloTargeting* to implement the approach (http://phylotargeting.fas.harvard.edu).

INTRODUCTION

The comparative method has played a major role in uncovering adaptive trait evolution in biological systems (Harvey and Pagel 1991; Martins 2000; Pagel 1999; Ridley 1983). The comparative method has revealed, for example, links between mating systems and sperm competition in primates (Harcourt et al. 1981) and other animals (Hosken 1997; Moller 1991). The comparative method also supported a model of sexual selection in which females choose males based on their ability to resist parasites (Hamilton and Zuk 1982), and it has been used to probe the origins of both parasitic and symbiotic associations (e.g., Hugot 1999; Lutzoni et al. 2001). More recently, comparative methods have been applied to study phylogenetic community ecology (Webb et al. 2002), for example in the context of the phylogenetic over-dispersion of mammalian communities (Cooper et al. 2008). The comparative method also can be used to address conservation issues (Fisher and Owens 2004), such as questions involving the factors that influence rates of extinction (Purvis et al. 2000b) and how the phylogenetic clumping of conservation threat status can lead to greater loss of phylogenetic diversity when species go extinct (Purvis et al. 2000a).

A comparative analysis requires data on a set of species relevant to a hypothesis of interest. Usually, however, data are available for only a fraction of the species in a clade, and data collection in both the field and laboratory is expensive and time-consuming. A proper selection of species to study is a non-trivial and multi-faceted problem (Garland 2001; Westoby 2002) that has rarely been addressed in a systematic way. Instead, species are often chosen either randomly or subjectively (Faustino 2008; Westoby 1999) because they are of "particular (and perhaps irrational) interest" (Garland 2001, p.119). Two problems are introduced when species are chosen in an unsystematic way. First, the full range of variation is not used to test the hypotheses. Second, taxonomic gap bias may occur, meaning that data collection has been focused on a few "popular" lineages. These different kinds of biases – incomplete variation and gap biases – can make a momentous difference to the conclusions

69 one draws. In studies of primates, for example, results of comparative research are likely to
70 change when the sample is tilted towards terrestrial species, rather than those that live in the
71 trees, because terrestrial species possess larger body masses, exhibit different locomotor
72 patterns, and live in larger social groups (Clutton-Brock and Harvey 1977; Martin 1990; Nunn
73 and van Schaik 2002).

74 To address these issues, methods are needed to quantify potential biases in comparative
75 datasets and to identify the species that should be studied in the future. Indeed, it is common
76 to read in write-ups of comparative research that further sampling is needed to validate the
77 findings, either because the sample sizes were small or the sample was biased towards
78 particular species within a clade (e.g., in the study of sleep patterns: Capellini et al. 2009;
79 Nunn et al. 2009; Roth et al. 2006). Unfortunately often, however, only general guidelines for
80 this selection process have been given, and these guidelines are often specific to the question
81 of interest (Westoby 2002). To our knowledge, no method yet exists that is flexible and
82 specific enough to address the crucial task of prioritizing future research in light of specific
83 hypotheses about the apportionment of variation in relation to one or more ecological factors.

84 Only a handful of studies have investigated ways of systematically identifying species to
85 study. For example, Ackerly (2000) compared the performance of different taxon sampling
86 strategies and found that their statistical performance differed substantially. One of the
87 algorithms he examined is based on the pairwise comparison approach (Felsenstein 1985,
88 p.13; Maddison 2000; Møller and Birkhead 1992; Oakes 1992; Purvis and Bromham 1997;
89 Read and Nee 1995) and identifies meaningful comparisons by selecting species pairs that
90 differ by a certain amount in the independent variable, following the suggestion of Westoby
91 (1999). Although it overestimates the magnitude of the correlation, Ackerly (2000) showed
92 that this design increases the statistical power to detect correlated evolution (see also Garland
93 2001 and Garland et al. 1993). One major weakness of the method is that the threshold for
94 when differences are “large” is arbitrary, dependent on the dataset, and must be set manually,

95 which limits its applicability considerably. Mitani et al. (1996) considered sampling strategies
96 in relation to testing competing hypotheses, while Read and Nee (1995) discussed the need to
97 identify pairs that contribute for or against hypotheses. Similarly, Maddison (2000) presented
98 a methodology for choosing species pairs in which each pair is "a comparison relevant for the
99 question of interest" (p. 198). However, his method is designed for binary rather than
100 continuously varying data, and it can only handle fully bifurcating trees and thus does not
101 provide enough flexibility for identifying meaningful comparisons with real data.
102 The method of pairwise comparisons has been used frequently to identify meaningful
103 comparisons. Several reasons exist for using pairwise comparisons. For example, the method
104 of pairwise comparison relies on fewer assumptions (Ackerly 2000; Hearn and Huber 2006;
105 Maddison 2000) than other methods. Thus, unlike phylogenetically independent contrasts
106 (PIC) (Felsenstein 1985; Garland et al. 1992; Harvey and Pagel 1991), pairwise comparison
107 does not require a specific model of evolution or the estimation of states at interior nodes. In
108 addition, some sets of species within a larger clade might not be directly comparable in
109 standard implementations of comparative methods, such as PIC. In mammalian sleep, for
110 example, some cetaceans sleep with only one half of their brains (Lyamin et al. 2008), making
111 it difficult to compare the measurements of sleep in cetaceans to other mammals. The method
112 of selecting specific pairwise comparisons provides a way to limit comparisons so that
113 cetaceans are compared only to other cetaceans, and non-cetaceans are compared only to non-
114 cetaceans. Similarly, some behavioral experiments might require similar sensory modalities or
115 cognitive ability among species in the dataset. Pairwise comparisons of some close relatives
116 may be more appropriate for selecting species for focused comparative experiments that take
117 these factors into account.
118 When using the method of pairwise comparisons, it is important that all pairs are
119 phylogenetically independent, i.e. no branches are shared among the comparisons (Felsenstein
120 1985; Maddison 2000). In Figure 2, for example, different sets of phylogenetically

121 independent pairs (which we call a "pairing," see Maddison 2000) are shown for each tree.
122 Thus, when selecting phylogenetically independent pairs, the selection of a particular pair
123 constrains which other pairs can be selected.

124 Here, we present a new approach, which we call "phylogenetic targeting," to
125 systematically identify the species to study in the future. Phylogenetic targeting is a taxon
126 sampling approach that aims to prioritize future research by identifying species that should be
127 studied in a target-oriented way under consideration of the specific hypotheses and data. It is
128 not a new way to analyze comparative data or a substitute for existing analysis methods, but
129 rather draws on existing methods in comparative biology. This method uses the pairwise
130 comparisons approach and is based on a scoring system that incorporates phylogeny and data
131 on variables relevant to testing hypotheses, specifically involving the predictor and response
132 variables in a comparative test. The predictor variables can include potentially confounding
133 variables or variables relevant to testing alternative hypotheses for an association. If external
134 information suggests that comparisons should be restricted taxonomically or in relation to
135 existing data, one can use the method to limit which species to compare.

136 After assigning a score for each pair of species, phylogenetic targeting uses a newly
137 developed algorithm to select the set of phylogenetically independent pairs of species that
138 offer greater statistical power to test the hypothesis once data have been collected on the
139 dependent variable. After collecting data, pairwise contrasts for the targeted species pairs can
140 be used to test hypotheses, or one can use standard comparative techniques for testing
141 correlated character evolution (Figure 1). This decision is up to the investigator and depends
142 on the actual hypotheses, data and analysis preferences (see Discussion). We use computer
143 simulations to assess the degree to which phylogenetic targeting increases statistical power for
144 detecting correlated trait evolution, as compared to random sampling of species. We also
145 implemented the method online (http://phylotargeting.fas.harvard.edu). We anticipate that the
146 general approach developed here for pairwise comparisons can be developed for use with

147 additional comparative methods, such as PIC or generalized least squares approaches, and we
148 discuss some of these potential extensions.
149
150 **METHODS**
151 The method requires a phylogeny and one or more explicit hypotheses that offer predictions
152 for how variation in one trait (X_1) correlates with variation in another trait that is common to
153 all the hypotheses and, because it is not known in all the species, is the "target" of the analysis
154 (Y_t) (Figure 1). We call this association between Y_t and X_1 the *primary hypothesis*. Additional
155 hypotheses, if desired, are implemented through traits $X_2 \ldots X_n$, which relate to competing
156 hypotheses or potentially confounding variables. The goal of the method is to identify species
157 that should be studied with regard to Y_t by using phylogenetic relationships and data already
158 collected for the X traits. Thus, a species cannot be included in a phylogenetic targeting
159 analysis if data on X are lacking for that species. We assume that larger evolutionary changes
160 in X_1 provide higher statistical power for comparative tests to test the hypotheses, because it
161 increases the available range of variation (Garland 2001; Garland et al. 2005; Westoby 1999;
162 Westoby et al. 1998). We also assume that the characters show a linear relationship. Different
163 targeting analyses are likely to focus on a primary hypothesis and various combinations of
164 alternative hypotheses, and both discrete and continuous traits can be used. Scores are
165 calculated so that higher values indicate more preferred species to study, based on user-
166 defined criteria involving control of confounding variables, testing of alternative hypotheses,
167 and availability of data on Y_t for one or more species in a clade.
168
169 **Calculating pairwise comparisons**
170 The analysis starts by calculating all possible $n * (n-1) / 2$ pairwise comparisons. In the tree
171 shown in Figure 2, for example, 15 comparisons can be constructed. The method thus does
172 not rely on using only pairs of sister species, as pairs of more distantly related species could

also offer compelling tests of the hypotheses (Maddison 2000; Read and Nee 1995; Westoby 1999). Pairwise comparisons with missing data in any of the traits except Y_t are excluded. In addition, certain species can be excluded manually from the analysis, for example in cases where an experiment can be applied to only certain species on the tree.

If discrete characters with more than two possible states are used, they can be treated as ordered (costs between different pairs of states are different, as a particular sequence exists in which the states must occur through evolution) or unordered (every state change is equal, as each state can directly be transformed into any other state) (Slowinski 1993).

Calculating scores for models with a single predictor (Y_t and X_1)

For predictions that only involve a primary hypothesis (i.e., only one independent variable), phylogenetic targeting uses a scoring system that maximizes the variability in X_1. In other words, species pairs are targeted that differ the most in X_1. If we were interested in hypotheses that involve body mass as an independent variable, for example, phylogenetic targeting gives pairs with the largest differences in body mass higher scores. Thus, pairwise comparisons with big differences in X_1 are scored more positively, whereas smaller differences are scored less positively. These contrasts are then standardized to the scale 0 to 1, with a difference of 0 assigned a score of 0 and the largest difference in all considered pairs assigned a score of 1. Note that even if no zero contrasts are found in the data, the method fixes this as the lowest contrast. All other differences are assigned a score between 0 and 1 by applying a linear scaling transformation. We call this the *score* of X_1.

If X_1 is an unordered discrete character, the score will be either 0 or 1 regardless of the actual difference in character state assignments, whereas the difference is scored on an interval between 0 and 1 in the case of an ordered character, with the maximum number of character steps scored as 1.

Calculating scores for models with covariates (Y_t, X_1, X_2 … X_n)

Models that incorporate additional traits enable the testing of different kinds of hypotheses (e.g., mutually exclusive and non-mutually exclusive), and they can be used to control for confounding variables. For each $X_2 \ldots X_n$, a separate scoring mechanism is defined in which larger contrasts have either a negative or a positive influence on the overall score. The decision for whether larger differences in each of the X_2 to X_n variables is scored higher or lower depends on whether the variables reflect confounding variables or a desire to distinguish among competing hypotheses. To simplify discussion in what follows, we consider a case in which only one additional variable is included; thus $Y_t = f(X_1, X_2)$. Further details on the specifics of scoring are given below.

To control for confounding variables, the goal is to minimize variation in the predictor variable that corresponds to the confounding variable of interest, i.e. X_2. Thus, pairwise comparisons in X_2 that make the absolute value of change in a particular confounding variable as small as possible are scored higher, whereas pairwise comparisons with bigger differences are scored lower ($Score_{NC}$, i.e. the score from standardizing the covariate for "no change"). The smallest pairwise contrast is assigned a score of 1, whereas the maximum pairwise contrast is assigned a score of 0. All other differences are assigned a score between 0 and 1.

To address mutually exclusive hypotheses, the goal is to maximize scores for X_2 that differ maximally from contrasts in X_1. Two different scoring options can be applied that both target big differences, but differ in how they score these differences. The first option scores differences in X_2 in the opposite direction as the difference in X_1 positively and differences in the same direction as X_1 negatively ($Score_{OD}$, i.e. the score from standardizing covariate in the "opposite direction"). The biggest difference in the opposite direction is assigned a score of 1, whereas the biggest difference in the same direction is assigned a score of -1. A difference of 0 is assigned a score of 0. The smallest pairwise contrast is always assigned 0 even if no pairwise comparison has a difference of 0 in this trait, as this ensures that all non-zero

differences are assigned a score different from 0. All other differences are assigned a score between -1 and 1 by applying a linear scaling transformation, which is calculated separately for positive and negative contrasts. The second option is the opposite of the first option; that is, differences in the opposite direction from the difference in X_1 are scored negatively and differences in the same direction are scored positively ($Score_{SD}$, i.e. the score from standardizing covariate in the "same direction"). For example, this option might be useful if an increase in X_1 is predicted to reduce Y_t while an increase in X_2 is predicted to increase Y_t. Thus, it is necessary to give higher scores to contrasts in the same direction for X_1 and X_2 to distinguish among the hypotheses.

For models with covariates, the direction of change for $X_2 \ldots X_n$ always refers to the direction of change in X_1, e.g. a positive value means that the direction of change is the same as in X_1. By doing so, we force the difference in X_1 (Δ_{raw}, see Table 1) to be positive and achieve consistency with other widely-used programs, such as CAIC (Purvis and Rambaut 1995) and PDAP-Mesquite (Midford et al. 2005). This "positivization assumption" also helps to make sense of the other trait differences and their directions when using the computer program, as it becomes possible to determine whether other pairwise comparisons are consistently positively or negatively associated with X_1 (e.g., if X_2 is positive, it must be in the same direction as X_1). Although not strictly necessary for the algorithms implemented here, this helps guide manual selection of contrasts in the web-based implementation of phylogenetic targeting.

Summed score and standardizing scores for branch lengths

For each pairwise comparison, the scores for all traits are summed up to define the *summed score* (see Table 1 for a case involving X_2 as a confounding variable, i.e. $Score_{NC}$). The summed score combines the information from all traits and thus represents the strength of

a pair for testing the hypotheses. For models with only Y_t and X_1, the summed score thus equals the score of X_1.

Regardless of the scoring model, the summed score can sometimes be uninformative when compared among different pairs because the more divergent two species are, the more likely it is that they evolved bigger differences. In other words, different pairs will have different expected amounts of change (i.e., variance). In our approach, we overcome this problem by normalizing the summed score by its expected variance (square root of the sum of the branch lengths that connect the two species) (Felsenstein 1985; Garland et al. 1992). We call this the *standardized summed score*. By doing so, all pairwise comparisons have a common variance as required by most statistical tests (see also Discussion).

Table 1 summarizes and applies the scoring system to the dataset in Figure 2, based on controlling for X_2 as a confounding variable ($Score_{NC}$). Different standardized summed scores would be obtained if we treated X_2 as representing a competing hypothesis, and depending on the expected direction of X_2 in the context of competing hypotheses (see columns for $Score_{SD}$ and $Score_{OD}$ in Table 1).

Availability variable

In addition to manually excluding species from an analysis, it is possible to define an "availability variable" to automatically exclude species or pairs in relation to the availability of data for Y_t. One can thus use the availability variable to identify other species that should be studied in the context of existing data on Y_t. An availability variable also provides a way to quickly "pinpoint" where the missing data points are in a phylogenetic context, which can help to identify biases in the distribution of the studied species.

The availability variable must be a discrete binary variable that identifies whether or not data are available for Y_t for a particular species. For example, consider the scenario in Figure 2, in which B_t is the availability variable. Possible options would be to only consider

pairs where data are available for both species that form the pair (exclusion of all pairs except s1-s5), for one species (exclusion of pairs s1-s5 and all combinations of s2, s3, s4 and s6), for at least one species (as before, but not s1-s5) and for none of the species (exclusion of the nine pairs with s1 and s5). This scoring procedure thus can be used in a variety of ways. For example, if the availability variable indicates that data are available for only a fraction of the species, the majority of the pairs will be excluded if the option is chosen to consider only pairs where one species has already been studied and data are needed for the other species. In such a case, only those pairs containing one studied species and one that has yet to be studied remain. It can thus be seen as an additional selection factor that effectively constrains the species that will be targeted.

Maximal pairing algorithm

The actual selection of species is performed by a dynamic programming algorithm that we call maximal pairing. The maximal pairing algorithm is a general optimization algorithm and selects pairs of species that are phylogenetically independent. In contrast to PIC, where pairs can also involve internal nodes on the tree, the maximal pairing algorithm selects only pairs between the tips of the tree. The selection of pairs is based on the summed score for each pair, and the algorithm determines the set of phylogenetically independent pairs that maximizes the sum of the individual summed scores (Table 1). This criterion is thus assumed to maximize the power to test the hypotheses given constraints on maintaining phylogenetic independence. With large datasets, it is difficult to find the maximal pairing manually, due to the large number of possible pairings and the complex phylogenetic dependence of pairs that must not share a branch (Figure 2). Despite some differences that involve execution time and representation of polytomies, the maximal pairing algorithm also works for polytomous trees (see Online Appendix A for more details).

For models that involve only X_1, for example, the maximal pairing generally selects pairs of closely related species that maximize differences in X_1, and those pairs are often distantly related to the other pairs that are selected. In a comparative test, such a design is considered to be especially powerful (Garland et al. 2005). If, however, an additional trait X_2 is used to control for confounding variables (thus scoring small differences in X_2 higher using Score_{NC}), the algorithm both maximizes differences in X_1 and minimizes differences in X_2. Conversely, if one aims to maximize differences in X_2 (thus scoring larger differences in X_2 opposite to X_1 higher with Score_{OD}), the algorithm maximizes differences in X_1 and maximizes differences in X_2 opposite in sign to X_1. Similar logic applies to Score_{SD}. It is worth noting, however, that due to the phylogenetic constraints and the standardizing of contrasts, the maximal pairing does not simply select the pairs with the most extreme character differences; instead, pairs with small differences among closely related species are also frequently selected.

Simulations

We compared the performance of phylogenetic targeting to random selection of species using simulations. The aim of the simulations was to generate data with known degrees of correlation between pairs of variables, and then to select subsets of species either randomly or using phylogenetic targeting. To perform the simulations, we first generated phylogenetic trees and character data using the GEIGER package (Harmon et al. 2008) in *R* (R Development Core Team 2009) according to a uniform birth-death process (b=0.15, d=0). We created 1500 random phylogenies for a series of N=50, 70, and 90 taxa. We then simulated character evolution for two continuously varying characters on each tree using five different models of evolution (Table 2) with character states (0,0) at the root of the tree. When simulating the non-Brownian motion models of evolution, we first transformed the tree in Geiger (Harmon et al. 2008), simulated traits on the transformed tree, and then analyzed the

data on the original tree, thus simulating a case where the branch lengths failed to accurately reflect trait evolution (see Online Appendix B). Characters were simulated with a variance of one and correlations of 0 and 0.5, respectively. This yielded 4500 datasets with varying numbers of species and known evolutionary correlations among the characters.

Using these data and phylogenies, we then selected subsets of species randomly and using phylogenetic targeting. In each simulation file, we selected the first simulated trait as X_1; the second variable was assumed to be Y_t. We also standardized the scores. The maximal pairing was then calculated, and we selected the six highest scoring pairs. We also randomly selected six phylogenetically independent pairs. To investigate whether the number of selected pairs impacts statistical performance, all analyses were repeated using 9 pairs and 12 pairs.

To evaluate statistical properties of both sampling approaches, we performed standard statistical tests based on the selected pairwise comparisons. For that, we used the character differences for X_1 and Y_t for the selected pairs and standardized them by their expected variance (square root of the sum of the branch lengths that connect the two species). We tested for a significant correlation between both characters using the correlation coefficient through the origin (Garland et al. 1992), with significance based on $\alpha = 0.05$ using a t-test with N-2 degrees of freedom. We determined Type I error rates (incorrectly rejecting a true null hypothesis of no association between traits) and statistical power (probability of rejecting a false null hypothesis) for both sampling approaches. Type I error rates were calculated as the proportion of significant results based on p=0.05 for datasets in which r=0, while statistical power was based on the proportion of significant results for datasets in which r=0.5.

In addition to tests based on pairwise comparisons, we performed tests based on the full set of independent contrasts. We did this because many users may be interested in using a full set of contrasts, yet the method operates by examining pairwise comparisons. Thus, understanding the statistical performance of phylogenetic targeting when used with PIC is an

353 important step and expands its application spectrum. After pruning the tree to the subset of
354 selected pairs, we calculated PIC (Felsenstein 1985) using the APE package (Paradis et al.
355 2004). We tested for a significant correlation between both characters using the methods
356 described in the previous paragraph.
357 We also tested how the inclusion of randomly selected, non-targeted species affects the
358 results. This simulates a common situation because data are often already available for some
359 species but missing for others. Specifically, we examined how including *k* random species
360 affects the results for tests based on pairwise comparisons and PIC (with *k* ranging from 2 to
361 10 in steps of 2). We included these additional species from the remaining set of species that
362 were not selected by phylogenetic targeting (and thus without using the availability variable).
363 Lastly, we analyzed how much of the original range of variation in the simulated data was
364 available after pruning the data to the selected species. This gives insights to the range of
365 variation that is available for hypothesis testing under the two sampling techniques.
366
367 **RESULTS**
368 **PhyloTargeting program**
369 We created a freely available computer program – *PhyloTargeting* – that implements the
370 phylogenetic targeting approach. It is web-based, takes the data as a Nexus file (Maddison et
371 al. 1997) and provides a user-friendly, interactive, step-by-step interface, a variety of analysis
372 options, and graphical visualizations of the results. The program is publicly available at
373 http://phylotargeting.fas.harvard.edu.
374
375 **Simulations**
376 The simulations revealed that phylogenetic targeting substantially increases the range of
377 biological variation that is sampled relative to random sampling (Figure 4). Phylogenetic
378 targeting also provided substantially higher statistical power for detecting a true relationship

(Figure 5). This held for both the pairwise tests and tests based on PIC. For the pairwise tests, Type 1 error rates for $\alpha = 0.05$ were elevated if the number of selected pairs was small, but decreased to the expected level when more pairs were selected. For the tests based on PIC, Type I error rates were close to the expected level in all scenarios. Importantly, Type 1 error rates under random sampling and phylogenetic targeting were generally indistinguishable. More details are provided in Online Appendix C.

Increasing the number of pairs that are selected by the sampling algorithms increased statistical power, as expected (Figure 5). For the pairwise tests, it also decreased Type 1 error rates. The number of taxa per tree, however, revealed a more surprising effect. Even when holding the number of pairs constant, the statistical power increased with the number of taxa in the clade under phylogenetic targeting, and Type 1 error rates did not increase (Figure 5). If species are selected randomly, however, power did not increase with increasing clade size.

When the true correlation was 0.5, mean values of *r* were elevated, and moreover increased with the number of species per tree (see Online Appendix C). Thus, a sampling regime based on phylogenetic targeting resulted in biased estimates of evolutionary trait correlations when $r \neq 0$, whereas a random selection of species resulted in no bias. Importantly, however, no bias was found when the true correlation was 0, as shown in the results for Type I error rates. Furthermore, the bias decreased substantially if additional, randomly selected species were included (see Discussion and Online Appendix C).

The results highlighted above are for a Brownian motion process of character evolution. For the alternative models that we tested (see Online Appendix B), results were comparable. However, for most of these analyses, Type 1 error rates were highly elevated and statistical power was reduced under the two sampling approaches and for PIC on the full tree (which we used as a control). Not surprisingly, the pairwise tests showed substantially less elevated Type 1 error rates if model assumptions were violated, possibly because the method of pairwise comparisons relies on fewer assumptions.

DISCUSSION

Comparative studies generally make use of available data. Here we show that the comparative approach can also be used to target species for future data collection. By applying the phylogenetic targeting concept, we can identify species that offer higher power to test predictions of a comparative hypothesis. Moreover, phylogenetic targeting provides a way to control for confounding variables when selecting species for further study, or to test competing hypotheses. The method will most likely be used to augment existing data, but it can also be applied to generate new datasets in the context of finite resources for data collection.

A major strength of the approach is that phylogenetic information is incorporated when selecting species to study (Garland 2001; Garland et al. 2005), thus ensuring that the selected pairs are phylogenetically independent of one another. This makes it possible to analyze the data using standard statistical methods (i.e., pairwise tests). However, the simulations revealed that compared to PIC, statistical power is reduced (see also Ackerly 2000). This may be due to the fact that for pairwise differences, the number of data points is reduced by a factor of approximately 2, because only the tips of the tree are contrasted and not the interior nodes of the tree. Furthermore, the bias in estimating the correlation coefficient is increased with pairwise comparisons. We thus advise users to analyze the selected species with standard comparative methods based on the full set of contrasts whenever possible instead of using the differences for the selected pairs directly.

The simulation results revealed that phylogenetic targeting provides many advantages compared to a random selection of species for detecting correlated trait evolution. Statistical power was strongly increased in all cases that we examined. Phylogenetic targeting used a higher percentage of the available range of variation for a character, as compared to random sampling of species. Thus, we can be more certain that the pattern holds generally across the

clade of organisms rather than, for example, only among the species that are larger in body size or more amenable to study. Surprisingly, the simulations also revealed that statistical power increased with the number of species per tree, even when the number of taxa selected for study remained constant. Type 1 errors, however, were always close to the nominal level and undistinguishable between phylogenetic targeting and random species sampling. Thus, applying the method to larger clades resulted in increased power without increasing the number of pairs examined, probably because having more taxa increased the magnitude of the differences that can be selected overall (which increased the ability to detect a correlation).

Phylogenetic targeting should be used with caution when one wants to determine the magnitude of a correlation. Similar to the pairwise approach of Westoby (1999), it overestimates the correlation coefficient (Ackerly 2000). This was true for both the pairwise tests and PIC, and the bias was stronger with the pairwise tests. The simulations also revealed that this overestimation increases with the number of species per tree, thus mirroring the increase in power. In the context of applying the method to real-world data in which data for Y_t are already available for some of the species, however, simulations confirmed that this bias decreases substantially with the number of randomly selected species for which data are already available. For most questions of interest that we envision, data are often available on Y_t for a number of species, often comprising a majority of the species in the dataset. When such data are available, inclusion of already available data in subsequent analysis after applying phylogenetic targeting is highly recommended. Alternatively, users can implement the availability variable option described above to more fully integrate decisions about future data collection with already studied species. Furthermore, as noted above, the bias is likely to decrease if additional traits representing confounding variables or alternative hypotheses are included in the analysis.

A few limitations and assumptions of phylogenetic targeting should be noted. Although the maximal pairing selects the set of species pairs that have the highest overall score

457 according to a user-defined scoring model, it may select species that are not directly
458 comparable in relation to a particular test, such as an experiment that involves testing
459 cognitive abilities. To overcome this possible weakness, our *PhyloTargeting* program
460 provides a way for the user to select pairs in which particular comparisons are possible and to
461 exclude other comparisons. Phylogenetic targeting must be used with caution if non-linear
462 relationships between the variables can be assumed, and we advise users to critically examine
463 the variables beforehand. Another critical issue is the phylogenetic tree, the representation of
464 polytomies (see Online Appendix B), and the branch lengths on which the species selection is
465 based. The selection of species can vary substantially between similar tree topologies due to
466 the fact that the maximal pairing algorithm strictly maximizes the overall score, which can
467 sometimes be heavily influenced by the topology. Branch lengths are assumed to be
468 proportional to the expected variance in the amount of evolutionary changes along each
469 branch (Brownian motion), which becomes an important assumption both in phylogenetic
470 targeting and in subsequent analyses. This is particularly true for PIC. If these assumptions are
471 violated, Type 1 error rate are inflated and statistical power is reduced (Diaz-Uriarte and
472 Garland 1996; Quader et al. 2004) . Indeed, the simulations confirmed this effect; for almost
473 all of the alternative models, Type 1 error rates were highly elevated. The only exception is
474 the early burst model, which yielded results very similar to those for Brownian motion
475 (Online Appendix C).

476 Because sister taxa will tend to be similar in many ways, confounding variables are
477 expected to be less of a problem in sister-species comparisons (Harvey and Pagel 1991;
478 Møller and Birkhead 1992). In our approach, however, more distantly related species pairs
479 can also be selected. That can be critical, because other, unmeasured confounding variables
480 may be introduced to the analysis. The comparison of distantly related species is comparable
481 to an experiment with multiple uncontrolled variables (Garland 2001; Garland and Adolph
482 1994). The more distantly related two species are, the more likely it is that such an effect

483 could bias the results. By including additional variables in the calculations, it is possible to
484 control for some confounds when measurements are available.

485 We recommend that users standardize pairs to meet statistical requirements of
486 subsequent statistical tests (i.e., equal variances among pairs). Standardization has not
487 typically been implemented for pairwise comparisons, but it is necessary if one wishes to use
488 parametric statistical tests that make assumptions about homoskedasticity. When contrasts are
489 standardized, distantly related pairs are less often selected. This may be useful if large
490 differences are only informative when the species are closely related (e.g., to control for
491 possibly unknown confounding variables), or when comparisons should be made between
492 closely related species (e.g., because of biological differences that limit comparability of
493 experimental results). Standardization thus affects the selection of pairs.

494 Another argument for standardization is that fewer traits should change on shorter
495 branches, and thus it helps control for confounding variables. However, standardization may
496 exaggerate evolutionary differences for close relatives when differences are due to sampling
497 error or within-species variation (Purvis and Webster 1999). It can thus overestimate the
498 importance of certain species pairs if they are close relatives. We may sometimes expect a
499 larger absolute change in some trait, regardless of its rate of change, to be more valuable in
500 testing a hypothesis than a small change over a short branch. For example, brain size that
501 increases by an order of magnitude might be a stronger test than a smaller amount of brain
502 change, even if it occurs over a small branch. Using the program that we provide, the choice
503 of standardization is left up to the user (with the default option to standardize scores), based
504 on his or her preferences, the assumptions of subsequent methods, and particulars of the
505 biological system.

506 Phylogenetic targeting works best for continuous traits, but it can also be used with
507 discrete traits. However, phylogenetic targeting purely based on discrete characters is more
508 challenging because the number of distinct differences is typically smaller. In such cases, it is

common to find that numerous pairs have the maximal possible score. This will ultimately result in multiple optimal solutions in the maximal pairing algorithm. However, as the current implementation returns only one optimal solution, it is difficult to evaluate its uniqueness. Possible workarounds would be to either add a continuous variable or to standardize contrasts, both of which help to generate variation in the scores and thus to decide among the possible pairs of taxa.

The maximal pairing algorithm falls in a class of general combinatorial optimization problems that are of considerable interest in comparative phylogenetics and bioinformatics more generally. Several modifications of this algorithm have practical importance as well. For example, the algorithm could be modified to select only a fixed number of pairs (given by the researcher), thus incorporating the fact that limited resources are available to select species for future study. This important variant has already been implemented elsewhere (see Arnold and Stadler 2010). It might also be desirable to take into account conservation status of different species, to ensure that species are studied before they go extinct. More generally, the selection of species could be based not solely on pairwise comparisons, but on the full set of contrasts, possibly in combination with examining the raw data space or regularly sampling character values along the entire range of a character of interest. Here, we laid down the foundations for systematically identifying species for future study. Many possible extensions and modifications of the approach are possible, particularly related to alternative ways of sampling species.

In summary, we provided a systematic method to select species for future study that offers greater statistical power to test adaptive hypotheses as compared to a random selection of species. With this method of phylogenetic targeting, it is also possible to control for confounding variables, to incorporate alternative hypotheses, and to make use of existing data on the trait of interest. It thus provides a way to guide the selection of species relative to *a*

priori hypotheses. Through our web-based computer program, other researchers are able to easily implement the approach in a flexible and user-friendly way.

Acknowledgements

We want to thank all people who contributed to this research, especially Peter F. Stadler, Liam Revell, and Luke J. Matthews. This research was supported by grant number BCS-0923791 from the National Science Foundation, the Max Planck Society, University of Leipzig and Harvard University.

ONLINE APPENDIX A: THE MAXIMAL PAIRING PROBLEM

The *Maximal Pairing Problem* (MPP) is the prototype of a class of combinatorial optimization problems with considerable interest in bioinformatics and comparative phylogenetics: Given an arbitrary phylogenetic tree T and weights ω_{xy} for the paths between any two pairs of species (x, y) (which measures the benefit or our amount of information contributed by including the comparison of species x with species y), what is the collection of phylogenetically independent paths between pairs of leaves (i.e., no edge is shared twice) that maximizes the total weight?

In what follows, we provide algorithmic details for the implemented version for how to compute the solution of the MPP, which we call *maximal pairing* (MP) (see also Arnold 2008; Arnold and Stadler 2010).

The algorithm proceeds from the root of the tree up to the leaves. Solutions of sub-problems (i.e., the MP of trees rooted at nodes other than the root node) are tabulated and thus do not have to be recalculated. The score for the MP for a particular tree rooted at *u,* denoted $S_{T(u)}$, can be decomposed into two cases. First, the MP of *T(u)* may exclusively consist of pairs that do not go through *u* itself. All pairs that contribute to $S_{T(u)}$ are thus located in the trees rooted at the children of *u*, denoted *chd(u)*. $S_{T(u)}$ therefore equals the sum of S_k for each $k \in$ *chd(u)*. To calculate $S_{T(u)}$, it is thus sufficient to recursively call all children of *u*.

The second case is more complex. Here, at least one pair, denoted r_u, with *u* as the least common ancestor belongs to the MP of T(u), and $S_{T(u)}$ is thus composed of the score of S_{r_u} and the sum of the scores from the MP of all leftover subtrees that arise when the branches from r_u are allocated in the tree, denoted *subtrees*(r_u). To calculate $S_{T(u)}$, however, we have to find the particular pair r_u that maximizes $S_{T(u)}$ for the second case (see also Figure A1). All subtrees *k* with $k \in$ *subtrees*(r_u) are then called recursively. The procedure becomes much more complex if polytomous nodes (degree > 2) are involved, due to the fact that more than

568 one pair can go through the polytomous node without violating phylogenetic independence. In
569 the current implementation, the MP algorithm calls polytomous nodes multiple times to find
570 the combination of pairs that maximizes the score of the MP for the second case by using a
571 brute force approach (for more details, see Arnold 2008).

572 These two distinct cases allow a decomposition of the initial problem into smaller
573 problems (dynamic programming). The recursions stop for subtrees with degree = 0, i.e. the
574 tips of the tree, as their score is always 0. Ultimately, this leads to the following recursion
575 formula:

576 $$S_u = \max \begin{cases} \sum_{k \in chd(u)} S_k \\ \max_{r_u} (S_{r_u} + \sum_{k \in subtrees(r_u)} S_k) \end{cases}$$

577 , with the notation introduced above. Figure A1 shows a graphical representation of the
578 recursion formula. After comparing the scores for both cases, the higher-scoring case is
579 selected, and the score and some additional information needed for the backtracing are
580 tabulated.

581 Finally, a backtracing procedure is applied to reconstruct the solution (i.e. the set of
582 phylogenetically independent pairs), based on the information collected in the forward
583 recursions.

584 For binary trees, the forward recursions can be computed in $O(n^3)$ time and $O(n^2)$
585 space. If the tree is balanced, only $O(n^2 \log_2 n)$ time is needed. Backtracing can be computed
586 in $O(n^2)$ time. For polytomous nodes p, execution time for the MP of the tree rooted at p is
587 increased exponentially by a factor 2^{d-2} that accounts for multiple calls of p (see above).
588 Execution time for polytomous trees can be improved to an overall polynomial-time algorithm
589 by building auxiliary graphs for each polytomous node and solving maximum weighted
590 matching problems (Arnold and Stadler 2010)

591 The MP algorithm works for arbitrary trees, including trees with polytomies. Hard and
592 soft polytomies are treated differently, as follows. If the polytomy is defined as hard (i.e. split
593 into more than two lineages), multiple pairs can go through the polytomous node without
594 violating phylogenetic independence. Polytomies that are defined as a series of zero-branches
595 (soft polytomies), however, are treated as a series of true dichotomies. Here, in most cases,
596 fewer pairs can be selected, due to the fact that no branch can be shared twice. Treating
597 polytomies as soft reduces execution time. Zero-length branches should be treated with
598 caution, however, since the arbitrary order of zero-branches might change the MP
599 considerably.
600

ONLINE APPENDIX B: ALTERNATIVE MODELS OF EVOLUTION FOR SIMULATIONS

We tested the narrow sense validity, in which the characters evolved on the randomly generated tree under Brownian motion, and then investigated the broad sense validity in which the characters evolved under different evolutionary models that were assumed to be unknown to the user. To implement different evolutionary models, we transformed the tree using the Geiger package (Harmon et al. 2008), evolved the characters with a particular model on the transformed tree under Brownian motion, and used the original tree for the subsequent steps. We investigated four different models that characterize stabilizing selection (the Ornstein-Uhlenbeck model) (Hansen 1997), an adaptive radiation model in which most change occurs early in the evolutionary history of the clade (Freckleton et al. 2003; Price 1997), a speciational model in which branches were equal, and a transformation of the tree corresponding to weaker levels of phylogenetic signal (Freckleton et al. 2002; Pagel 1999). Table B1 provides more details on the models and their parameters.

ONLINE APPENDIX C: SIMULATION RESULTS

All simulation results (including the results not highlighted in the manuscript) are provided in the file “Simulation results.xls”.

REFERENCES

Ackerly, D. D. 2000. Taxon sampling, correlated evolution, and independent contrasts. Evolution 54:1480-1492.

Arnold, C. 2008. Phylogenetic Targeting: A Systematic Approach and Computer Program for Targeting Research Effort in Comparative Evolutionary Biology, University of Leipzig, Leipzig.

Arnold, C., and P. F. Stadler. 2010. Polynomial algorithms for the Maximal Pairing Problem: Efficient phylogenetic targeting on arbitrary trees. Algorithms for Molecular Biology 5:25.

Capellini, I., P. McNamara, B. T. Preston, C. L. Nunn, and R. A. Barton. 2009. Does Sleep Play a Role in Memory Consolidation? A Comparative Test. PLoS ONE 4:e4609.

Clutton-Brock, T. H., and P. H. Harvey. 1977. Primate ecology and social organization. Journal of Zoology 183:1-39.

Cooper, N., J. Rodriguez, and A. Purvis. 2008. A common tendency for phylogenetic overdispersion in mammalian assemblages. Proceedings of the Royal Society B-Biological Sciences 275:2031-2037.

Diaz-Uriarte, R., and T. Garland. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. Systematic Biology 45:27-47.

Faustino, C. E. S. 2008. Designing a shipboard line transect survey to estimate cetacean abundance off the Azores Archipelago,Portugal, St. Andrews University, St Andrews, Fife, Scotland, UK.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125:1-15.

Fisher, D. O., and I. P. F. Owens. 2004. The comparative method in conservation biology. Trends in Ecology & Evolution 19:391-398.

643 Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative
644 data: A test and review of evidence. American Naturalist 160:712-726.
645 Freckleton, R. P., M. Pagel, and P. H. Harvey. 2003. Comparative methods for adaptive
646 radiations, Pages 391-407 *in* T. M. Blackburn, and K. J. Gaston, eds. Macroecology:
647 Concepts and consequences. Oxford, Blackwell.
648 Garland, T. 2001. Phylogenetic comparison and artificial selection - Two approaches in
649 evolutionary physiology, Pages 107-132 *in* R. C. Roach, P. D. Wagner, and P. H.
650 Hackett, eds. From Genes to the Bedside (Advances in Experimental Biology and
651 Medicine). New York, Kluwer Academic/Plenum Publishers.
652 Garland, T., A. F. Bennett, and E. L. Rezende. 2005. Phylogenetic approaches in comparative
653 physiology. Journal of Experimental Biology 208:3015-3035.
654 Garland, T., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of
655 covariance by computer simulation. Systematic Biology 42:265-292.
656 Garland, T., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative
657 data using phylogenetically independent contrasts. Systematic Biology 4:18-32.
658 Garland, T., Jr., and S. C. Adolph. 1994. Why not to do two-species comparative studies:
659 limitations on inferring adaptation. Physiological Zoology 67:797-828.
660 Hamilton, W. D., and M. Zuk. 1982. Heritable true fitness and bright birds: a role for
661 parasites? Science 218:384-387.
662 Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation.
663 Evolution 51:1341-1351.
664 Harcourt, A. H., P. H. Harvey, S. G. Larson, and R. V. Short. 1981. Testis weight, body
665 weight and breeding system in primates. Nature 293:55-57.
666 Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER:
667 investigating evolutionary radiations. Bioinformatics 24:129-131.

668 Harvey, P. H., and M. D. Pagel. 1991, The Comparative Method in Evolutionary Biology:
669 Oxford Series in Ecology and Evolution. Oxford, Oxford University Press.
670 Hearn, D., and M. Huber. 2006. The ancestral distance test: What relatedness can reveal about
671 correlated evolution in large lineages with missing character data and incomplete
672 phylogenies. Systematic Biology 55:803-817.
673 Hosken, D. J. 1997. Sperm Competition in Bats. Proceedings of the Royal Society of London
674 - Series B: Biological Sciences 264:385-392.
675 Hugot, J. P. 1999. Primates and their pinworm parasites: The Cameron hypothesis revisited.
676 Systematic Biology 48:523-546.
677 Lutzoni, F., M. Pagel, and V. Reeb. 2001. Major fungal lineages are derived from lichen
678 symbiotic ancestors. Nature 411:937-940.
679 Lyamin, O. I., P. R. Manger, S. H. Ridgway, L. M. Mukhametov, and J. M. Siegel. 2008.
680 Cetacean sleep: An unusual form of mammalian sleep. Neuroscience and
681 Biobehavioral Reviews 32:1451-1484.
682 Maddison, D. R., D. L. Swofford, and W. P. Maddison. 1997. Nexus: An extensible file
683 format for systematic information. Systematic Biology 46:590-621.
684 Maddison, W. P. 2000. Testing character correlation using pairwise comparisons on a
685 phylogeny. Journal of Theoretical Biology 202:195-204.
686 Martin, R. D. 1990, Primate Origins and Evolution. London, Chapman and Hall.
687 Martins, E. P. 2000. Adaptation and the comparative method. Trends in Ecology and
688 Evolution 15:296-299.
689 Midford, P. E., T. Garland, and W. P. Maddison. 2005.PDAP Package of Mesquite, version
690 1.08(2).
691 Mitani, J. C., J. GrosLouis, and J. H. Manson. 1996. Number of males in primate groups:
692 Comparative tests of competing hypotheses. American Journal of Primatology 38:315-
693 332.

694 Moller, A. P. 1991. Sperm competition, sperm depletion, paternal care, and relative testis size
695 in birds. American Naturalist 137:882-906.
696 Møller, A. P., and T. R. Birkhead. 1992. A pairwise comparative method as illustrated by
697 copulation frequency in birds. American Naturalist 139:644-656.
698 Nunn, C. L., P. McNamara, I. Capellini, P. Preston, and R. A. Barton. 2009. Primate sleep in
699 phylogenetic perspective *in* P. McNamara, R. A. Barton, and C. L. Nunn, eds.
700 Evolution of sleep: Phylogenetic and Functional Perspectives. Cambridge, Cambridge
701 University Press.
702 Nunn, C. L., and C. P. van Schaik. 2002. Reconstructing the behavioral ecology of extinct
703 primates, Pages 159-216 *in* J. M. Plavcan, R. F. Kay, W. L. Jungers, and C. P. v.
704 Schaik, eds. Reconstructing Behavior in the Fossil Record. New York, Kluwer
705 Academic/ Plenum.
706 Oakes, E. J. 1992. Lekking and the evolution of sexual dimorphism in birds: comparative
707 approaches. American Naturalist 140:665-684.
708 Pagel, M. 1999. Inferring the historical patterns of biological evolution. Nature 401:877-884.
709 Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution
710 in R language. Bioinformatics 20:289-290.
711 Price, T. 1997. Correlated evolution and independent contrasts. Philosophical Transactions of
712 the Royal Society, London 352:519-529.
713 Purvis, A., P. M. Agapow, J. L. Gittleman, and G. M. Mace. 2000a. Nonrandom extinction
714 and the loss of evolutionary history. Science 288:328-330.
715 Purvis, A., and L. Bromham. 1997. Estimating the transition/transversion ratio from
716 independent pairwise comparisons with an assumed phylogeny. Journal of Molecular
717 Evolution 44:112-119.

718 Purvis, A., J. L. Gittleman, G. Cowlishaw, and G. M. Mace. 2000b. Predicting extinction risk
719 in declining species. Proceedings of the Royal Society of London Series B-Biological
720 Sciences 267:1947-1952.
721 Purvis, A., and A. Rambaut. 1995. Comparative analysis by independent contrasts (CAIC):
722 an Apple Macintosh application for analysing comparative data. Computer
723 Applications in the Biosciences 11:247-251.
724 Quader, S., K. Isvaran, R. E. Hale, B. G. Miner, and N. E. Seavy. 2004. Nonlinear
725 relationships and phylogenetically independent contrasts. Journal of Evolutionary
726 Biology 17:709-715.
727 R Development Core Team. 2009.R: A language and environment for statistical computing. R
728 Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-
729 project.org.
730 Read, A. F., and S. Nee. 1995. Inference from binary comparative data. Journal of Theoretical
731 Biology 173:99-108.
732 Ridley, M. 1983, The explanation of organic diversity: the comparative method and
733 adaptations of mating. Oxford, Clarendon.
734 Roth, T. C., J. A. Lesku, C. J. Amlaner, and S. L. Lima. 2006. A phylogenetic analysis of the
735 correlates of sleep in birds. Journal of Sleep Research 15:395-402.
736 Slowinski, J. B. 1993. "Unordered" Versus "Ordered" Characters. Systematic Biology:155-
737 165.
738 Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and
739 community ecology. Annual Review of Ecology and Systematics 33:475-505.
740 Westoby, M. 1999. Generalization in functional plant ecology: the species sampling problem,
741 plant ecology strategy schemes, and phylogeny, Pages 847–872 *in* F. I. Pugnaire, and
742 F. Valladares, eds. Handbook of functional plant ecology. New York, M. Dekker.
743 —. 2002. Choosing species to study. Trends in Ecology & Evolution 17:587-587.

744 Westoby, M., S. A. Cunningham, C. Fonseca, J. Overton, and I. J. Wright. 1998. Phylogeny
745 and variation in light capture area deployed per unit investment in leaves: designs for
746 selecting study species with a view to generalizing, Pages 539–566 *in* H. Lambers, H.
747 Poorter, and M. M. I. V. Vuren, eds. Variation in growth rate and productivity of
748 higher plants. Leiden, The Netherlands, Backhuys Publishers.
749
750

FIGURES

Figure 1. Flow chart for applying phylogenetic targeting. Phylogenetic targeting is essentially a taxon sampling technique to systematically guide future data collection.

Figure 2. Three out of the 15 possible pairings for an example tree. Paired species are highlighted in black. One pairing has three pairs, ten pairings two pairs, and four only one pair. In all pairings, pairs are phylogenetically independent, and no additional pair can be added without violating the requirement of phylogenetic independence.

Figure 3. Example dataset and phylogeny for applying phylogenetic targeting. The tree shows continuously varying traits X_1, X_2, Y_t and a binary trait B_t indicating whether the species has already been studied in relation to Y_t. Two species have already been studied regarding Y_t, and data on Y_t are missing for four species. The goal is to identify which of the four unstudied species should be targeted for studying Y_t.

Figure 4. Results from the simulations. Simulation results for the percentage of the used range of variation for X_1 when species pairs are selected using phylogenetic targeting (dark grey) and randomly (light grey) are shown. The x-axis plots the effects of the number of pairs that have been selected (6, 9, and 12). Contrast standardization is turned on.

Figure 5. Selected results from the simulations under Brownian motion. Type I errors and statistical power for correlation tests based on pairwise comparisons (PC, left category) and phylogenetically independent contrasts (PIC, right category) are shown for phylogenetically targeted sampling ("PT") and random taxon sampling ("R"). The first three bars in each category represent Type I error rates (based on 50, 70, and 90 species tree; from left to right),

776 and the last three bars represent statistical power (also based on 50, 70, and 90 species tree;
777 from left to right). Contrast standardization is turned on, and six pairs were selected.
778
779 Figure A1. Graphical representation of the recursion formula of the maximal pairing
780 algorithm for bifurcating nodes. Calculation of the maximal pairing proceeds recursively from
781 the root to the tips. For each internal node, two distinct cases can be distinguished that allow a
782 decomposition of the initial problem into smaller problems (dynamic programming). The
783 higher-scoring case is selected and the recursion proceeds. Note that for polytomous nodes, a
784 different algorithm is used (not shown here). See text for details.
785

TABLES

TABLE 1. ILLUSTRATION OF THE SCORING SYSTEM AND THE MAXIMAL PAIRING, APPLIED TO FIGURE 2.

Pairwise comparison	X_1		X_2				Summed score	Sum of branch lengths	Standardized summed score
	Δ_{Raw}	Score	Δ_{Raw}	Score					
				$Score_{NC}$	$Score_{SD}$	$Score_{OD}$			
s1-s2*	0.5	0.385	-3	0.831	-0.171	0.171	1.216	6	0.496
s1-s3	0.8	0.615	-1.5	0.916	-0.086	0.086	1.531	6	0.625
s1-s4	1.3	1	-2.7	0.848	-0.154	0.154	1.848	6	0.755
s1-s5	1	0.769	14.8	0.169	0.831	-0.831	0.938	8	0.332
s1-s6	0.6	0.462	9.6	0.461	0.539	-0.539	0.922	8	0.326
s2-s3	0.3	0.231	1.5	0.916	0.084	-0.084	1.146	4	0.573
s2-s4	0.8	0.615	0.3	0.983	0.017	-0.017	1.599	4	0.799
s2-s5	0.5	0.385	17.8	0	1	-1	0.385	8	0.136
s2-s6	0.1	0.077	12.6	0.292	0.708	-0.708	0.369	8	0.13
s3-s4*	0.5	0.385	-1.2	0.933	-0.069	0.069	1.317	2	0.931

s3-s5	0.2	0.154	16.3	0.084	0.916	-0.916	0.238	8	0.084
s3-s6	0.2	0.154	-11.1	0.376	-0.634	0.634	0.53	8	0.187
s4-s5	0.3	0.231	-17.5	0.017	-1	1	0.248	8	0.088
s4-s6	0.7	0.538	-12.3	0.309	-0.703	0.703	0.847	8	0.3
s5-s6*	0.4	0.308	5.2	0.708	0.292	-0.292	1.016	2	0.718

NOTE.– Δ_{Raw} = raw difference of trait values (see Figure 2). See scoring section for details on $Score_{NC}$, $Score_{SD}$, and $Score_{OD}$. Calculation of the summed score based on the score of X_1 and the $Score_{NC}$ scoring option for X_2; sum of branch lengths according to the tree in Figure 2. Pairs that are selected in the maximal pairing are indicated by * in the leftmost column.

TABLE B1. MODELS OF EVOLUTION USED IN THE SIMULATIONS.

Model of evolution	**Description of the model**	**Parameters in the GEIGER package**
Brownian motion	constant-rate random-walk model	None
Ornstein-Uhlenbeck	random-walk model with a central tendency, so that phenotypes tend to evolve towards one "optimal" value[1]	α = 0.5, 1, and 2
Adaptive radiation / Early burst	rate of evolution decays exponentially through time	endRate=0.3 and 0.6
Speciational/ Punctuated	all branches have length 1	None
Lambda transformation	The parameter λ is a scaling parameter that can be used to estimate phylogenetic signal. Decreasing the value of λ has the effect of gradually eliminating phylogenetic structure. Under Brownian motion, λ takes the value 1.0 by default. If the Brownian motion assumption is violated, however, λ will significantly depart from 1.0.	λ=0.3 and 0.6

NOTE.– [1] here: the ancestral state for the character

Fig. 1

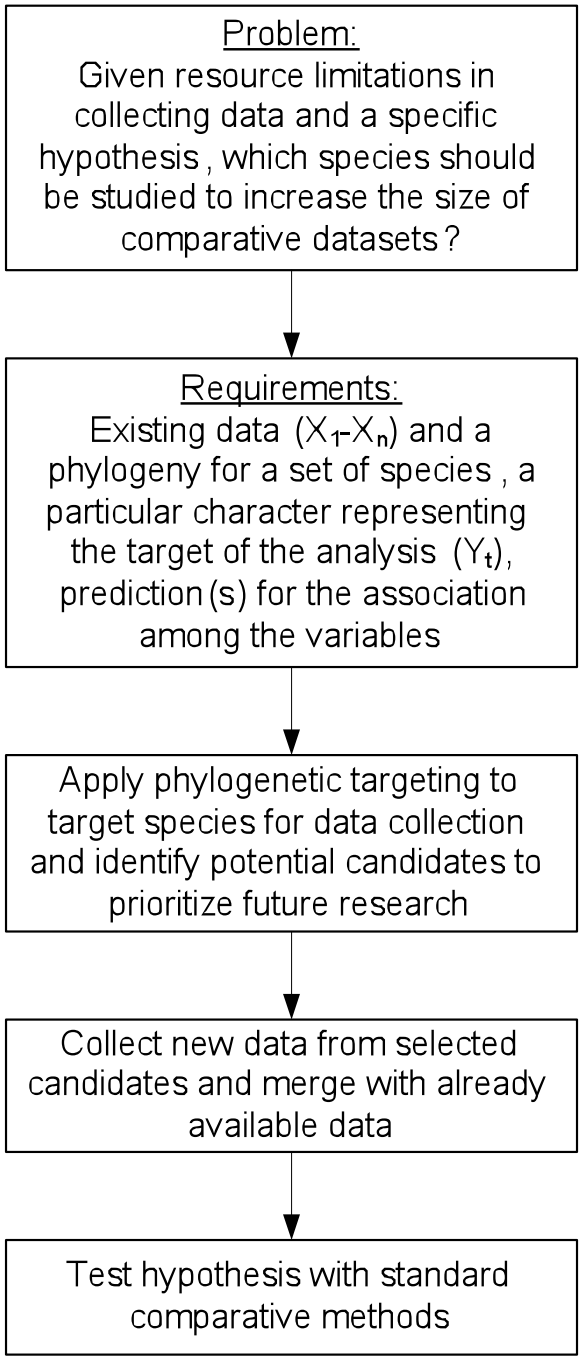

Problem:
Given resource limitations in collecting data and a specific hypothesis , which species should be studied to increase the size of comparative datasets ?
Requirements:
Existing data (X_1-X_n) and a phylogeny for a set of species , a particular character representing the target of the analysis (Y_t), prediction (s) for the association among the variables
Apply phylogenetic targeting to target species for data collection and identify potential candidates to prioritize future research
Collect new data from selected candidates and merge with already available data
Test hypothesis with standard comparative methods

Fig. 2

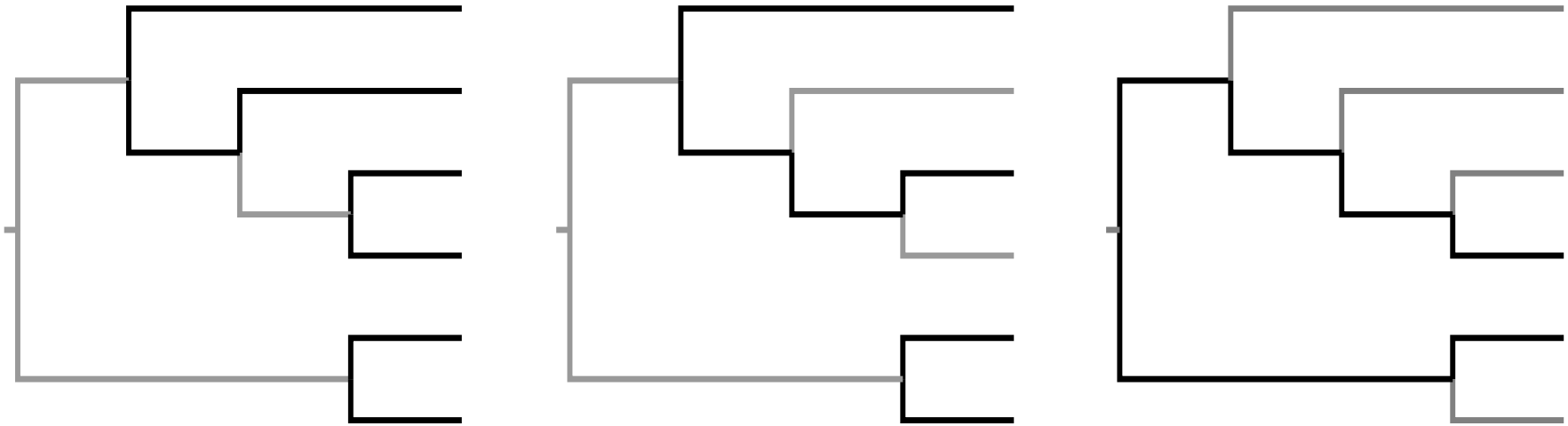
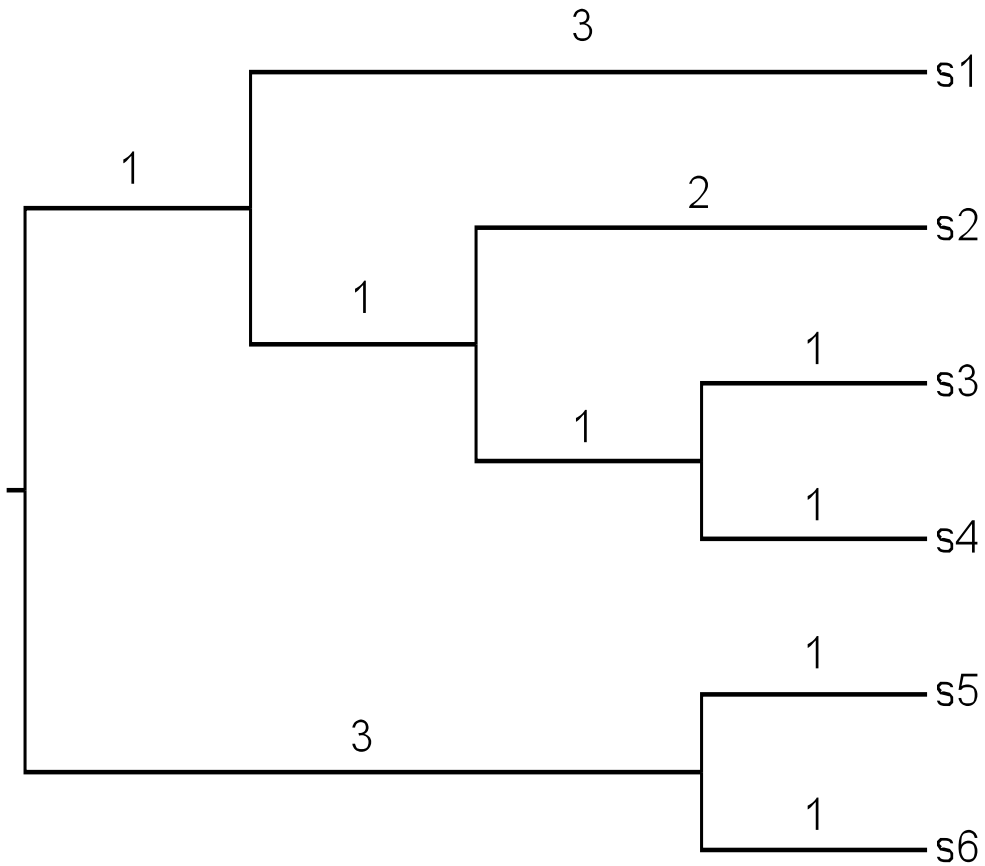


Fig. 3



	X_1	X_2	Y_t	B_t
s1	-0.4	10.5	5.5	1
s2	0.1	7.5	?	0
s3	0.4	9.0	?	0
s4	0.9	7.8	?	0
s5	0.6	25.3	6.9	1
s6	0.2	20.1	?	0

Fig. 4

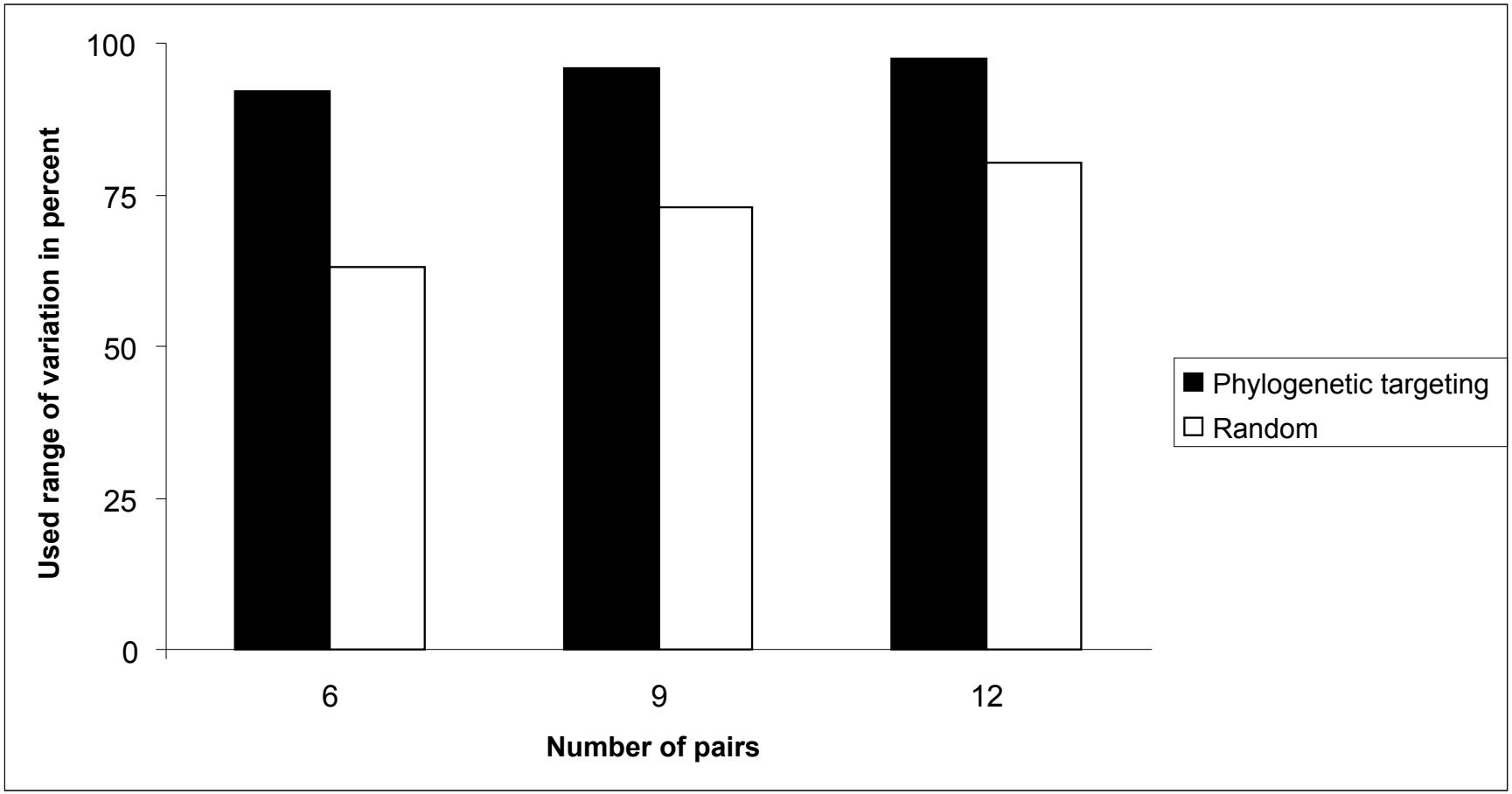

100
75
50
25
0
Used range of variation in percent
6
9
12
Number of pairs
Phylogenetic targeting
Random

Fig. 5

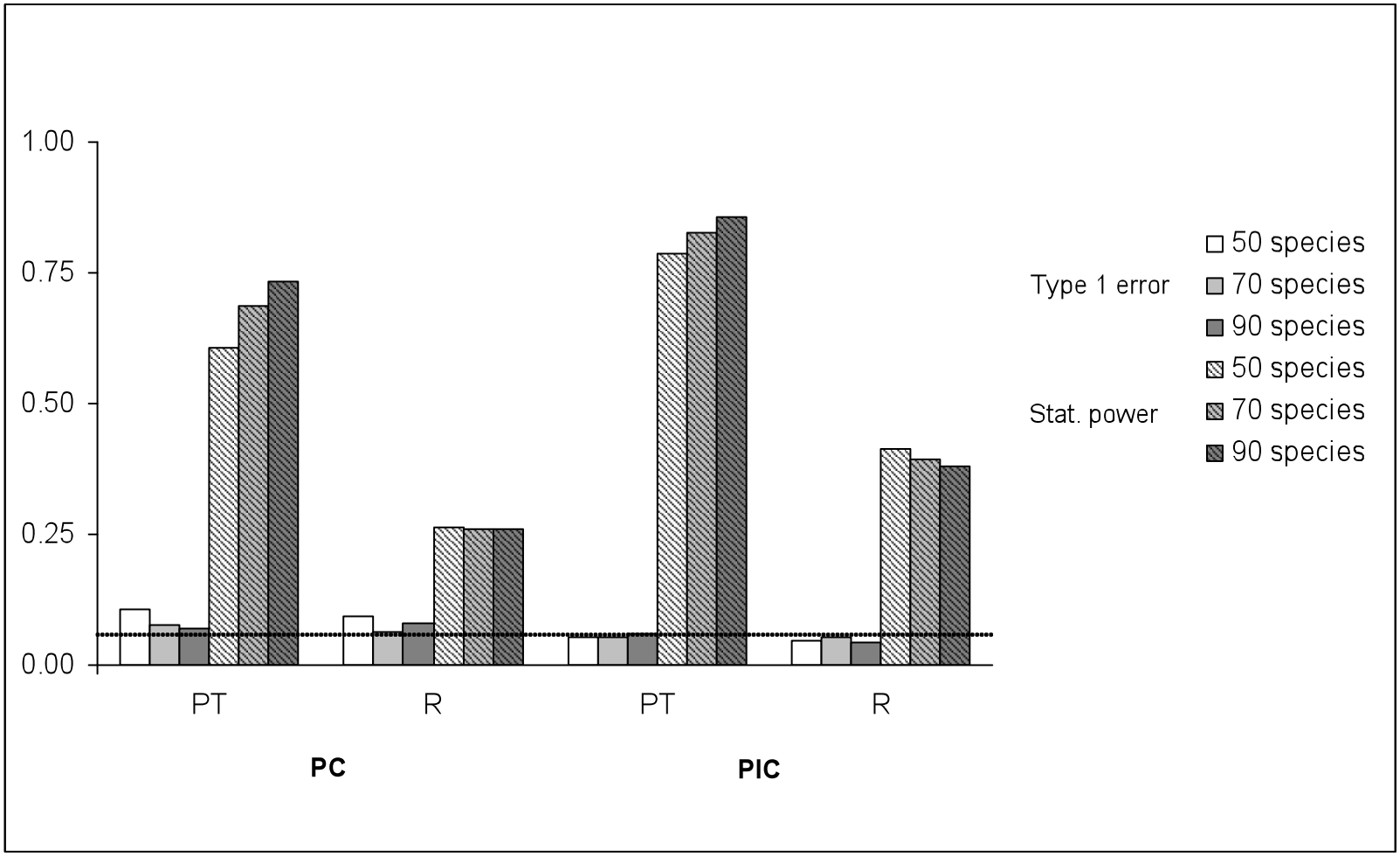

1.00
0.75
0.50
0.25
0.00
PT
R
PT
R
PC
PIC
Type 1 error
Stat. power
50 species
70 species
90 species
50 species
70 species
90 species

Fig. A1