doi:10.1111/evo.12520

Investigating evolutionary lag using the species-pairs evolutionary lag test (SPELT)

Charles L. Nunn[1,2] and Natalie Cooper[3,4]

[1]*Department of Evolutionary Anthropology, Duke University, Durham, North Carolina 27708*

[2]*E-mail: clnunn@duke.edu*

[3]*School of Natural Sciences, Trinity College Dublin, College Green, Dublin, Ireland*

[4]*Trinity Centre for Biodiversity Research, Trinity College Dublin, Dublin, Ireland*

Received July 12, 2013
Accepted August 9, 2014

For traits showing correlated evolution, one trait may evolve more slowly than the other, producing evolutionary lag. The species-pairs evolutionary lag test (SPELT) uses an independent contrasts based approach to detect evolutionary lag on a phylogeny. We investigated the statistical performance of SPELT in relation to degree of lag, sample size (species pairs), and strength of association between traits. We simulated trait evolution under two models: one in which trait *X* changes during speciation and the lagging trait *Y* catches up as a function of time since speciation; and another in which trait *X* evolves in a random walk and the lagging trait *Y* is a function of *X* at a previous time period. Type I error rates under "no lag" were close to the expected level of 5%, indicating that the method is not prone to false-positives. Simulation results suggest that reasonable statistical power (80%) is reached with around 140 species pairs, although the degree of lag and trait associations had additional influences on power. We applied the method to two datasets and discuss how estimation of a branch length scaling parameter (κ) can be used with SPELT to detect lag.

KEY WORDS: **Brownian motion, computer simulation, macroevolution, phylogeny, phylogenetic comparative methods.**

The comparative method plays a major role in evolutionary biology, with the history of life serving as a "natural experiment" to test hypotheses about adaptation (Ridley 1983; Harvey and Pagel 1991; Doughty 1996; Nunn 2011). Many comparative studies focus on investigating correlated trait evolution, using, for example, the method of independent contrasts (Felsenstein 1985; Garland et al. 1992). In such a study, the investigator tests for associations among two or more traits using a regression model that accounts for phylogenetic nonindependence. Another set of questions involves the evolutionary process (Pagel 1997, 1999; Blomberg et al. 2003; Butler and King 2004). Recently, for example, researchers have developed and applied scaling parameters that represent different models of trait change, such as an early burst model of evolution that corresponds to expectations of trait evolution under an adaptive radiation model (Blomberg et al. 2003; Cooper and Purvis 2010; Harmon et al. 2010). Similarly, researchers have developed methods to investigate rate variation (McPeek 1995; O'Meara et al. 2006; Thomas et al. 2006) and evidence for directional evolution (Pagel 1997, 1999).

Here, we consider another type of evolutionary pattern: evolutionary lag (Deaner and Nunn 1999). We define evolutionary lag as correlated evolution in two traits where one trait evolves more slowly than the other (this is sometimes referred to as phylogenetic inertia, but we avoid this term because it has many different interpretations, see Blomberg and Garland 2002; Losos 2008). Biological explanations for evolutionary lag include low genetic diversity, developmental constraints, or pleiotropic effects that exist in one trait but not others (Harvey and Pagel 1991; Blomberg and Garland 2002). In addition to basic interest in the evolutionary process, lag is important in light of anthropogenic changes that expose species to new selective pressures, possibly resulting in extinction. For example, a species may lack genetic variation related to immune defenses in the context of an emerging disease (e.g., facial tumor disease in Tasmanian devils, Siddle et al. 2007), or environmental change may occur more rapidly than evolutionary responses (e.g., the polar bear in a warming climate).

To investigate evolutionary lag, Deaner and Nunn (1999) developed an approach based on paired independent contrasts on

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

a dated phylogeny (see also Burt 1989; Giannini and Goloboff 2010). We call this method the species-pairs evolutionary lag test, or SPELT. Deaner and Nunn (1999) implemented SPELT using paired comparisons of extant species (Fig. 1). Imagine two traits, *X* and *Y*, where changes in *Y* lag behind changes in *X*. They calculated pairwise contrasts in *X* and *Y* for phylogenetically independent pairs of extant species on the phylogeny, then regressed pairwise contrasts in *Y* on pairwise contrasts in *X* through the origin. Pairwise contrasts in *X* were forced to be positive, with the direction of subtraction kept the same for the pairwise contrasts in *Y*; thus, the magnitude of change in *Y* will be a function of the degree of lag, change in *X*, and the amount of time available for *Y* to catch up to *X*. With this in mind, they regressed residuals from this model against the divergence times for the sister species pairs. If changes in *Y* lag behind changes in *X*, more recently diverged species pairs should have more negative residuals because contrasts in *Y* should be smaller than contrasts in *X*. Conversely, sister species that diverged longer ago should have positive residuals because more time has elapsed for changes in *Y* to take place. Thus, evolutionary lag is predicted to produce a positive association between the residuals and divergence time.

SPELT assumes that species pairs exhibit variation in branch lengths that encompass meaningful variation in lag for that system; that is, the tree should include some branches that are short enough for lag to be expressed, and others in which time was available for the lagging trait to evolve. SPELT also assumes that the phylogeny is known, including the timing of the split between the paired species. SPELT should have greater power to detect lag when change in *X* occurs immediately after the lineages split from a common ancestor, with *Y* then "catching up" to the change in *X* later. This situation corresponds to expectations under a speciational model of evolution with very low extinction rates, in which trait change in *X* is concentrated around speciation events (Harvey and Pagel 1991; Garland et al. 1993). Such a model may occur when speciation and trait evolution are associated with niche shifts (e.g., in adaptive radiations; Schluter and Nagel 1995), when founder effects are important for one trait, or in human cultural traits that indicate group identity, such as language (Atkinson et al. 2008).

It should be noted that phylogenetically independent pairwise contrasts are not the same as phylogenetically independent contrasts. We focus here on "cherries," which represent pairs of extant species with adjacent tips. Thus, the number of pairs depends on the shape of the tree (ranging from *N* species divided by 2 for a completely balanced tree, and 1 for a completely unbalanced tree), and is not equal to $N - 1$ as in phylogenetically independent contrasts. Using only these pairwise contrasts on an ultrametric tree has advantages in this context: only the date of one node is needed (rather than three dates for internal contrasts involving a pair of deeper nodes, Deaner and Nunn 1999), the lengths of

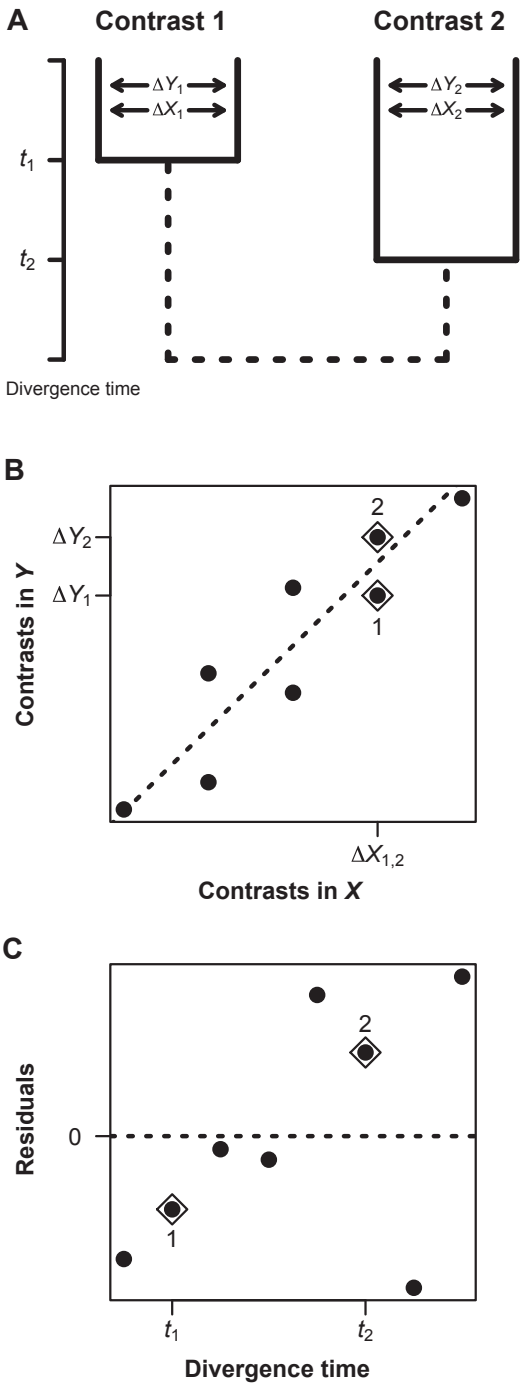

Figure 1. **Overview of SPELT. (A) The method examines pairwise comparisons (contrasts) involving extant taxa, that is, tips of the tree. The tree has branch lengths, *t*. (B) With a set of pairwise comparisons, the direction of subtraction is set so that *X* is positive. The same direction is used for *Y*. Contrasts are thus obtained, with *Y* regressed on *X* through the origin. Vertical residuals are recorded for the next step. In this example, comparing the two contrasts, change in Y_2 is greater relative to change in X_2, resulting in a positive residual, due to more time being available for Y_2 to catch up to the initial change in X_2 (i.e., $t_2 > t_1$). (C) Residuals are regressed on branch lengths, *t*. Positive effects are consistent with expectations under evolutionary lag.**

branches on either side of the node are equal, and actual measures of the traits are used rather than reconstructed values.

Several studies have applied SPELT to evolutionary questions, with mixed success. Deaner and Nunn (1999) applied the method to study primate brain size, but found no evidence that changes in brain size lag behind changes in body size. More recently, Isler et al. (2008) applied the method to investigate endocranial volume (ECV) in primates, which provided larger sample sizes. They also failed to find evidence for evolutionary lag. In a study of primate social systems, Lindenfors et al. (2004) found that changes in the number of males lags behind changes in the number of females. Finally, Thorén et al. (2006) failed to find evidence of lag in canine size (and canine dimorphism) relative to body mass (and body mass dimorphism) in primates.

Two important questions arise with regard to interpreting these findings. To what extent does SPELT have statistical power to detect evolutionary lag, especially given that only pairwise contrasts are used? And, is the method prone to false-positives when in fact lag did not occur? Respectively, these refer to Type II and Type I error rates. Understanding these error rates is crucial for interpreting previous results and for future applications of SPELT.

We address these questions using a simulation approach. By using two models of evolutionary lag, we aim to draw attention to situations under which SPELT is most likely to detect evolutionary lag, such as when most changes in X occur at speciation. Thus, one of our models simulates change in X under a speciational model (Garland et al. 1993). The other model simulates evolutionary change in trait X using a constant-variance or Brownian motion model, with trait Y as a function of the state of X at a previous time point.

Drawing on these findings, we also apply SPELT to reanalyze two previous studies of evolutionary lag in primates: one that failed to find evidence for lag (Isler et al. 2008) and another that found evidence for lag (Lindenfors et al. 2004). We reran these analyses to investigate how results change when using a more recent phylogeny (Arnold et al. 2010), to implement the procedure more systematically using an R script that we developed (available at https://github.com/nhcooper123/SPELT), and to investigate the use of a branch-length scaling method to assess the underlying assumptions of SPELT. For the latter, we obtained the maximum likelihood estimate of κ for each empirical example. This phylogenetic scaling parameter raises branch lengths to the value κ, resulting in equal branch lengths under a speciational model of evolution when $\kappa = 0$ (Pagel 1999, 2002). Thus, we expected that for the study that failed to detect evolutionary lag (Isler et al. 2008), ECV (trait Y) and body mass (trait X) would have similar values of κ, whereas for the study that did detect evolutionary lag (Lindenfors et al. 2004), the estimate of κ for the number of females (trait X) would be substantially lower than the estimate of κ for the number of males (trait Y), and thus more indicative of speciational change in the predictor variable.

Methods and Materials

To simulate evolutionary lag, we examined the correlated evolution of two traits on two descendent lineages from a common ancestor. We varied the number of paired sister species comparisons (N) and the degree to which Y lagged behind X (L). For one of the models, we also varied the degree of association between X and Y. All simulations were run in R version 2.15.1 (R Development Core Team 2009). For the two different models of evolutionary change, we ran 1000 simulations for every set of parameter combinations, which involved varying N from 20 to 200 (in multiples of 30), B from 0.5 to 3 (multiples of 0.5), L for Model 1 from 0 to 0.7 (multiples of 0.1), and L for Model 2 from 0 to 20 (multiples of 2).

MODEL 1: THRESHOLD MODEL OF LAG

This model assumed that changes occurred during speciation, and that the Y trait "caught up" to the X trait in a user-defined time period. To simulate trait evolution from a common ancestor of two sister species, s_1 and s_2, we drew a number from a uniform distribution on the interval [0,1] to reflect the branch length (t) leading from each species to their common ancestor. We then drew two random numbers from a normal distribution with a mean of 0 and variance equal to 0.03 (see below), where one number was assigned to s_1 and the other to s_2 (and thus served as the values of X for the two species). Under this model, the variance in trait change is not proportional to branch length, as expected under a speciational model of change (Fig. S1A).

We then determined the values for trait Y in s_1 and s_2 based on a user-defined slope, B, under the assumption that Y is a function of X, t, and L_1. If t was greater than the lag time (L_1), it was assumed that Y had caught up to change in X. Hence, Y was a linear function of X plus a small amount of error, ε, which was drawn from a normal distribution with variance of 0.03:

$$Y = BX + \varepsilon. \quad (1)$$

If t was less than L_1, then trait Y continued to lag behind trait X, and it was assumed to do so as a linear function of t. To obtain trait Y for the pair of species, we incorporated L_1:

$$Y_L = \frac{BXt}{L_1} + \varepsilon. \quad (2)$$

Thus, the effect of X on Y is weaker on shorter branch lengths (small t relative to L_1), as expected under evolutionary lag, and tending to generate a stronger relationship between t and Y (Fig. S1B). This procedure was repeated for each species pair,

producing N pairs of species with known branch lengths and trait values (and thus $2N$ species datapoints). Contrasts were calculated for each species pair; as expected, Y contrasts on branches less than L_1 tended to be smaller than those on branches greater than L_1 (Fig. S1C).

With these simulated data, we assessed evolutionary lag by regressing contrasts in Y on positivized contrasts in X through the origin, obtaining residuals, and then regressing those residuals on t (not through the origin, Fig. S1D). If the association between the residuals and t was significantly positive in a one-tailed test with $\alpha = 0.05$, we counted that simulation as showing evidence for evolutionary lag.

Note that the variance of the random number (0.03 in the simulations above) is arbitrary, and when the effect of X on Y is strong, a given level of variation is expected to have weaker effects on the statistical properties. For the simulations run here, we repeated analyses with variance set to 0.01 and 0.1. These simulations produced qualitatively similar results.

MODEL 2: INCREMENT-BASED LAG

In this model, correlated evolutionary change occurs in the two traits, but the lagged (and dependent) trait Y evolves relative to a previous value of X, determined by the integer L_2. Specifically, we assumed that $Y = B_{[t-L_2]} + \varepsilon$ when the branch length t was larger than L_2, and $Y = \varepsilon$ otherwise, where ε is normally distributed with a mean of 0 and SD of 1. Thus, Y is expected to lag behind X prior to L_2—specifically, change in Y will be random with respect to X on branches shorter than L_2—and for branches that are longer than L_2, Y is based on a previous value of X rather than the current value of X. As shown in Figure S2, trait X evolves under Brownian motion, with variance accumulating with time, and trait Y shows a similar pattern. This figure also reveals an association between X and t and evidence for lag when regressing residuals on t.

We implemented this model by generating a branch length as a random draw from a uniform distribution ranging from 0 to 1. We then rounded this number to two decimal places and multiplied it by 100, giving t units of evolutionary change along each branch ranging from 1 to 100. We then generated a vector of t normally distributed random numbers, also with a mean of zero and SD of 1, separately for two branches leading to two species, s_1 and s_2. The trait value for X was simply the sum of these t random numbers. For the lagging trait Y, the trait value was the sum of the first $t - L_2$ random numbers in the vector, plus statistical noise (ε) represented as a random normal deviate from zero with SD of 1. When $t < L_2$, we used the noise parameter ε as the estimate of Y.

To provide additional explanation of this procedure, consider the following example. Imagine we draw a value from a uniform distribution of 0.1217. To obtain the integer branch length, we round this to two decimal places, that is, 0.12, then multiplied this by 100 to get $t = 12$ units of evolutionary change. We then generate a vector of 12 normally distributed random numbers. The sum of these numbers is the trait value for X at the end of the branch, and the trait value for Y was the sum of the first $12 - L_2$ numbers plus ε. L_2 is an integer that varies from 0 to 20 in our simulations. For values of $L_2 \geq 12$, trait Y is equal to ε.

The underlying idea we seek to capture with Model 2 is that trait X evolves under a constant-variance Brownian motion model in which variance accumulates with time (see Fig. S2), and the value of Y is a function of a previous value of X when lag occurs (or the current value of X without lag, when $L_2 = 0$). When $t > L_2$, one can think of the previous value occurring on the branches leading to the sister species. When $t < L_2$, this time point occurred on the branch prior to the split of the two species, which results in the same trait value for both species (plus statistical error centered on zero). Because the Y traits are the same on the deeper branch, this shared value subtracts out to zero when calculating the Y contrast, and thus need not be estimated (while the error is included).

STATISTICAL ANALYSES

We examined the statistical output graphically and using general linear models that included key parameters as predictor variables (Model 1: N, B, L_2; Model 2: N, L_2). We compared two statistical models—one with no interactions, and another with all interactions—and selected among these based on differences in the Akaike information criterion (AIC), requiring at least two AIC units to favor one model over the other. All statistical analyses were conducted in R version 2.15.1 (R Development Core Team 2009).

R CODE AND EMPIRICAL TESTS

Our R package to implement SPELT is available at https://github.com/nhcooper123/SPELT, with instructions for downloading directly to R and code for simulations available in the online Supporting Information. We used this code to reanalyze two comparative studies of primates: Isler et al.'s (2008) study of lag in female ECVs relative to female body mass, and Lindenfors et al.'s (2004) study of lag in the number of males relative to the number of females. To run these analyses, we obtained a dated consensus tree from Version 3 of the *10kTrees* website (Arnold et al. 2010). After matching up the species in the data and trees, we had a dataset of 168 species for the analysis of ECV and 116 species for analysis of group composition. Data were log10-transformed prior to analysis. We estimated the parameter κ (Pagel 1999, 2002) using the caper package (Orme et al. 2011) in R (R Development Core Team 2009). Code for the simulations is available in the online Supporting Information.

Table 1. Statistical results from simulations of Model 1.

Variable	Estimate	SE	*t*-Statistic	*P*-Value
Intercept	0.04	0.07	0.53	0.59
Number of species pairs (N)	−0.05	0.03	−1.48	0.14
Slope (B)	−0.52	0.16	−3.24	0.001
Lag (L_1)	−2.24	0.45	−4.96	<0.0001
Interaction ($N \times B$)	0.47	0.07	6.41	<0.0001
Interaction ($N \times L_1$)	1.52	0.21	7.25	<0.0001
Interaction ($B \times L_1$)	8.04	1.03	7.84	<0.0001
Interaction ($N \times B \times L_1$)	−4.18	0.48	−8.8	<0.0001

Analyses were run using only tests of statistical power (462 simulations), but analyses of the full dataset produced similar results. To deal with zero values, we added one to each variable prior to running the $\log_{10}$ transformation. The AIC for a model with the interaction terms was substantially smaller (−1852) than a model without the interactions (−1742), and log-transformation helped meet the assumptions of the general linear model and considerably reduced the AIC. We assessed the effects of interactions using interaction plots that confirmed that increases in *N*, *B*, and L_1 increase statistical power once interactions are taken into account, despite the negative coefficients in this table (see Supporting Information).

Results

MODEL 1: THRESHOLD MODEL OF LAG

Running the simulations produced 336 different combinations of parameters, each of which was investigated using 1000 simulations. Of these combinations, 42 involved no lag, and thus provide an estimate of Type I error rates. In this subset of simulations, the mean Type I error rate was close to expected, at 5.2%, and the percentage of significant findings ranged from 3.8% (for $B = 1.5$ and $N = 50$) to 6.8% (for $B = 1.0$ and $N = 50$).

The remaining 294 simulations provide an estimate of statistical power under different strengths of association between X and Y (i.e., B), degree of lag (L_1), and number of species pairs (N). We found that all three variables positively affected the percentage of significant results using SPELT ($R^2 = 0.75$), with statistical power increasing as lag (L_1) increased, when the slope (B) relating X and Y increased (making the effects of random noise weaker), and when more pairs of species (N) were available for comparison. The model with interactions was favored, however, and the individual terms had negative coefficients due to significant interactions among the variables (Table 1, $R^2 = 0.87$), which indicate that values of one simulated parameter influence the effects of other parameters. Using interaction plots, we confirmed that power increased with N, B, and L_1 (see Figs. S3–S5). Figure 2 shows how statistical power varies with each of the variables independently. It appears that the method reaches a median 80% statistical power when $L_1 > 0.3$, $B > 1.5$, and $N > 110$. Even at the highest N of 200 species pairs, however, a wide range of statistical power outcomes was found, again highlighting the importance of other variables under this model of evolutionary lag and trait change.

MODEL 2: INCREMENT-BASED LAG

The parameter combinations used for Model 2 resulted in 77 simulation runs. Of these, seven involved $L_2 = 0$, that is, no lag. Across these seven runs, the mean Type I error rate was 4.8%, and thus close to the expected value of 5%. Type I error ranged from 3.7% (for 200 species pairs) to 6% (for 20 species pairs).

The remaining 70 simulations provide assessment of statistical power with varying degree of lag and sample size. The AIC for a model with the interaction term was slightly larger (−362.2) than a model without the interaction (−362.6). Because both models produced similar results, the results from the simpler model are presented here ($R^2 = 0.94$). Statistical power increased as lag and sample size increased (N: $b = 0.19$, $t_{67} = 29$, $P < 0.0001$; L_2: $b = 0.15$, $t_{67} = 18.6$, $P < 0.0001$). As with Model 1, we found that both degree of lag and sample size affected statistical power, yet power was nearly always greater than 80% when sample sizes exceeded 140 species pairs (Fig. 2D).

EXAMPLE 1. ECV IN PRIMATES

For the analysis of ECV in primates, SPELT returned 55 pairwise comparisons, with nodes ranging in age from 0.17 to 18.7 million years (mean of 3.25 million years). We found a strong positive association between pairwise contrasts in female primate body mass and pairwise contrasts in ECV (regression through the origin: $b = 0.45$, $t_{53} = 10.9$, $P < 0.0001$). When testing for evolutionary lag, residual ECV tended to increase with increasing branch length (Fig. 3A), but the result was not statistically significant ($b = 0.002$, $t_{52} = 1.29$, $P = 0.20$). When fitting κ, we found that estimates were similar for the two traits, and close to the expected value of 1 under Brownian motion ($\log_{10}$ female body mass: $\kappa_{ML} = 0.91$, 95% CI: 0.73–1.09; $\log_{10}$ ECV: $\kappa_{ML} = 1.14$, 95% CI: 0.95–1.32).

EXAMPLE 2. THE NUMBER OF MALES IN PRIMATE GROUPS

For the analysis of group composition in primates, the SPELT function returned 39 pairwise contrasts, with nodes ranging in age from 0.44 to 18.7 million years (mean of 4.9 million years). We found a strong positive association between pairwise contrasts in the number of males and pairwise contrasts in the number of females (regression through the origin: $b = 0.66$, $t_{38} = 6.06$, $P < 0.0001$). When testing for evolutionary lag, residual number of males tended to increase with increasing branch length (Fig. 3B), but the result was not statistically significant ($b = 0.0096$, $t_{37} = 1.05$, $P = 0.30$). When fitting κ, we found that estimates were similar for the two traits ($\log_{10}$ number of females:

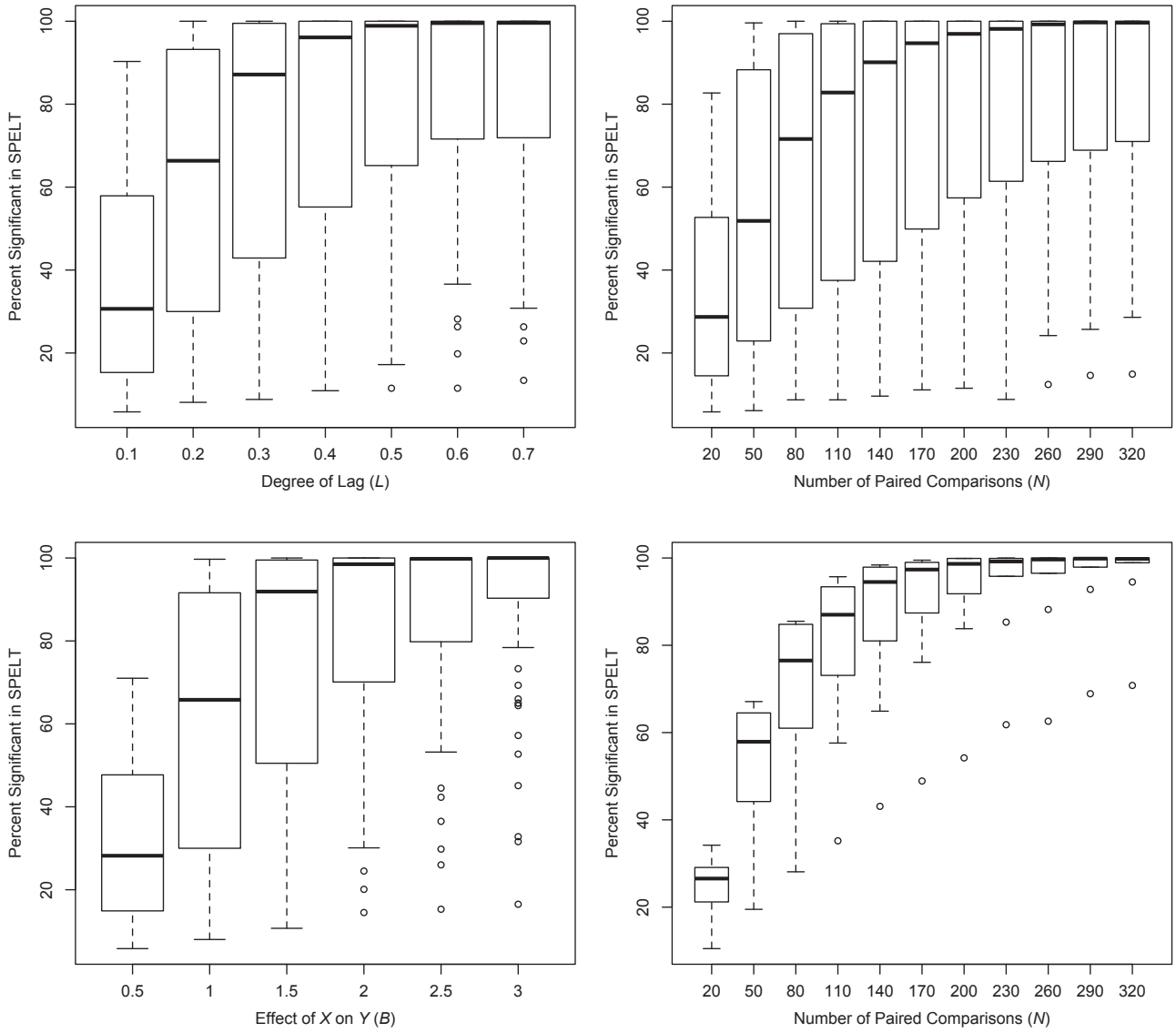

Figure 2. Statistical Power of SPELT. Under Model 1, the power to detect evolutionary lag increases with increasing lag (L_1, panel A), the effect of *X* on *Y* (*B*, panel B), and increasing sample size (*N*, panel C). Under Model 2, the power also increased with sample size (*N*, panel D) and degree of lag (L_2, see Fig. S5).

$\kappa_{ML} = 0.42$, 95% CI: 0.11–0.72; $\log_{10}$ number of males: $\kappa_{ML} = 0.39$, 95% CI: 0.10–0.67).

Discussion

Our simulation tests revealed that SPELT has reasonable Type I error rates that are very close to expected values of 5%. We also found that the power to detect evolutionary lag increased with sample size (number of species pairs) and the degree of lag that was simulated (equivalent to an effect size). The results suggest that reasonable power is obtained with approximately 140 species pairs, although the power depends further on the degree of lag and association among the traits. Given the large phylogenies and informatics databases that are now available, these sample sizes—that is, about 400 species—are possible for some well-studied groups. However, our simulations indicate that interpreting results from studies of fewer than 140 species pairs may be difficult. The association among the traits (*B*) is also easy to estimate (and included as output by the code we provide). It is harder to estimate the degree of lag in the context of the two models we used for assessing statistical power.

Previous studies of lag have generally failed to find significant results, with one exception: Lindenfors et al. (2004) found that the number of males lags behind the number of females in primates, which is consistent with the socioecological model

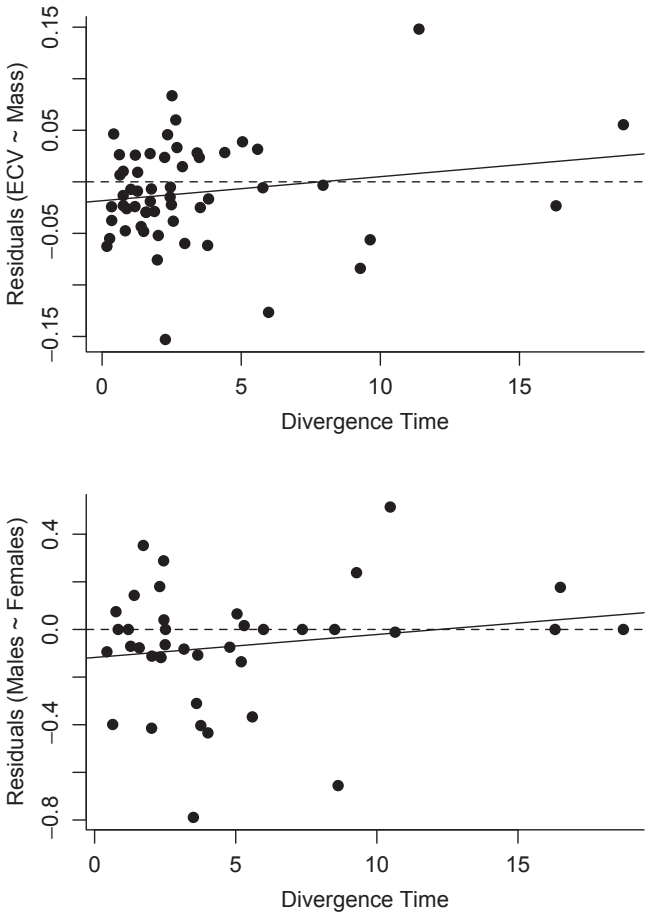

Figure 3. Application of SPELT to two empirical examples. Panel A shows results from retesting for lag in ECV relative to body mass (Isler et al. 2008), whereas panel B retested for lag in the number of males relative to the number of females (Lindenfors et al. 2004). Although slopes were in the predicted positive direction in both cases, in neither situation was the test significant.

that females map onto resources, and males then map onto females (Nunn 1999). Although we confirmed the lack of significant lag effects in one previous study (Isler et al. 2008), we failed to confirm the presence of evolutionary lag in male group size in primates. This inconsistency in results could be due to several differences, including an updated phylogeny, the use of a more systematic method for identifying pairwise contrasts via the SPELT R package, and the low power expected given the sample size involved in this test (39 contrasts; Lindenfors et al. 2004 had 37). In general, however, it appears that few traits exhibit lag on a scale that is detectable on phylogenies. Even if rare, it is valuable to have methods, such as SPELT, to test hypotheses involving evolutionary lag. The R package developed here makes implementation of SPELT substantially easier and repeatable.

Several recent developments in comparative biology deserve mention in relation to implementing SPELT. First, evolutionary lag shares similarities with phylogenetic niche conservatism (PNC), defined as the tendency of species to retain characteristics of their fundamental niche over time (Wiens and Graham 2005). In some of the macroevolutionary models underlying PNC, we also see one trait lagging behind the other, for example, if PNC is identified by the conserved trait having a lower rate of evolution than another trait (e.g., Cooper et al. 2010), the result will be consistent with evolutionary lag as defined here. Other methods for detecting PNC (or "phylogenetic inertia" sensu Hansen et al. 2008) are often based on the Ornstein–Uhlenbeck (OU) model of stabilizing selection (Felsenstein 1988; Hansen et al. 2008; Lavin et al. 2008). The OU model is a modification of the Brownian model, where trait values evolve toward an optimal phenotype, which itself evolves by Brownian motion (Lande 1976; Felsenstein 1988; Hansen 1997). The OU model incorporates a parameter that explicitly measures how the rate of evolution of the focal trait to the optimum is lagged relative to the change in trait optimum (Hansen et al. 2008). Traits with high values of this parameter may be consistent with evolutionary lag as defined here. However, the main difference between these methods and evolutionary lag is that lag makes an explicit assumption that one trait lags behind the other via an unspecified mechanism. In the OU model, the mechanism underlying lag is assumed to be "phylogenetic inertia" (Hansen et al. 2008; Labra et al. 2009), but the lag is between a trait and its optimum, rather than between two different traits.

Second, new methods to control for phylogenetic uncertainty could be used in applying SPELT. Specifically, we assumed that the true-dated phylogeny is known, yet phylogenetic topology and branch lengths are never known with certainty. Bayesian methods could be used to investigate key parameters—such as the association between branch lengths and residuals—to produce a Bayesian posterior probability distribution of this coefficient (Pagel and Lutzoni 2002). One could then assess the degree of support for evolutionary lag based on, for example, the proportion of regression coefficients that are greater than zero or whether the 95% credible interval includes zero. It is also possible to investigate whether particular trees from a Bayesian posterior distribution provide stronger support for evolutionary lag, which may provide insights into the reasons for those differences.

With our demonstration that SPELT has appropriate statistical performance, an important step for the future involves applying the method to additional traits in which lag has been hypothesized to occur. Although we failed to find evidence for lag in the example data we investigated, the code that we provide to implement SPELT will help in those endeavors, especially now that the statistical performance of the method has been documented. We expect that evolutionary lag should be most detectable for traits that covary with traits showing a speciational evolutionary pattern, including those involved in the speciation process. Thus, estimation

of phylogenetic scaling parameters should also play an important role in future analyses using this method.

ACKNOWLEDGMENTS

We thank R. Griffin for help with the figures and references, and two anonymous referees for helpful suggestions. This project was initiated during the AnthroTree Workshop, which is supported by the National Science Foundation (BCS-0923791) and the National Evolutionary Synthesis Center (NSF grant EF-0905606). NC is supported by the European Commission CORDIS Seventh Framework Program (FP7) Marie Curie CIG grant, proposal number: 321696.

DATA ARCHIVING

The doi for our data is 10.5061/dryad.qb6h8.

LITERATRE CITED

Arnold, C., L. J. Matthews, and C. L. Nunn. 2010. The *10kTrees Website*: a new online resource for primate phylogeny. Evol. Anthropol. 19:114–118.

Atkinson, Q. D., A. Meade, C. Venditti, S. J. Greenhill, and M. Pagel. 2008. Languages evolve in punctuational bursts. Science 319:588.

Blomberg, S. P., and T. Garland. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. J. Evol. Biol. 15:899–910.

Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745.

Burt, A. 1989. Comparative methods using phylogenetically independent contrasts. Pp. 33–53 *in* P. H. Harvey and L. Partridge, eds. Oxford surveys in evolutionary biology. Oxford Univ. Press, Oxford, U.K.

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164:683–695.

Cooper, N., and A. Purvis. 2010. Body size evolution in mammals: complexity in tempo and mode. Am. Nat. 175:727–738.

Cooper, N., W. Jetz, and R. Freckleton. 2010. Phylogenetic comparative approaches for studying niche conservatism. J. Evol. Biol. 23:2529–2539.

Deaner, R. O., and C. L. Nunn. 1999. How quickly do brains catch up with bodies? A comparative method for detecting evolutionary lag. Proc. R. Soc. Lond. B Biol. Sci. 266:687–694.

Doughty, P. 1996. Statistical analysis of natural experiments in evolutionary biology: comments on recent criticisms of the use of comparative methods to study adaptation. Am. Nat. 148:943–956.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

———. 1988. Phylogenies and quantitative characters. Ann. Rev. Ecol. Syst. 19:445–471.

Garland, T., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Syst. Biol. 4:18–32.

Garland, T., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Syst. Biol. 42:265–292.

Giannini, N. P., and P. A. Goloboff. 2010. Delayed-response phylogenetic correlation: an optimization-based method to test covariation of continuous characters. Evolution 64:1885–1898.

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351.

Hansen, T., J. Pienaar, and S. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62:1965–1977.

Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. Bryan Jennings, K. H. Kozak, M. A. McPeek, F. Moreno Roark, and T. J. Near. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385–2396.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford Univ. Press, Oxford, U.K.

Isler, K., E. Christopher Kirk, J. Miller, G. A. Albrecht, B. R. Gelvin, and R. D. Martin. 2008. Endocranial volumes of primate species: scaling analyses using a comprehensive and reliable data set. J. Hum. Evol. 55:967–978.

Labra, A., J. Pienaar, and T. Hansen. 2009. Evolution of thermal physiology in liolaemus lizards: adaptation, phylogenetic inertia, and niche tracking. Am. Nat. 174:204–220.

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334.

Lavin, S., W. Karasov, A. Ives, K. Middleton, and T. Garland Jr. 2008. Morphometrics of the avian small intestine compared with that of nonflying mammals: a phylogenetic approach. Physiol. Biochem. Zool. 81:526–550.

Lindenfors, P., L. Froberg, and C. L. Nunn. 2004. Females drive primate social evolution. Proc. R. Soc. Lond. B 271:S101–S103.

Losos, J. 2008. Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. Ecol. Lett. 11:995–1003.

McPeek, M. A. 1995. Testing hypotheses about evolutionary change on single branches of a phylogeny using evolutionary contrasts. Am. Nat. 145:686–703.

Nunn, C. L. 1999. The number of males in primate social groups: a comparative test of the socioecological model. Behav. Ecol. Sociobiol. 46:1–13.

———. 2011. The comparative approach in evolutionary anthropology and biology. Univ. of Chicago Press, Chicago, IL.

O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922–933.

Orme, D., R. Freckleton, G. Thomas, T. Petzoldt, S. Fritz, and N. Isaac. 2011. Caper: comparative analyses of phylogenetics and evolution in R. Available at http://R-Forge.R-project.org/projects/caper/.

Pagel, M. 1997. Inferring evolutionary processes from phylogenies. Zool. Scr. 26:331–348.

———. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

———. 2002. Modelling the evolution of continuously varying characters on phylogenetic trees: the case of hominid cranial capacity. Pp. 269–286 *in* N. MacLeod and P. L. Forey, eds. Morphology, shape and phylogeny. Taylor and Francis, Lond.

Pagel, M., and F. Lutzoni. 2002. Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation. Pp. 148–161 *in* M. Lässig and A. Valleriani, eds. Biological evolution and statistical physics. Springer-Verlag, Berlin.

R Development Core Team. 2009. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ridley, M. 1983. The explanation of organic diversity: the comparative method and adaptations of mating. Clarendon, Oxford, U.K.

Schluter, D., and L. M. Nagel. 1995. Parallel speciation by natural selection. Am. Nat. 146:292–301.

Siddle, H. V., A. Kreiss, M. D. B. Eldridge, E. Noonan, C. J. Clarke, S. Pyecroft, G. M. Woods, and K. Belov. 2007. Transmission of a fatal clonal tumor by biting occurs due to depleted MHC diversity in a threatened carnivorous marsupial. Proc. Natl. Acad. Sci. 104:16221–16226.

Thomas, G. H., R. P. Freckleton, and T. Székely. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proc. R. Soc. B Biol. Sci. 273:1619–1624.

Thorén, S., P. Lindenfors, and P. M. Kappeler. 2006. Phylogenetic analyses of dimorphism in primates: evidence for stronger selection on canine size than on body size. Am. J. Phys. Anthropol. 130:50–59.

Wiens, J. J., and C. H. Graham. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. Ann. Rev. Ecol. Evol. Syst. 36:519–539.

Associate Editor: E. Abouheif
Handling Editor: J. Conner

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Example output from Model 1 (threshold model with speciational change in X).

Figure S2. Example output from Model 2 (increment-based lag with Brownian change in X).

Figure S3. Shows how the value of the slope (B) influences the relationship between number of species pairs (N) and statistical power.

Figure S4. Shows how the degree of lag (L_1) influences the association between the number of species pairs and statistical power.

Figure S5. Shows how the value of the slope (B) influences the relationship between lag (L_1) and statistical power.