DISENTANGLING EVOLUTIONARY CAUSE-EFFECT RELATIONSHIPS WITH PHYLOGENETIC CONFIRMATORY PATH ANALYSIS

AchazvonHardenberg 1,2,3 and Alejandro Gonzalez-Voyer 3,4

1 Alpine Wildlife Research Centre, Gran Paradiso National Park, Degioz 11, 11010 Valsavarenche (Aosta), Italy
2 E-mail: achaz.hardenberg@pngp.it
3 Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC), Av. Américo Vespucio 1, 41092 Sevilla, Spain

Confirmatory path analysis is a statistical technique to build models of causal hypotheses among variables and test if the data conform with the causal model. However, classical path analysis techniques ignore the nonindependence of observations due to phylogenetic relatedness among species, possibly leading to spurious results. Here, we present a simple method to perform phylogenetic confirmatory path analysis (PPA). We analyzed simulated datasets with varying amounts of phylogenetic signal in the data and known underlying causal structure linking the traits to estimate Type I error and power. Results show that Type I error for PPA appeared to be slightly anticonservative (range: 0.047–0.072) but path analysis models ignoring phylogenetic signal resulted in much higher Type I error rates, which were positively related to the amount of phylogenetic signal (range: 0.051 for $\lambda = 0.0916$ for $\lambda = 1$). Further, the power of the test was not compromised when accounting for phylogeny. As an example of the application of PPA, we revisited a study on the correlates of aggressive broodmate competition across seven avian families. The use of PPA allowed us to gain greater insight into the plausible causal paths linking species traits to aggressive broodmate competition.

**KEYWORDS:** Broodmate aggression, confirmatory path analysis, d-sep test, generalized least squares, phylogenetic comparative analyses.

The phylogenetic comparative method has become widely used to address questions related to long-term evolutionary processes by analyzing datasets collected across multiple species and incorporating information about the varying degrees of relatedness among them (Felsenstein 1985; Harvey and Pagel 1991; Freckleton et al. 2002). Such comparative analyses often include numerous variables, which may be directly or indirectly related to the trait of interest, yielding a complex, multivariate network of associations, in which the distinct variables may present different effect sizes. Evolutionary biologists employing the comparative method have come to accept, with some resignation, that one inevitable consequence of the use of such methods is that they must banish the idea of causality altogether (although one particular method does allow to determine contingency, see Pagel and Meade 2006). The results are generally interpreted as allowing—albeit to identify a subset of variables that are weakly correlated in fashion with differences in the trait of interest between two groups of species. Indeed, the fully randomized experiment.

3 These authors contributed equally to this work.
ideal means by which to test hypotheses and explore causal relationships among variables (Fisher 1926). However, many evolutionary questions regarding causality are simply impossible to address using fully randomized experiments and alternative methods that have not been adopted (Felsenstein 1985; Harvey and Pagel 1991; Martins 2000; Freckleton et al. 2002).

One of these methods, confirmatory path analysis, has been specifically developed to test prespecified causal hypotheses represented as directed acyclic graphs (DAGs) and thus as a set of structural equations (Shipley 2000b). Basically, path analysis posits that correlational relationships between characters imply an unresolved causal structure, because the causal processes generating the data impose constraints on the patterns of correlation such that such display (Shipley 2000a). Standard path analysis methods, such as those implemented in structural equation models (SEM), therefore compare the observed covariance matrix with the covariance matrix predicted by the tested causal model. Alternatively, the d-sep test, developed by Shipley (2000b), tests the conditional probabilistic independencies implied in the DAG of the hypothesized causal model. As has been well discussed in the literature, however, data points in multispecies analyses cannot be considered as independent from a statistical point of view because the differing degrees of shared ancestry among species will influence the expected similarity in trait values (Felsenstein 1985; Harvey and Pagel 1991; Garland et al. 1992; Freckleton et al. 2002). The consequences of not accounting for phylogenetic effects in statistical analyses of multispecies data are, among others, artificially inflated numbers of degrees of freedom, incorrectly estimated variances, and increased Type I error rates of significance tests (Felsenstein 1985; Harvey and Pagel 1991; Martins and Garland 1991; Martins et al. 2002; Rohlf 2006). All these problems can be compounded in path analysis because of the requirement of testing multiple structural equations (in the case of SEM) or all the conditional probabilistic independencies that must be true for the causal model to be correct (in the case of the d-sep test). Path analysis models addressing evolutionary questions using multispecies data, but which ignore the underlying phylogenetic relationships among species, may therefore fail to detect the “true” causal structure between the variables. Attempts to use path analysis on multispecies datasets have been previously reported in the literature. However, most of these analyses failed to account explicitly for phylogeny (Soleti et al. 2005, 2010) or did not specify the methods used to account for phylogeny. Recently, Santos and Cannatella (2011) used phylogenetic independent contrasts (Felsenstein 1985; Garland et al. 1992) as the data entered into SEM. This approach allowed them to account for the phylogenetic nonindependence of the data arising from phylogenetic relatedness. However, independent contrasts assume that the data being analyzed evolves following a strict Brownian motion model of evolution and performance can be compromised if the assumption is not met (Revell 2010); furthermore, independent contrasts also assume strictly linear relationships between trait values (Quade et al. 2004). Here, we propose an alternative approach combining path analysis with phylogenetic generalized least squares (PGLS) methods (Martins and Hansen 1997). The advantage of PGLS is that it can incorporate distinct models of trait evolution, can combine continuous and categorical variables in a single model without the need to code dummy variables, and provides the value of the y-intercept (Martins and Hansen 1997). Further, a key advantage of using PGLS is that it would allow for path analyses to be undertaken using taxon-specific trait values rather than contrasts, facilitating interpretation of the results. Finally, in PGLS an evolutionary parameter is estimated simultaneously with the model fit. Therefore, this parameter is determined by the amount of phylogenetic signal in the residual error (the residual of the model to be precise) and hence the necessary correction for the expected covariance in trait values resulting from phylogenetic relatedness, given the evolutionary model (Freckleton et al. 2002; Revell 2010). This is an important advantage because in some instances data may present a phylogenetic structure that is intermediate between that predicted by the evolutionary model and absence of phylogenetic correlation in the data (Freckleton et al. 2002; Revell 2010). Under such circumstances, PGLS models have been shown to outperform independent contrasts (Revell 2010).

Our proposed method for phylogenetic confirmatory path analysis (hereafter called PPA), integrates PGLS with the d-sep test developed by Shipley (2000b). This method exploits the concept of conditional independence (Pearl 1988; Verma and Pearl 1988) to predict the minimal set of conditional probabilistic independencies that must be true if the causal model is correct. The predicted independencies can thus be tested using various statistical tests, according to the nature of the data at hand, and the probabilities of these tests can be combined using Fisher’s C-test (Shipley 2000a), which reflects the deviation of the data from the causal model to be correct. The d-sep test is very general in that it can be used for small sample sizes (because the inferential tests are not asymptotic), nonnormally distributed data (although the phylogenetic comparative methods will use the assumption of normal distribution of the phylogenetically transformed residuals), and nonlinear functional relationships. The only disadvantage of the d-sep test is that it cannot be used with causal models including latent (i.e., not measured) variables (Shipley 2000a, b). Shipley (2009) showed how confirmatory path analysis by d-sep tests can be generalized to deal with data having an underlying hierarchical or multilevel structure. Here, we generalize the method further to deal with multispecies data, which are not independent because of phylogenetic relationships among species. We use simulation to explore the consequences of
Methods
INTEGRATING THED-SEPTEST WITH PGLS
Shipley (2009) showed how the d-septest can be combined with generalized linear mixed models (GLMM) and provides detailed instructions to do this within the open source statistical environment R (R Development Core Team 2011) using the package “nlme” (Pinheiro et al. 2011). We bring this idea one step further, showing that the same procedure as in Shipley (2009) can be used to combine the d-septest with PGLS, and thus perform PPA. Although the method was already described in detail elsewhere, for didactic reasons, we present here the four steps involved in the d-sept test for confirmatory path analysis, with additional details about how to combine it with PGLS (for a more detailed account on the procedure for non-phylogenetic path analysis and on the statistical background, we refer readers to Shipley 2000b, 2009). The first step in any path analysis (phylogenetic or not) is to describe the hypothesized causal relationships among the measured variables using a DAG. Typically, in a DAG, measured variables are represented as boxes (called vertices in the jargon of graph theory) and causal links are represented as directed arrows (called edges) joining the vertices. A vertex from which an edge originates is called parent. Figure 1 shows an example of DAGs describing two alternative models of possible cause–effect relationships among five variables. The second step consists in using the concept of d-separation (Pearl 1988; Verma and Pearl 1988) to predict the minimal set of conditional probabilistic independence constraints (called the basis set), which must all be true for the causal model to be correct. In practice, to obtain the basis set, one has to list all pairs of non-adjacent variables, that is, those not directly joined by an edge.

Thus, for the model in Figure 1A the list would be \{X_1, X_3, \{X_1, X_4\}, \{X_1, X_5\}, \{X_2, X_3\}, \{X_2, X_4\}, \{X_3, X_5\}\}. Then, eliminate the parent variables of either non-adjacent variables in the previous list, that is \{X_2, X_3\}. The final list is \{X_2, X_3\}. Therefore, the basis set of the d-separation statement describing the probabilistic independence between the two non-adjacent variables, conditioned on the parent variables of both; that is, for the model in Figure 1A, the basis set would be \{X_1, X_3\}, \{X_2, X_3\}, \{X_1, X_4\}, \{X_1, X_5\}, \{X_2, X_3\}.

In the case of data with an underlying phylogenetic structure, such linear models can be easily fit using the PGLS approach implemented in R using the package “nlme,” already used by Shipley (2009) in the context of GLMM, and the package ape (Paradis et al. 2004). More specifically, the above conditional independence statement (and all the others in the basis set) can be analyzed using generalized least squares models where the correlation structure of the data is given by the expected covariance of speciestraits given the phylogenetic tree and evolutionary model (for details on the code and function of the analyses see Paradis 2006). The last step consists in testing whether the predicted basis set of conditional independencies is fulfilled in the observational data. This is done by combining all the values of \rho (i.e., the probabilities that the non-adjacent variables in the basis set are statistically independent conditional on their parent variables) using Fisher’s C statistic

$$C = -2 \sum_{i=1}^{k} \ln(p_i)$$

where \(k\) is the number of independent tests in the basis set. When the model is correct, the C statistic follows a \(\chi^2\) distribution with \(2k\) degrees of freedom. The path model is thus considered to fit the data when the C statistic is not significant (\(P > 0.05\)) (Shipley 2000a, 2004).

Unfortunately, different causal models can fit the same data and, therefore, some form of model selection procedure is needed to identify the best fitting, and thus most likely, causal model among these accepted path models. Shipley (2000a) proposed

---

**Figure 1.** Directed acyclic graphs describing two alternative models of possible cause–effect relationships among five variables.
an approach based on testing the difference in Fisher's $C$ statistics of two competing nested models, which follows a $\chi^2$ distribution with $\Delta df = df_{model1} - df_{model2}$. The basis model is rejected in favor of the nested model when the probability associated with $C$ is lower than the chosen significance level ($\alpha = 0.05$). This approach, however, can be used only when comparing truly nested models, that is, when the parameters fixed to 0 in the first model are a subset of the fixed parameters in the second model. An appealing alternative, which can also be used for selecting among non-nested models (provided the dataset is always the same for all models in the set), we propose to use the Information Theory approach recently applied, in the setting of a non-phylogenetic path analysis, by Cardon et al. (2011). An information criterion modified for small sample sizes and adapted to path analysis (C statistic Information Criterion [CICc]) can be calculated as follows (Cardon et al. 2011):

$$CICc = C + 2q \times \frac{n}{(n - q - 1)}$$

where $C$ is Fisher's $C$ statistic, $n$ is the sample size, and $q$ is the number of parameters that is given by the total number of variables used to build the models (a constant within the same set of models we are comparing), plus the number of edges linking them (which can change for each model compared). Model selection, as well as subsequent model averaging, can thus follow standard information theory procedures, whose detailed description is outside the scope of the present article (for excellent accounts on these procedures, we refer the reader to Burnham and Anderson 1998 and Grueber et al. 2011). Although Cardon et al. (2011) call this information criterion $CICc$, we prefer to call it $CICc$ to avoid confusion with the original Akaike Information Criterion that is based on the maximum likelihood of the data rather than on the $C$ statistic of the $d$-sep test. However, while this approach has been previously used in the context of confirmatory path analysis with the d-sepmethod (Cardon et al. 2011), the proposed $CICc$ statistic is still lacking formal proof. It should therefore be used with caution, until further studies confirm its validity.

SIMULATIONS

We used a simulation-based approach to investigate the consequence of ignoring phylogenetic relatedness when undertaking path analysis using the d-sepmethod (Shipley 2000b). We simulated evolution of five hypothetical traits using a prespecified covariance matrix among the traits determining a specific trait model (the same model depicted in Figure 1A and used as an example in the previous section). Simulations were run under six different scenarios spanning a continuum from all to strongly phylogenetic signal in the simulated data; in other words, from traits evolving along a star phylogeny, where trait evolution for each species is independent, to traits evolving following a Brownian motion model, where the degree of similarity between species traits is inversely proportional to the distance to the nearest common ancestor. For the scenario of strong phylogenetic signal, traits were simulated to evolve on an ultrametric phylogeny under a Brownian motion model. For the five remaining scenarios, we used the parameter lambda ($\lambda$) (Freckleton et al. 2002) to transform the phylogenetic tree prior to trait evolution. The $\lambda$ parameter is taken as equal to 1 and 0, which gives high values of $\lambda$ (i.e., 0.8, 0.6, 0.4, 0.2, and 0) prior to simulating trait evolution and tests of conditional dependencies on the transformed tree. For each of the six scenarios we simulated 1000 datasets, each with an underlying phylogenetic tree of fixed, arbitrary size of 100 species. Each simulation of trait evolution was done using a different simulated phylogeny; hence, our simulations also incorporated the effects of varying phylogenetic topology. At each iteration, we calculated Fisher's $C$ statistic and obtained a distribution of $P$-values determined by the Type I error (i.e., the probability of rejecting the null hypothesis, in this case the tested model, when it is true, testing the predicted set of conditional independence consistent with the true underlying causal model depicted in Fig. 1A) and the power (i.e., 1 — the Type I error, the probability of not rejecting the tested model when it is actually false, testing the predicted set of conditional independencies of a “correct” causal model depicted in Fig. 1B). These simulations were run both for d-sep tests ignoring phylogenetic effects and for the phylogenetically explicit d-sep test. All simulations and analyses were done in R (R Development Core Team 2011) using the packages “ape” (Paradis et al. 2004), “nlme” (Pinheiro et al. 2011), and “geiger” (Harmon et al. 2008). Scripts used for the simulations are provided as Supporting information.

AVIAN BROODMATE COMPETITION DATA

As an empirical example of PPA, we revisit the question of which factors favor the evolution of aggressive sibling competition in birds (see Gonzalez-Voyer et al. 2007). In their study, Gonzalez-Voyer et al. (2007) analyzed the correlation between five behavioral and life-history traits—feeding method, feeding rate, clutch size, egg size, and length of the nestling period—and two measures of aggressive competition: incidence and intensity. Incidence of aggression was the percentage of broods in which aggression was reported and was measured on a 4-point scale. Intensity of aggression was scored on a 4-point scale by five judges independently, on the basis of qualitative and quantitative descriptions of broodmate aggression in the primary literature, and the median was used as the score (Gonzalez-Voyer et al. 2007). Feeding method was a continuous variable expressed as the
had been competition between the parent hick (i.e., feeding to the old v lesting period) had been highly and significantly correlated (Gonzalez-Voyer et al. 2007). Egg size was used as a proxy for nesting body size at hatching with which it was highly and significantly correlated (Gonzalez-Voyer et al. 2007). Finally, average length of the nesting period was the number of days separating hatching from fledging and was log transformed (for further details on variables see Gonzalez-Voyer et al. 2007). Because length of the nestling period and egg size were significantly correlated, in the original analyses egg size was omitted to avoid problems of multicollinearity, however, the authors did find that when replacing length of the nestling period by egg size, the later was not significantly correlated with either measure of aggressive competition (Gonzalez-Voyer et al. 2007), suggesting there is no direct association between egg size and aggressive competition. In the original study, Gonzalez-Voyer et al. (2007) included 69 species from seven different bird families, however, data on egg size was not available for one species (Haliaeetus vociferoides) so the dataset analyzed here includes 68 species. Because Gonzalez-Voyer et al. (2007) did not find any significant relationship between feeding rate and either measure of aggressive competition and data were not available for 27 species we did not include this trait in the phylogenetic path analyses. For our results to be comparable with the original study, the analyses were done using the dataset by Gonzalez-Voyer et al. (2007). The dataset used for this study is available online as Supporting information.

Due to the methodological limitations of the time, questions remained unanswered. For instance, although the PGLS analyses suggested there was no direct association between egg size and aggressive competition, egg size could influence aggression through its effect on clutch size and length of the nestling period. Egg size, clutch size, and length of the nestling period are known to be associated with lifetime-history trade-offs between offspring number and offspring size (see Bennett and Owens 2002). Second, the authors found a significant negative correlation between clutch size and intensity of aggression but the directionality of the relationship was unresolved: smaller clutches could favor the evolution of aggressive strife (Drummond 2002); alternatively, smaller clutches could be favored because species in which aggressive competition has evolved are in need of food (Godfrey and Parker 1992; Gonzalez-Voyer et al. 2007). Direct feeding method (i.e., when food passes directly from the parent to the chick’s bill) had been proposed to favor aggressive competition because it lowers dominance-broodmate attacks and intimated competitors and hence monopolizes the food. On the other hand, when food is deposited on the nest floor, it was assumed that aggressive competition was not efficient for food monopolization and hence would not be favored by selection (see Mock 1984, 1985). However, the hypothesis that aggression is correlated with clutch size was supported by Pelican and cattle egret data (see Mock’s data and Drummond 1993; Gonzalez-Voyer and Drummond 2007). Following the steps described in the section “Integratingthegest with PGLS,” we tested the alternative causal hypotheses using PPA.

Results

SIMULATIONS

As expected, path analyses undertaken ignoring phylogenetic structure in the data presented very high nominal Type I error rates (see Table 1), with the exception of the simulations scenario under null phylogenetic signal. Low Type I error rates in this last scenario are unsurprising because the data no longer presented any phylogenetic signal and hence analyses using ordinary least squares (OLS) methods are fully justified. On the other hand, PPA presented much lower nominal error rates (see Table 1), although in some cases these were slightly higher than the conventional 0.05 level. PPA outperformed path analysis ignoring phylogeny in all scenarios except one (see Table 1). The only scenario in which path analysis ignoring phylogeny presented a lower Type I error rate was the scenario in which phylogenetic signal was added to the data. As could be expected, the Type I error rate for path analysis ignoring phylogeny decreased as the phylogenetic signal in the simulated data decreased reaching its lowest value when the phylogenetic signal was null, at which point the Type I error rates of both methods converge (see Table 1). Power was much more

<table>
<thead>
<tr>
<th>λ</th>
<th>PGLS</th>
<th>OLS</th>
<th>PGLS</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.068</td>
<td>0.051</td>
<td>0.967</td>
<td>0.964</td>
</tr>
<tr>
<td>0.2</td>
<td>0.072</td>
<td>0.092</td>
<td>0.951</td>
<td>0.944</td>
</tr>
<tr>
<td>0.6</td>
<td>0.054</td>
<td>0.491</td>
<td>0.962</td>
<td>0.947</td>
</tr>
<tr>
<td>0.8</td>
<td>0.065</td>
<td>0.741</td>
<td>0.959</td>
<td>0.956</td>
</tr>
<tr>
<td>1</td>
<td>0.047</td>
<td>0.916</td>
<td>0.954</td>
<td>0.979</td>
</tr>
</tbody>
</table>

Table 1. Type I error and power for Ford-Deppe path analysis models using PGLS Ordinary least squares and data from five hypothetical traits simulated under six different phylogenetic signal scenarios (for λ ranging from 0 to 1). Values on these simulations are provided in the main text.
similar between phylogenetic and nonphylogenetic path analysis methods, indicating that both have relatively similar capability to detect wrong model (see Table 1).

**PPA Analysis of the Evolution of Aggressive Sibling Competition in Birds**

Our first PPA model (model A) tests the directed graph depicted in Figure 2A. This directed graph describes a multiple regression model in which intensity or incidence of aggression (IA) directly depends from egg size (ES), proxy for body size, clutch size (CS), feeding method (FM), and length of the nesting period (L). This model, however, differs from the PGLS model tested in Gonzalez-Voyer et al. (2007) as it implies no covariance among the independent variables. We use this simple model as a starting point to investigate the possible causal effects linking the variables previously suggested to be related (directly or through other variables) with intensity or incidence of aggression (Drummond 2002; Gonzalez-Voyer et al. 2007). The result of the test and the corresponding CIC values of the model are listed in Table 2. The basis set of the conditional independence constraints predicted by model A and all other PPA models presented in this article as well as their associated P-values obtained with PGLS are provided as Supporting Information. Model A is clearly rejected by the data, and looking at the individual separation statements implied by the model we can see that the assumed dependencies between CS and L, CS and ES, as well as L and ES are false (see Supporting Information). We have tested the alternative hypothesis that ES is not directly linked with aggressive competition, but instead is the causal parent of CS and L, leaving the other cause-effect relationships in model A (model B, Figure 2). This model is not rejected by the data using intensity or incidence of aggression as the dependent variable (P-value of Fisher’s test > 0.05, see Table 2), and thus we accept it as a possible explanation of the cause-effect relationships among the variables.

*Figure 2.* Directed acyclic graphs of the tested hypothetical cause-effect models of the relationships among egg size (ES), clutch size (CS), length of the nesting period (L), feeding method (FM), and two indices of aggressive sibling competition (intensity of aggression and incidence of aggression; both labeled in the graphs) in 68 bird species.
Table 2. Summary of the PPA model results for the 14 hypothetical cause-effect models depicted in Figure 3 including intensity of aggression (a), incidence of aggression (b) as proxies of aggressive siblings competition in 68 bird species. The best set of models, with a ∆CICc < 2 is represented in bold.

(a) Intensity of aggression

<table>
<thead>
<tr>
<th>Model</th>
<th>C</th>
<th>k</th>
<th>q</th>
<th>P-value</th>
<th>CICc</th>
<th>ΔCICc</th>
<th>Wi</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>5.28</td>
<td>4</td>
<td>11</td>
<td>0.727</td>
<td>31.994</td>
<td>0</td>
<td>0.293</td>
</tr>
<tr>
<td>B</td>
<td>9.37</td>
<td>5</td>
<td>10</td>
<td>0.497</td>
<td>33.230</td>
<td>1.236</td>
<td>0.158</td>
</tr>
<tr>
<td>L</td>
<td>10.01</td>
<td>5</td>
<td>10</td>
<td>0.439</td>
<td>33.880</td>
<td>1.886</td>
<td>0.114</td>
</tr>
<tr>
<td>M</td>
<td>11.18</td>
<td>5</td>
<td>10</td>
<td>0.343</td>
<td>35.043</td>
<td>3.049</td>
<td>0.064</td>
</tr>
<tr>
<td>D</td>
<td>12.26</td>
<td>5</td>
<td>10</td>
<td>0.268</td>
<td>36.122</td>
<td>4.128</td>
<td>0.037</td>
</tr>
<tr>
<td>F</td>
<td>13.02</td>
<td>5</td>
<td>10</td>
<td>0.222</td>
<td>36.880</td>
<td>4.886</td>
<td>0.026</td>
</tr>
<tr>
<td>N</td>
<td>16.64</td>
<td>6</td>
<td>9</td>
<td>0.163</td>
<td>37.748</td>
<td>5.754</td>
<td>0.017</td>
</tr>
<tr>
<td>O</td>
<td>15.57</td>
<td>6</td>
<td>9</td>
<td>0.212</td>
<td>39.427</td>
<td>7.433</td>
<td>0.007</td>
</tr>
<tr>
<td>E</td>
<td>19.04</td>
<td>6</td>
<td>9</td>
<td>0.087</td>
<td>40.145</td>
<td>8.151</td>
<td>0.005</td>
</tr>
<tr>
<td>H</td>
<td>26.65</td>
<td>6</td>
<td>9</td>
<td>0.008</td>
<td>47.749</td>
<td>15.755</td>
<td>0.000</td>
</tr>
</tbody>
</table>

(b) Incidence of aggression

<table>
<thead>
<tr>
<th>Model</th>
<th>C</th>
<th>k</th>
<th>q</th>
<th>P-value</th>
<th>CICc</th>
<th>ΔCICc</th>
<th>Wi</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>5.68</td>
<td>5</td>
<td>10</td>
<td>0.841</td>
<td>29.547</td>
<td>0</td>
<td>0.339</td>
</tr>
<tr>
<td>K</td>
<td>4.16</td>
<td>4</td>
<td>11</td>
<td>0.842</td>
<td>30.875</td>
<td>1.328</td>
<td>0.174</td>
</tr>
<tr>
<td>N</td>
<td>9.77</td>
<td>6</td>
<td>9</td>
<td>0.636</td>
<td>30.879</td>
<td>1.332</td>
<td>0.174</td>
</tr>
<tr>
<td>D</td>
<td>9.13</td>
<td>5</td>
<td>10</td>
<td>0.520</td>
<td>32.989</td>
<td>3.442</td>
<td>0.061</td>
</tr>
<tr>
<td>L</td>
<td>9.52</td>
<td>5</td>
<td>10</td>
<td>0.484</td>
<td>33.375</td>
<td>3.828</td>
<td>0.050</td>
</tr>
<tr>
<td>M</td>
<td>9.79</td>
<td>5</td>
<td>10</td>
<td>0.459</td>
<td>33.654</td>
<td>4.107</td>
<td>0.043</td>
</tr>
<tr>
<td>C</td>
<td>13.60</td>
<td>6</td>
<td>9</td>
<td>0.327</td>
<td>34.708</td>
<td>5.161</td>
<td>0.026</td>
</tr>
<tr>
<td>F</td>
<td>15.15</td>
<td>5</td>
<td>10</td>
<td>0.127</td>
<td>39.010</td>
<td>9.463</td>
<td>0.003</td>
</tr>
<tr>
<td>G</td>
<td>19.24</td>
<td>6</td>
<td>9</td>
<td>0.083</td>
<td>40.343</td>
<td>10.796</td>
<td>0.002</td>
</tr>
<tr>
<td>H</td>
<td>23.63</td>
<td>6</td>
<td>9</td>
<td>0.023</td>
<td>44.742</td>
<td>15.195</td>
<td>0.000</td>
</tr>
<tr>
<td>A</td>
<td>59.14</td>
<td>6</td>
<td>9</td>
<td>3.23</td>
<td>83.000</td>
<td>53.453</td>
<td>0.000</td>
</tr>
</tbody>
</table>

C, Fisher’s C statistics; k, number of independence claims; q, number of parameters; ∆CICc, difference in CICc from the best fitting model; Wi, CICc weights.

We have shown how PPA can be easily conducted integrating PGLS with the ‘sephMethod’ developed by Shipley (2000b). Using simulations, we showed that PPA correctly identifies the true causal structure of a model with reasonable Type I error rates. On the contrary, path analysis using OL method was shown to be the true signal presented Type I error rates which increased with the level of phylogenetic signal in the data. Type I error rates of ordinary path analysis are comparable to those of PPA only when lambda = 0, which is unsurprising because the data no longer presented any phylogenetic signal and hence analyses using OL methods are fully justified (Freckleton et al., 2002; Revell, 2010). Power was similar in both using PPA and path analysis, ignoring phylogenetic signal. High power in path analysis that ignores phylogenetic relationships is to be expected.
Indeed, a consequence of ignoring phylogenetic nonindependence is higher Type I error rates (Martins and Garland 1991; Rohlf 2006), therefore OLS path analysis simulation incorrectly identify a higher number of significant correlations, which reduces the frequency of non-significant separation and a consequence of their every model. In sum, the results of ignoring phylogenetic nonindependence in path analysis is relatively higher power, but very high Type I error rates. Our simulations clearly indicate that when conducting a confirmatory path analysis on data with an underlying phylogenetic signal and ignoring this signal, the probability of rejecting the true causal model when should have been accepted are unacceptably high, making any inference on the hypothesized underlying causal structure impossible. PA on the contrary efficiently accounts for the added phylogenetic correlation in the data and allows correctly discriminate between correct and wrong hypothesized causal models.

An empirical example of the application of PPA was the analysis of the evolutionary correlate of aggressive sibling strife in birds (Gonzalez-Voyer et al. 2007). PPA confirmed the results of the previous study, identifying the hypothesized causal model linking length of the nesting period and feeding method to aggressive competition (for both incidence and intensity of aggression), as well as the link between clutch size and intensity of aggression (Gonzalez-Voyer et al. 2007). However, PPA allowed us to identify other causal relationships that could not be tested by the correlational analyses previously undertaken. For example, egg size (approx. for nestling brood size and clutch size) was not included in the multiple regression models in the previous study due to the problem of multicollinearity. By applying PPA, we were able to show that egg size represents an indirect causal link with aggressive competition through its effect on clutch size and length of the nesting period. Theoretical arguments had previously suggested that larger nesting size at hatching might enable chicksto efficiently use aggression to intimidate siblings (Drummond 2002). Our results suggest that the hypothesis need to be reframed, as a direct causal link between egg size and aggression is not supported. However, egg size does appear to have an indirect influence on aggressive competition through its effect on clutch size and length of the nesting period. Comparative studies have shown that there is a life-history trade-off in birds between clutch size and egg size, which would explain the causal link between egg size and clutch size (Bennett and Owens 2002). There is also a positive relationship between egg size and fledging age across species, in other words nestlings hatching from larger clutches also tend to present longer nesting periods (Bennett and Owens 2002). According to the sibling competition hypothesis, increased growth rate and hence shorter nesting periods would be favored by siblings (Worschulk and Jackson 1979). However, the hypothesis has been criticized and a comparison of growth rates and lengths of nesting periods in eagles found no support (Bortolotti 1986). Our results show that indeed a link between length of the nesting period and sibling competition. Results of the PPA also suggest there is a causal link between clutch size and feeding method. Such a link had not previously been envisaged by theoretical studies. Its possibility that this link reflects the fact that more than half (51.4%) of the 68 species included in the analysis are Accipitrinae, which tend to have small clutches and also present a developmental switch from direct to indirect feeding. This potential novel link between clutch size and feeding method will need to be analyzed further. Finally, PPA allowed us to determine the directionality of the causal link between clutch size and intensity of aggression. The previous analysis had suggested a significant correlation between the aforementioned traits, however it remained unclear whether small clutches were a cause or consequence of intense sibling aggression. PPA has allowed us to propose that intense aggressive competition among siblings is favored by small clutch size.

In conclusion, we strongly suggest that PPA should be used when undertaking path analysis on multispecies datasets. Use of PPA will result in much reduced Type I error rates compared to path models ignoring phylogenetic structure in the data whereas power will not be compromised. Our empirical example of the application of PPA demonstrated how application of the method...
allowed us to propose novel causal hypotheses between species traits and the evolution of aggressive sibling strife. Our results suggest that large nesting size at hatching indirectly favors evolution of aggressive competition through its effects on clutch size and length of the nesting period. Furthermore, the results confirmed the causal link between clutch size and intensity of aggression and allowed us to determine the directionality of the causal link, proposing that small clutches favor more intense aggressive competition.

ACKNOWLEDGMENTS

We are grateful to S. Blanchet for useful discussions about the implementation of the information criterion approach in the framework of path analysis. Many thanks are due to L. Kubatko, B. Shipley, and anonymous reviewers for useful comments and suggestions to the manuscript. AV’s support in his research activity by the Gran Paradiso National Park. Part of the work was done while AV was hosted as a visiting research fellow at the National Centre for Statistical Ecology, University of Kent, UK, supported by the European Union, the Autonomous Region Aosta Valley, and the Italian Ministry of Work and Social Providence.

AG-V was funded by a Juan de la Cierva postdoctoral fellowship. This work was partly funded by project CGL 2010–21250 (to AG-V) of the Spanish Ministry of Economy and Competitiveness. The authors declare that no conflict of interest exists.

LITERATURE CITED


Supporting Information
The following supporting information is available for this article:

Table S1. R code used for simulations.
Table S2. Aggressive broodmate competition data used for the phylogenetic path analyses.
Table S3. Basis sets of conditional independence constraints.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.