

# DISENTANGLING EVOLUTIONARY CAUSE-EFFECT RELATIONSHIPS WITH PHYLOGENETIC CONFIRMATORY PATH ANALYSIS

Achaz von Hardenberg<sup>1,2,3</sup> and Alejandro Gonzalez-Voyer<sup>3,4</sup>

<sup>1</sup>Alpine Wildlife Research Centre, Gran Paradiso National Park, Degioz 11, 11010 Valsavarenche (Aosta), Italy

<sup>2</sup>E-mail: achaz.hardenberg@pngp.it

<sup>4</sup>Department of Integrative Ecology, Estación Biológica de Doñana (EBD-CSIC), Av Américo Vespucio s/n, 41092 Sevilla, Spain

Received February 13, 2012

Accepted August 5, 2012

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 a 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  $\lambda = 0$  to 0.916 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 revisit 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.

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

The phylogenetic comparative method has become a widely used tool 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 all together (although one particular method does allow to determine contingency, see Pagel and Meade 2006). The results are generally interpreted as allowing—at best—to identify a subset of variables that evolve in a correlated fashion or that differences exist in the trait of interest between two groups of species. Indeed, the fully randomized experiment is the

<sup>3</sup>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 have to be 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 observed data impose constraints on the patterns of correlation that such data 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 independences 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 number 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 become 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 (Sol et al. 2005, 2010) or did not specify the method 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 the authors to undertake the path analysis accounting for the statistical 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 (Quader 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 model fit. The role of this parameter is to determine the amount of phylogenetic signal in the data (in the residuals 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).

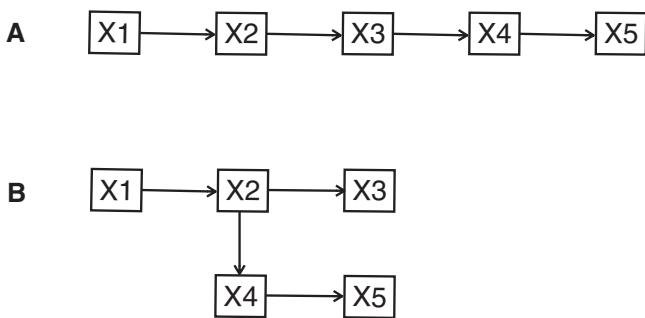
Our proposed method for phylogenetic confirmatory path analysis (hereafter called PPA), integrates PGLS with the d-sep test developed by Shipley (2000b). This method exploits the concept of d-separation (Pearl 1988; Verma and Pearl 1988) to predict the minimal set of conditional probabilistic independencies that must all 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 correlational structure predicted if the causal model is correct. The d-sep test is a very general test that can be used for small sample sizes (because the inferential tests are not asymptotic), nonnormally distributed data (although the phylogenetic comparative methods we will use here do assume normal distribution of the phylogenetically transformed residuals), and nonlinear functional relationships. The only disadvantage of d-sep tests is that they 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 simulations to explore the consequences of

ignoring phylogeny when undertaking confirmatory path analysis by d-sep tests. Finally, we revisit a previously published analysis of the evolutionary correlates of aggressive sibling strife in birds (Gonzalez-Voyer et al. 2007) as an empirical example of the implementation of the method.

## Methods

### INTEGRATING THE D-SEP TEST WITH PGLS

Shipley (2009) showed how the d-sep test 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-sep test with PGLS and thus perform a PPA. Although the method was already described in detail elsewhere, for didactic reasons, we present here the four steps involved in the d-sep test for confirmatory path analysis, with additional details about how to combine it with PGLS (for a more detailed account on the procedure for nonphylogenetic 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 a 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 nonadjacent variables, that is, those not directly joined by an edge.



**Figure 1.** Directed acyclic graphs describing two alternative models of possible cause-effect relationships among five variables.

Thus, for the model in Figure 1A the list would be  $[(X1, X3), (X1, X4), (X1, X5), (X2, X4), (X2, X5), (X3, X5)]$ . Then one lists the parent variables of either nonadjacent variables in the previous list, that is,  $[\{X2\}, \{X3\}, \{X4\}, \{X1, X3\}, \{X1, X4\}, \{X2, X4\}]$ . Simply combining these two lists, one obtains the basis set of the d-separation statement describing the probabilistic independence between the two nonadjacent variables, conditioned on the parent variables of both; that is, for the model in Figure 1A, the basis set would be  $[(X1, X3)\{X2\}, (X1, X4)\{X3\}, (X1, X5)\{X4\}, (X2, X4)\{X1, X3\}, (X2, X5)\{X1, X4\}, (X3, X5)\{X2, X4\}]$ . Following the notation of Shipley (2004),  $(X1, X3)\{X2\}$  indicates that  $X1$  is probabilistically independent from  $X3$  conditional on the variable  $X2$  whereas  $(X2, X4)\{X1, X3\}$  indicates that  $X2$  is probabilistically independent from  $X4$  conditional on the variables  $X1$  and  $X3$ . We leave it to the reader to derive the basis set for the model in Figure 1B. The third step, in the context of this article, consists in testing each conditional independence, derived from the d-sep statements in the basis set, by linear models of the type (taking as an example the first d-sep statement of the basis set listed above):  $X3 = X2 + X1$ , to calculate the probability ( $p_i$ ) that the partial regression coefficient associated with  $X1$  is 0 (i.e., the effect of  $X1$  on  $X3$  conditional on  $X2$ ). 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 species traits 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 combining all the values of  $p_i$  (i.e., the probabilities that the nonadjacent variables in the basis set are statistically independent conditional on their parent variables) using Fisher’s  $C$  statistic

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

where  $k$  is the number of independence 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 the set of accepted path models. Shipley (2000a) proposed

an approach based on testing the difference in Fisher's  $C$  statistics of two competing nested models, which follows a  $\chi^2$  distribution with  $\Delta\text{df} = \text{df}_{\text{model1}} - \text{df}_{\text{model2}}$ . 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. As an appealing alternative, which can also be used for selecting among nonnested 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 nonphylogenetic 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):

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

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 readers to Burnham and Anderson 1998 and Grueber et al. 2011). Although Cardon et al. (2011) call this information criterion AICc, 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-sep method (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 analyses using the d-sep method (Shipley 2000b). We simulated evolution of five hypothetical traits using a prespecified covariance matrix among the traits determining a specific path 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 null to strong phylogenetic signal in the simulated data; or in other words from traits evolving along a star phylogeny, where trait evolution for each species is in-

dependent, 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 the simulated 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 can take any value between 1 and 0, where high values indicate strong phylogenetic signal and low values indicate low phylogenetic covariance in the data (see Freckleton et al. 2002). The simulated phylogeny was transformed based on values of  $\lambda$  ranging from 0.8 to 0 (i.e., 0.8, 0.6, 0.4, 0.2, and 0) prior to simulating trait evolution and tests of conditional independencies done using the untransformed tree. For each of the six scenarios we simulated 1000 datasets, each with an underlying phylogenetic tree of a 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 to determine the level of 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 independencies consistent with the "true" underlying causal model depicted in Fig. 1A) and the power (i.e., 1—the Type II error, the probability of not rejecting the tested model when it is actually false, testing the predicted set of conditional independencies of a "wrong" 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 (see Gonzalez-Voyer et al. 2007). Feeding method was a continuous variable expressed as the

proportion of the nestling period (from hatching until fledging) during which feeding is direct, ranging from 0 (indirect feeding throughout the nestling period) to 1 (direct feeding throughout the nestling period). For species with a developmental transition in feeding method, the proportion was calculated on the basis of the average age at which chicks switched from one method to the other. Clutch size was used as a proxy for brood size at hatching with which it was highly and significantly correlated (Gonzalez-Voyer et al. 2007). Egg size was used as a proxy for nestling body size at hatching with which it was highly and significantly correlated (Gonzalez-Voyer et al. 2007). Finally, average length of the nestling 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 same topology as in 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 nestling periods are known to be associated through life-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 in species in which aggressive competition has evolved to reduce the costs (Godfray and Parker 1992; Gonzalez-Voyer et al. 2007). Direct feeding method (i.e., when food passes directly from the parent's to the chick's bill) had been proposed to favor aggressive competition because it al-

lowed dominant broodmates to attack and intimidate competitors and hence monopolize the food. On the other hand, when food is deposited on the nest floor (indirect feeding) 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 had been criticized (Drummond 2001a,b) and field studies with pelicans and cattle egrets did not support it (Pinsón and Drummond 1993; Gonzalez-Voyer and Drummond 2007). Following the steps described in the section "Integrating the d-sep test with PGLS," we tested these alternative causal hypotheses using PPA.

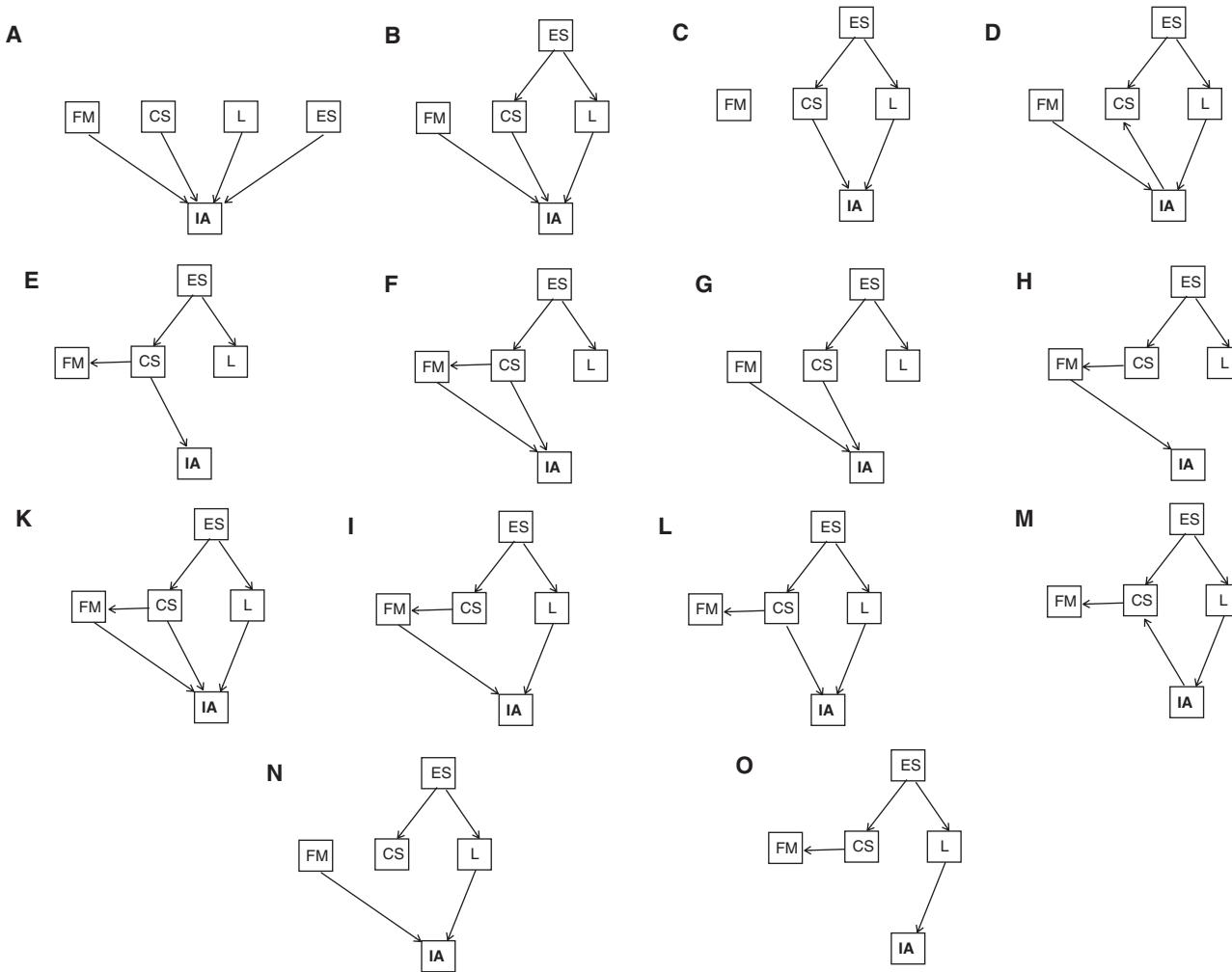
## Results

### 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 simulation 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 than that of the phylogenetic path analysis was when data were simulated without any phylogenetic signal whatsoever. As could be expected, the Type I error rate for path analysis ignoring phylogeny decreased as the phylogenetic signal in the simulated data also 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 1.** Type I error and power for d-sep path analysis models using PGLS or ordinary least squares on data from five hypothetical traits simulated under six different phylogenetic signal scenarios (for  $\lambda$  ranging from 0 to 1). Details on the simulations are provided in the main text.

$\lambda$	Correct path model (Type I error)		Wrong path model (Power)	
	PGLS	OLS	PGLS	OLS
0	0.068	0.051	0.967	0.964
0.2	0.072	0.092	0.951	0.944
0.4	0.065	0.253	0.973	0.945
0.6	0.054	0.491	0.962	0.947
0.8	0.065	0.741	0.959	0.956
1	0.047	0.916	0.954	0.979



**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 IA in the graphs) in 68 bird species.

similar between phylogenetic and nonphylogenetic path analysis methods, indicating that both have relatively similar capability to detect a 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 depend from egg size (ES; a 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 (Drum-

mond 2002; Gonzalez-Voyer et al. 2007). The results of the d-sep test and the corresponding CICc value 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 d-separation statements implied by the model we can see that the assumed independencies between CS and L, CS and ES, as well as L and ES are false (see Supporting information). We thus 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 as in model A (model B, Fig. 2). This model is not rejected by the data using intensity or incidence of aggression as the dependent variable (*P*-value of Fisher's *C* test > 0.05, see Table 2), and thus we accept it as a possible explanation of the cause-effect relationships among the variables.

**Table 2.** Summary of the PPA model results for the 14 hypothetical cause-effect models depicted in Figure 3 including intensity of aggression (a) or incidence of aggression (b) as proxies of aggressive sibling competition in 68 bird species. The best set of models, with a  $\Delta\text{CICc} < 2$  is represented in bold.

(a) Intensity of aggression							
Model	C	k	q	P-value	CICc	$\Delta\text{CICc}$	Wi
<b>K</b>	<b>5.28</b>	<b>4</b>	<b>11</b>	<b>0.727</b>	<b>31.994</b>	<b>0</b>	<b>0.293</b>
<b>I</b>	<b>8.83</b>	<b>5</b>	<b>10</b>	<b>0.548</b>	<b>32.693</b>	<b>0.699</b>	<b>0.207</b>
<b>B</b>	<b>9.37</b>	<b>5</b>	<b>10</b>	<b>0.497</b>	<b>33.230</b>	<b>1.236</b>	<b>0.158</b>
<b>L</b>	<b>10.01</b>	<b>5</b>	<b>10</b>	<b>0.439</b>	<b>33.880</b>	<b>1.886</b>	<b>0.114</b>
M	11.18	5	10	0.343	35.043	3.049	0.064
C	14.11	6	9	0.294	35.212	3.218	0.059
D	12.26	5	10	0.268	36.122	4.128	0.037
F	13.02	5	10	0.222	36.880	4.886	0.026
N	16.64	6	9	0.163	37.748	5.754	0.017
G	17.11	6	9	0.145	38.212	6.218	0.013
O	15.57	6	9	0.212	39.427	7.433	0.007
E	19.04	6	9	0.087	40.145	8.151	0.005
H	26.65	6	9	0.008	47.749	15.755	0.000
A	59.14	6	9	$3.23 \times 10^{-8}$	83.000	51.006	0.000

(b) Incidence of aggression							
Model	C	k	q	P-value	CICc	$\Delta\text{CICc}$	Wi
<b>I</b>	<b>5.68</b>	<b>5</b>	<b>10</b>	<b>0.841</b>	<b>29.547</b>	<b>0</b>	<b>0.339</b>
<b>K</b>	<b>4.16</b>	<b>4</b>	<b>11</b>	<b>0.842</b>	<b>30.875</b>	<b>1.328</b>	<b>0.174</b>
<b>N</b>	<b>9.77</b>	<b>6</b>	<b>9</b>	<b>0.636</b>	<b>30.879</b>	<b>1.332</b>	<b>0.174</b>
B	8.25	5	10	0.604	32.110	2.563	0.094
D	9.13	5	10	0.520	32.989	3.442	0.061
L	9.52	5	10	0.484	33.375	3.828	0.050
M	9.79	5	10	0.459	33.654	4.107	0.043
O	10.28	6	9	0.591	34.146	4.599	0.034
C	13.60	6	9	0.327	34.708	5.161	0.026
F	15.15	5	10	0.127	39.010	9.463	0.003
G	19.24	6	9	0.083	40.343	10.796	0.002
E	21.56	6	9	0.043	42.662	13.115	0.000
H	23.63	6	9	0.023	44.742	15.195	0.000
A	59.14	6	9	$3.23 \times 10^{-8}$	83.000	53.453	0.000

C, Fisher's C statistics; k, number of independence claims; q, number of parameters;  $\Delta\text{CICc}$ , difference in CICc from the best fitting model; Wi, CICc weights.

Model B suggests that nestling size at hatching (through its proxy Egg size) may indeed have an indirect influence on aggressive competition through its effects on length of the nestling period and clutch size. As there is, however, controversy in the literature regarding the effects of feeding method and length of the nestling period on aggressive competition (see Bortolotti 1986; Drummond 2001a,b; Gonzalez-Voyer and Drummond 2007), and there could be an effect of clutch size on feeding method, we hypothesized and tested 12 other causal models (models C–O,

Fig. 2) with different possible combinations of the causal links among L, CS, FM, and IA. The results of all these PPA models and their relative explanatory power, expressed as CICc weights (Wi), compared to each other including the previously described models, are summarized in Table 2. Among these models, we also specifically tested the hypothesis that intensity and incidence of aggression have a causal effect on clutch size, an inverse causal relationship to that assumed in the other models, that is, smaller clutches are influenced by intensity or incidence of aggression rather than the other way round (model D and M; Fig. 2). Both model D and M are not rejected by the data, applying Fisher's C test, and thus provide a plausible explanation, using intensity as well as incidence of aggression in the model (Table 2). However, looking at the difference in CICc values between these and the best fitting model ( $\Delta\text{CICc}$ ), the former appear to perform poorly compared to models in which the link between IA and CS is in the other direction (i.e., clutch size influences sibling aggression). PPA thus allowed us to determine that the most likely direction of the causal relationship between clutch size and intensity of aggression is clutch size influencing intensity of aggression, hence as predicted, smaller clutches favor the evolution of more intense aggressive sibling strife (Drummond 2002; Gonzalez-Voyer et al. 2007). All of the best fitting models, with a  $\Delta\text{CICc} < 2$ , predict a strong causal link between L and IA (for both intensity and incidence of aggression). A causal link between CS and FM appears to be supported also by the best models (for both intensity and incidence of aggression) and with some support between FM and IA. The causal links between CS and IA instead are only supported when IA represents intensity of aggression but not incidence, which is in accord with the results of Gonzalez-Voyer et al. (2007). The standardized path coefficients with standard errors and their 95% confidence intervals, averaged among the best fitting models (with  $\Delta\text{CICc} < 2$ ), are provided in Table 3.

## Discussion

We have shown how PPA can be easily conducted integrating PGLS with the d-sep method 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 OLS methods and thus ignoring the underlying phylogenetic signal presented Type I error rates which increase 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 OLS methods are fully justified (Freckleton et al. 2002; Revell 2010). Power was similarly high both using PPA and path analysis ignoring phylogenetic signal. High power in a path analysis that ignores phylogenetic relationships is to be expected.

**Table 3.** Standardized path coefficients (Coeff.) with standard errors (SE) and their lower and upper 95% confidence intervals (L95%CI and U95%CI, respectively), averaged among the best fitting models (with  $\Delta\text{AICc} < 2$ ) obtained after model selection for models including intensity of aggression (intensity) or incidence of aggression (incidence) as proxies of aggressive sibling competition in 68 bird species.

Intensity					Incidence			
Path	Coeff.	SE	L95%CI	U95%CI	Coeff.	SE	L95%CI	U95%CI
CS -> IA	-0.27	0.13	-0.52	-0.02	-0.09	0.13	-0.34	0.16
L -> IA	0.33	0.14	0.06	0.60	0.35	0.11	0.14	0.56
FM -> IA	-0.16	0.12	-0.38	0.07	-0.16	0.10	-0.36	0.03
CS -> FM	-0.22	0.13	-0.49	0.04	-0.22	0.14	-0.49	0.05

ES, egg size; CS, clutch Size; L, length of the nesting period; FM, feeding method; IA, intensity or incidence of aggression.

Indeed, a consequence of ignoring phylogenetic nonindependence is higher Type I error rates (Martins and Garland 1991; Rohlf 2006), therefore OLS path analysis models incorrectly identify a higher number of significant correlations, which reduces the frequency of nonsignificant d-separation and as a consequence tend to reject every model. In sum, the end result of ignoring phylogenetic nonindependence in path analysis is relatively high 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 it should have been accepted are unacceptably high, making any inference on the hypothesized underlying causal structure impossible. PPA on the contrary efficiently accounts for the added phylogenetic correlations in the data and allows to correctly discriminate between correct and wrong hypothesized causal models.

As an empirical example of the application of PPA, we revisited the analysis of the evolutionary correlates 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 nestling 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 correlative analyses previously undertaken. For example, egg size (a proxy for nestling body size at hatching) was not included in the multiple regression models in the previous study to avoid problems of multicollinearity. By applying PPA, we were able to show that egg size presents an indirect causal link with aggressive competition through its effect on clutch size and length of the nestling period. Theoretical arguments had previously suggested that larger nestling size at hatching might enable chicks to efficiently use aggression to intimidate siblings (Drummond 2002). Our results suggest that the hypothesis needs 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 nestling 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 eggs also tend to present longer nestling periods (Bennett and Owens 2002). According to the sibling competition hypothesis, increased growth rate and hence shorter nestling periods would be favored by sibling strife (Werschul and Jackson 1979). However, the hypothesis has been criticized and a comparison of growth rates and lengths of nestling periods in eagles found no support for it (Bortolotti 1986). Our results show there is indeed a link between length of the nestling 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. It is possible that this link reflects the fact that more than half (51.4%) of the 68 species included in the analysis were Accipitridae, 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 analyses had identified 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 nestling size at hatching indirectly favors evolution of aggressive competition through its effect on clutch size and length of the nestling 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 the 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 an anonymous reviewer for useful comments and suggestions to the manuscript. AVH is supported in his research activity by the Gran Paradiso National Park. Part of the work was done while AVH 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 CGL2010-21250 (to AG-V) of the Spanish Ministry of Economy and Competitiveness. The authors declare that no conflict of interest exists.

#### LITERATURE CITED

- Bennett, P. M., and I. P. F. Owens. 2002. Evolutionary ecology of birds. Oxford Univ. Press, Oxford, U.K.
- Bortolotti, G. R. 1986. Evolution of growth rates in eagles: sibling competition vs energy considerations. *Ecology* 67:182–194.
- Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York.
- Cardon, M., G. Loot, G. Grenouillet, and S. Blanchet. 2011. Host characteristics and environmental factors differentially drive the burden and pathogenicity of an ectoparasite: a multilevel causal analysis. *J. Anim. Ecol.* 80:657–667.
- Drummond, H. 2001a. A revaluation of the role of food in broodmate aggression. *Anim. Behav.* 61:517–526.
- . 2001b. The control and function of agonism in avian broodmates. *Adv. Study Behav.* 30:261–301.
- . 2002. Begging versus aggression in avian broodmate competition. Pp. 337–360 in J. Wright and M. L. Leonard, eds. *The evolution of nestling begging: competition, cooperation and communication*. Kluwer Academic Press, Dordrecht, The Netherlands.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- Fisher, R. A. 1926. The arrangement of field experiments. *J. Ministry Agric.* 33:503–513.
- Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* 160: 712–726.
- Garland, T. J., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst. Biol.* 41:18–32.
- Godfray, H. C. G., and G. A. Parker. 1992. Sibling competition, parent-offspring conflict and clutch size. *Anim. Behav.* 43:473–490.
- Gonzalez-Voyer, A., and H. Drummond. 2007. Is broodmate aggression really associated with direct feeding? *Behaviour* 144:373–392.
- Gonzalez-Voyer, A., T. Szekely, and H. Drummond. 2007. Why do some siblings attack each other? Comparative analysis of aggression in avian broods. *Evolution* 61:1946–1955.
- Grueber, C. E., S. Nakagawa, R. J. Laws, and I. G. Jamieson. 2011. Multimodel inference in ecology and evolution: challenges and solutions. *J. Evol. Biol.* 24:699–711.
- Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24: 129–131.
- Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford Univ. Press, Oxford, U.K.
- Martins, E. P. 2000. Adaptation and the comparative method. *Trends Ecol. Evol.* 15:296–299.
- Martins, E. P., and T. Garland. 1991. Phylogenetic analyses of the correlated evolution of continuous characters – a simulation study. *Evolution* 45:534–557.
- Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* 149:646–667.
- Martins, E. P., J. A. Diniz-Filho, and E. A. Housworth. 2002. Adaptive constraints and the phylogenetic comparative method: a computer simulation test. *Evolution* 56:1–13.
- Mock, D. W. 1984. Siblicidal aggression and resource monopolization in birds. *Science* 225:731–733.
- . 1985. Siblicidal brood reduction: the prey size hypothesis. *Am. Nat.* 125:327–343.
- Pagel, M., and A. Meade. 2006. Bayesian analysis of correlated evolution of discrete characters by Reversible-Jump Markov Chain Monte Carlo. *Am. Nat.* 167:808–825.
- Paradis, E. 2006. Analysis of phylogenetics and evolution with R. Springer, New York.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Pearl, J. 1988. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, San Mateo, CA.
- Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and the R Development Core Team. 2011. nlme: linear and nonlinear mixed effects models. R package version 3.1–102.
- Pinson, D., and H. Drummond. 1993. Brown pelican siblicide and the prey-size hypothesis. *Behav. Ecol. Sociobiol.* 32:111–118.
- Quader, S., K. Isvaran, R. E. Hale, B. G. Miner, and N. E. Seavy. 2004. Nonlinear relationships and phylogenetically independent contrasts. *J. Evol. Biol.* 17:709–715.
- R Development Core Team. 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Revell, L. J. 2010. Phylogenetic signal and linear regression on species data. *Meth. Ecol. Evol.* 1:319–329.
- Rohlf, F. J. 2006. A comment on phylogenetic correction. *Evolution* 60:1509–1515.
- Santos, J. C., and D. C. Cannatella. 2011. Phenotypic integration emerges from aposematism and scale in poison frogs. *Proc. Natl. Acad. Sci. U.S.A.* 108:6175–6180.
- Shipley, B. 2000a. Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference. Cambridge Univ. Press, Cambridge, U.K.
- . 2000b. A new inferential test for path models based on directed acyclic graphs. *Struct. Equ. Modeling* 7:206–218.
- . 2004. Analyzing the allometry of multiple interacting traits. *Persp. Plant Ecol. Evol. Syst.* 6:235–241.
- . 2009. Confirmatory path analysis in a generalized multilevel context. *Ecology* 90:363–368.

- Sol, D., L. Lefebvre, and J. D. Rodriguez-Treijeiro. 2005. Brain size, innovative propensity and migratory behaviour in temperate Palaearctic birds. *Proc. R. Soc. B* 272:1433–1441.
- Sol, D., N. Garcia, A. Iwaniuk, K. Davis, A. Meade, W. A. Boyle, and T. Székely. 2010. Evolutionary divergence in brain size between migratory and resident birds. *PLoS One* 5:e9617.
- Verma, T., and J. Pearl. 1988. Causal networks: semantics and expressiveness. Pp. 352–359 in *Proceedings of the 4th workshop on uncertainty in artificial intelligence*. Mountain View, CA. Schachter, R., Levitt, T. S., and Kanal, L. N., eds. *Uncertainty in artificial intelligence*. Vol. 4. Pp. 69–76. Elsevier, Amsterdam.
- Werschkul, D. F., and J. A. Jackson. 1979. Sibling competition and avian growth rates. *Ibis* 121:97–102.

Associate Editor: L. Kubatko

## *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.