JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12869

Discovering phenotypic causal structure from nonexperimental data

J. OTSUKA

Department of Philosophy, Kobe University, Kobe, Japan

Keywords:

causal models;
natural selection;
phenotypic integration;
quantitative genetics;
theory.

Abstract

The evolutionary potential of organisms depends on how their parts are structured into a cohesive whole. A major obstacle for empirical studies of phenotypic organization is that observed associations among characters usually confound different causal pathways such as pleiotropic modules, interphenotypic causal relationships and environmental effects. The present article proposes causal search algorithms as a new tool to distinguish these different modes of phenotypic integration. Without assuming an *a priori* structure, the algorithms seek a class of causal hypotheses consistent with independence relationships holding in observational data. The technique can be applied to discover causal relationships among a set of measured traits and to distinguish genuine selection from spurious correlations. The former application is illustrated with a biological data set of rat morphological measurements previously analysed by Cheverud *et al.* (*Evolution* 1983, 37, 895).

Introduction

The variational properties and evolutionary potential of phenotypic characters are largely determined by the underlying developmental mechanism. Darwin, in his discussion on 'correlation of growth', has emphasized that causal connections among distinct traits influence and sometimes constrain evolutionary changes. In the quantitative genetics literature, the study of constraints on adaptive evolution has focused on the additive genetic covariance matrix (the **G** matrix) of the Lande equation $\Delta\bar{\mathbf{z}} = \mathbf{G}\boldsymbol{\beta}$ (Lande, 1979). In particular, eigenvectors of **G** associated with large eigenvalues are interpreted to represent directions to which adaptive response is less constrained (Arnold *et al.*, 2001; Blows & Hoffmann, 2005; Blows, 2007; Walsh & Blows, 2009). Conversely, if one or more of the eigenvalues is zero, evolutionary trajectories are restricted to a lower-dimensional hypersurface in the adaptive landscape, meaning that there are phenotypic combinations unattainable by the population provided that its **G** matrix remains constant. The constraint in this context is a purely phenomenological concept – the **G** matrix or its eigenstructure facilitates predicting *how* a population will respond to a given selective pressure, without explaining *why* on the physiological or developmental ground. In other words, it does not reveal the causal basis of evolutionary changes.

An alternative approach is to explicitly model the causal structure over phenotypic characters using a directed graph or *path diagram* (Wright, 1920; Li, 1975; Lynch, 1988; Crespi & Bookstein, 1989; Mitchell-Olds & Bergelson, 1990; Mitchell, 1992). The technique has been applied in life-history studies to distinguish indirect from direct fitness contributions of traits, with an aim to provide a more detailed picture of selection than that captured by selection gradient $\boldsymbol{\beta}$ that focuses solely on the proximate causes of fitness. Path analysis can also be used to explicate the constraints summarized by the **G** matrix in terms of direct causal relationships among phenotypes. For a set of n traits, a linear path model can be represented by an $n \times n$ matrix whose ij element is the path coefficient of the ith trait on the jth trait and is zero if there is no edge between them. Let **B** be such a matrix and define $\boldsymbol{\Phi} = (\mathbf{I} - \mathbf{B})^{-1}$, where **I** is an identity matrix. Then, the **G** matrix can be written as:

$$\mathbf{G} = \boldsymbol{\Phi}\mathbf{G}_{\varepsilon}\boldsymbol{\Phi}^{\mathrm{T}}, \qquad (1)$$

where $\mathbf{G}_{\varepsilon}$ is the covariance of the additive genetic components *not* attributable to direct causal relationships

Correspondence: Jun Otsuka, Department of Philosophy, Kobe University, Rokko-dai 1-1, Nada, Kobe, Japan. Tel.: +81 (0)78 803 5591; fax: +81 (0)78 803 5589; e-mail: junotk@gmail.com

among traits and the superscript T is a matrix transpose (Gianola & Sorensen, 2004). Inserting this into the Lande equation and rearranging yield

$$\Delta\bar{\mathbf{z}} = \mathbf{\Phi}\mathbf{G}_{\varepsilon}\boldsymbol{\eta}, \quad (2)$$

where the *extended selection gradients* $\boldsymbol{\eta} = \mathbf{\Phi}^{\mathrm{T}}\boldsymbol{\beta}$ measure the total (direct and indirect) effects of traits on the relative fitness (Morrissey, 2014).

Morrissey's equation (2) describes the change in mean phenotype as a function of the phenotypic causal structure ($\mathbf{\Phi}$) as well as selective pressures ($\boldsymbol{\eta}$) and the covariance of exogenous genetic values ($\mathbf{G}_{\varepsilon}$). To fit the model to an actual population, however, one must first identify the path diagram, which is not a trivial task. The heuristic use of repeated multiple regression to construct a path diagram (Mitchell-Olds & Bergelson, 1990) is problematic for a regression coefficient inevitably confounds a direct causal relationship between traits with genetic associations (Kempthorne, 1978). Confirming a causal link with a manipulative experiment or assuming it on an *a priori* ground does not solve the issue because it does not eliminate the possibility of genetic confounding that may bias the estimation of the path coefficient. In general, a phenotypic correlation may arise in multiple, nonexclusive manners including a direct causal (developmental or life history) link, environmental/genetic/phenotypic confounding, and linkage disequilibrium. Although equation (2) requires one to separately estimate these possible sources of phenotypic associations, no regression method can achieve this owing to insufficient degrees of freedom (Cowley & Atchley, 1992). Further, the construction of a path diagram presupposes the orientation of each causal edge, which is underdetermined from a symmetric correlational relationship.

Confounding also represents a serious obstacle in evolutionary predictions based on Lande's equation, one of its vital assumptions being that selection gradients $\boldsymbol{\beta}$ reflect all and only direct causal contributions from the phenotype to fitness, free from any unobserved environmental, phenotypic or genetic confounding (Mitchell-Olds & Shaw, 1987; Hadfield, 2008; Morrissey *et al.*, 2010). This cannot be guaranteed by the regression method alone. In another context, the $\mathbf{G}$ matrix has been used not only as a statistical summary of evolutionary potential, but also to infer the underlying genetic architecture or developmental *modules* (e.g. Atchley, 1984; Riska, 1986; Cheverud, 1996; Hansen, 2006; Hansen & Houle, 2008; Polly, 2008). A variety of methods have been proposed to detect a modular structure from genetic or phenotypic correlations (e.g. Cheverud & Buikstra, 1981; Cheverud *et al.*, 1983; Zelditch, 1988; Phillips & Arnold, 1999; Mezey *et al.*, 2000; Magwene, 2001, 2008; Mitteroecker & Bookstein, 2007, 2008), many of which are variants of the exploratory or confirmatory factor analysis. The underlying assumption of factor analysis and other proposals to study genetic modules is that all phenotypic associations derive from pleiotropic genes with no horizontal (phenotypic) causal relations. This assumption, again, remains untested by these statistical methods.

Hence, identifying the causal structure of an evolving population proves both essential and challenging in either approach. The difficulty is familiar with the dictum 'correlation is not causation' – statistical information underdetermines causal relationships. Although this is true in general, recent decades have seen rapid developments of causal search algorithms that attempt to infer causal facts from statistical data (Pearl, 2000; Spirtes *et al.*, 2000). Rather than directly fitting free parameters of an *a priori* model, these algorithms exploit the patterns of conditional independence to narrow down the range of causal hypotheses consistent with the observed data. A reduction in the hypothesis space facilitates the determination of what model to fit with the data quantitatively. Although causal modelling has been applied in ecology (Shipley, 2000, 2010) and breeding literature (Valente *et al.*, 2010, 2011; Rosa *et al.*, 2011), its implication to the study of evolution and relative significance vis-a-vis the existing methods are not well documented. The present article aims to address this deficiency. Following a brief introduction of the basic idea of causal search, this article explores its potential in identifying a path model and distinguishing genuine selection from spurious correlations. Using the data reported by Cheverud *et al.* (1983) as input, it will be demonstrated that the algorithm creates an ontogenetic process model that corresponds well with their findings. The method presented herein, however, is not 'a silver bullet'. The last section discusses its limitations and precautions in biological applications.

Constraint-based causal search

Theoretical background

The goal of causal search is to construct a path diagram from observational data with no or minimum assumptions regarding a specific form of the causal structure to be identified. The task consists of determining the presence or absence of a direct causal connection (adjacency) for each pair of variables and orienting causal edges. This section describes how the search algorithm manages these two subtasks. A more complete treatment can be found in Spirtes *et al.* (2000) or Pearl (2000). For an accessible introduction to biological contexts, see Shipley (2000, 2010).

The basic idea of causal search is that a probability distribution is generated from the underlying causal structure and as such contains some trace of the latter in the form of conditional independence. As an example, suppose Fig. 1 represents the true causal structure underlying four observed variables $\{X_1, X_2, X_3, X_4\}$. It is then expected that although X_4 correlates with its

 J. EVOL. BIOL. doi: 10.1111/jeb.12869

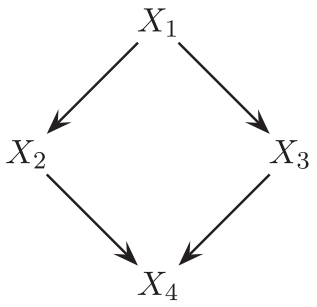


Fig. 1 Example of collider where two edges $X_2 \rightarrow X_4$ and $X_4 \leftarrow X_3$ collide at X_4.

remote cause X_1, this dependence will be cleared when conditioned upon the intermediates X_2 and X_3; in formal notation, $X_1 \not\perp_p X_4$ and $X_1 \perp_p X_4 \mid \{X_2, X_3\}$, where $\perp_p$ denotes probabilistic independence and $\not\perp_p$ dependence. The generative structure thus constrains possible forms of distribution in such a manner that causal (in) dependence between variables implies their probabilistic (in)dependence. The notion of causal independence is more formally defined by *d-separation* (Pearl, 1988, 2000). A path p is said to be d-separated by a set of variables **Z** if

1 p contains a chain $i \rightarrow m \rightarrow j$ or a fork $i \leftarrow m \rightarrow j$ such that the middle variable m is in **Z**, or

2 p contains a collider $i \rightarrow m \leftarrow j$ such that the middle variable m or any of its causal descendants is not in **Z**.

We also say that variables X and Y are d-separated by **Z** if every path between X and Y is d-separated by **Z** and denote this as $X \perp_d Y \mid \mathbf{Z}$. Note that whereas conditional independence ($\perp_p$) is a probabilistic notion that can be tested vis-a-vis observed data, d-separation ($\perp_d$) is a topological feature of an unknown path diagram. A distribution is called *Markov* with respect to a given path diagram if d-separation relationships in the latter entail probabilistic independence holding in the former such that $\mathbf{X} \perp_d \mathbf{Y} \mid \mathbf{Z} \Rightarrow \mathbf{X} \perp_p \mathbf{Y} \mid \mathbf{Z}$ for any choice of variables **X**, **Y** and **Z**. When the opposite direction holds (probabilistic dependence implies d-separation), the distribution is called *faithful* or *stable*.

When do these two conditions hold? The Markov condition requires that all common causes that simultaneously affect more than one variable in the data set be also measured; that is, there is no confounding. Suppose in the previous example, X_1 and X_4 were confounded by a latent variable. Then, they would no longer become independent even conditioned on the observed intermediates (i.e. $X_1 \not\perp_p X_4 \mid \{X_2, X_3\}$), despite that they are d-separated (i.e. $X_1 \perp_d X_4 \mid \{X_2, X_3\}$) in the original diagram that omits the confounder. Because such confounding is expected to be the rule rather than the exception with phenotypic records, the Markov condition can hardly be assumed in the present context.

Faithfulness is violated when there are multiple causal pathways whose effects cancel each other. In Fig. 1, this occurs when two pathways $X_1 \rightarrow X_2 \rightarrow X_4$ and $X_1 \rightarrow X_3 \rightarrow X_4$ exert the opposite influences such that the net effect of X_1 on X_4 is zero. If the system is linear, this amounts to $\boldsymbol{\beta}_{41 \mid 2} = -\boldsymbol{\beta}_{41 \mid 3}$, where $\boldsymbol{\beta}_{41 \mid k}$ is the partial regression coefficient of X_4 on X_1 given X_k, $k \in \{2, 3\}$. Although not impossible, one may expect such a fine-tuned cancellation to be unlikely (Spirtes *et al.*, 2000) and even if it occurs, it would not remain invariant under slight changes in the parameters (Pearl, 2000).

The causal search algorithm described below assumes faithfulness, but not the Markov property of an input distribution. The algorithm consists of three steps, adjacency, orientation and follow-up. The first step of the algorithm constructs a nondirected graph that represents what variable is causally connected to another. Under the faithfulness assumption, an observed probabilistic independence implies the pair of variables to be d-separated. Using this property, the adjacency step determines, for any pair of variables in the data set, whether they are causally disconnected by testing their conditional independence. More precisely,

1 For each pair of variables X and Y, find a set of variables $S(X, Y)$ (possibly empty) such that $X \perp_p Y \mid S(X, Y)$. Connect X and Y with an undirected edge if and only if there is no such set.

Provided that all statistical decisions of independence are made correctly, this rule yields an *adjacency graph* in which the *absence* of an edge between a pair of variables guarantees that they do not cause each other directly nor have a latent common cause. However, without the Markov condition, the *presence* of an edge does not necessarily represent a direct causal relationship – it only means that they are inseparably correlated and this association may be due to a confounding factor.

Given an adjacency graph, the next step is to orient its undirected edges. The orientation process makes the use of a specific structure called *collider* where two edges 'collide' at the middle node such that $X \rightarrow Z \leftarrow Y$ (see the second part of the definition of d-separation). A unique property of the collider is that the variables at both ends (X, Y) become *dependent* conditioned on the middle (Z). For example, the collider $X_2 \rightarrow X_4 \leftarrow X_3$ in Fig. 1 yields $X_2 \perp_p X_3 \mid X_1$ but also $X_2 \not\perp_p X_3 \mid \{X_1, X_4\}$. An illustration of collider in the context of evolution is the linkage disequilibrium induced by selection, where selection creates a correlation between two uncorrelated loci L_1 and L_2 both affecting fitness W. In such cases, $L_1 \rightarrow W \leftarrow L_2$ forms a collider and conditioning on W by selection induces a genetic correlation. Colliders whose two ends are not connected to each other directly are called *unshielded*. Of course, we do not know which part of a given adjacency graph forms a collider, or even an unshielded collider. But because only unshielded colliders can produce the unique pat-

tern of independence just described, one can infer colliding points and orient edges around them based on the results of statistical tests. Using a wildcard symbol '$*$' to represent either the arrowtail '$-$' or the head '$>$', the second rule for edge orientation is as follows:

2 For each triplet X, Z, Y such that X is adjacent to Z, Z is adjacent to Y and X is *not* adjacent to Y, orient $X*\!-\!* Z *\!-\!* Y$ as $X *\!\rightarrow Z \leftarrow\!* Y$ if and only if Z is not in S(X, Y).

Because all the unshielded colliders in the true graph must be identified by this rule, the remaining ternary relationships can be further oriented such that they do not create a new unshielded collier:

3 If there is a subgraph $X *\!\rightarrow Y *\!-\!* Z$ with X and Z being nonadjacent, orient this as $X *\!\rightarrow Y *\!\rightarrow Z$.

Finally, if we can assume that the true causal structure is *acyclic* or *recursive*, that is, if no variable causes itself, the following rule is applicable:

4 If there is a directed path from X to Y where each edge is pointing towards Y and they are directly connected by an edge $X*\!-\!* Y$, orient the edge as $X *\!\rightarrow Y$.

Otherwise, X and Y form a cycle and are causes of themselves, contrary to the assumption of acyclicity.

These rules are implemented by the FCI (Spirtes *et al.*, 2000) and IC* (Pearl, 2000) algorithms. From a sample distribution summarized by a covariance matrix or cell counts, these algorithms output a *partial ancestor graph* or PAG, which may contain four types of edges with the following meanings (Spirtes *et al.*, 1995):

- $X \rightarrow Y$: X causes Y. They may or may not be confounded.
- $X \leftrightarrow Y$: X and Y are confounded, with no direct causal relationship in between.
- X o$\rightarrow$ Y: Either $X \rightarrow Y$ or $X \leftrightarrow Y$.
- X o–o Y: A causal relationship is suspected between X and Y; however, the algorithm could not determine its form.

There are two caveats in interpreting directed edges in a PAG. First, although an edge $X \rightarrow Y$ implies X to be a cause of Y, the relation may be *indirect*; that is, the causal influence of X on Y may be mediated by other variables in the data set. Further, directed edges, in general, do not exclude possible confounding factors, that is, although $X \rightarrow Y$ indicates X to be a cause of Y, they may further be confounded by other unobserved factors. There are, however, cases where such possibilities can be eliminated, as we will see below. These complications do not arise under the Markov condition, that is when there is no confounding variable.

Figure 2 illustrates an application of the above rules. The true causal structure is on the left, with $\{X_1, X_2\}$ and $\{X_3, X_4\}$ being confounded by latent factors. In distributions faithful to this graph, the pairs of variables $\{X_1, X_4\}$ and $\{X_2, X_3\}$ are independent given X_2 and X_1, respectively (i.e. $X_1 \perp_p X_4 \mid X_2$ and $X_2 \perp_p X_3 \mid X_1$), which enables the first step of the algorithm to build the adjacency graph (b). The same pairs, however, become dependent when conditioned on different variables, X_3 and X_4, respectively ($X_1 \not\perp_p X_4 \mid X_3$ and $X_2 \not\perp_p X_3 \mid X_4$). These facts are used in Step 2 to identify the unshielded colliders $X_1 *\!\rightarrow X_3 \leftarrow\!* X_4$ and $X_2 *\!\rightarrow X_4 \leftarrow\!* X_3$ in the PAG (c). The third and fourth rules are not applicable in this example, leaving the remaining edges undetermined (denoted by the empty circles). Although the result is partial, this example illustrates how facts regarding the underlying causal structure can be learned from statistical data alone. More can be inferred if the time order is known, as we will see below.

Causal search with the mixed model

As mentioned above, the Markov condition cannot be assumed in most phenotypic records owing to possible genetic or environmental confounding. Valente *et al.* (2010, 2011) have proposed a use of the standard mixed model to statistically eliminate genetic associations to restore the Markov property. A linear path model for a trait vector $\mathbf{z}$ with random additive genetic effects $\mathbf{u}$ and residual deviations $\mathbf{e}$ can be written as

$$\begin{aligned}\mathbf{z} &= \mathbf{B}\mathbf{z} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e} \\ &= \boldsymbol{\Phi}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Phi}(\mathbf{u} + \mathbf{e})\end{aligned} \tag{3}$$

where $\mathbf{B}$ is a path coefficient matrix, $\mathbf{X}$ is a design matrix, $\boldsymbol{\beta}$ is a vector of fixed effects and $\boldsymbol{\Phi} = (\mathbf{I} - \mathbf{B})^{-1}$ as defined above. Under the standard assumptions of the multitrait mixed model (e.g. $\mathbf{u}$ and $\mathbf{e}$ are normally distributed and independent of each other), the phenotypic covariance matrix is

$$\begin{aligned}\mathrm{Var}(\mathbf{Z}) &= \boldsymbol{\Phi}\mathrm{Var}(\mathbf{U})\boldsymbol{\Phi}^{\mathrm{T}} + \boldsymbol{\Phi}\mathrm{Var}(\mathbf{E})\boldsymbol{\Phi}^{\mathrm{T}} \\ &= \mathbf{G} + \mathbf{R}\end{aligned} \tag{4}$$

Valente *et al.* (2010) note that $\mathbf{G}$ and $\mathbf{R}$ are the additive genetic and residual covariance matrices estimated by a mixed model. The residual component thus contains information regarding the phenotypic causal structure ($\boldsymbol{\Phi}$) and serves as an input to a causal search algorithm. If Var($\mathbf{E}$) is diagonal, the conditional (residual) distribution $P(\mathbf{Z} \mid \mathbf{u})$ satisfies the Markov condition even if the unconditional $P(\mathbf{Z})$ does not, which allows one to use a more powerful search algorithm such as the IC or PC algorithm (Pearl, 2000; Spirtes *et al.*, 2000). The parameters of the obtained path model can then be estimated using Bayesian Markov chain Monte Carlo method (Gianola & Sorensen, 2004; Valente *et al.*, 2010, 2011).

This approach will be particularly effective under pervasive genetic associations, which may hide direct causal relationships among phenotypes. The advantage is balanced by increased standard errors in the estimated residual covariances, which may impair statistical

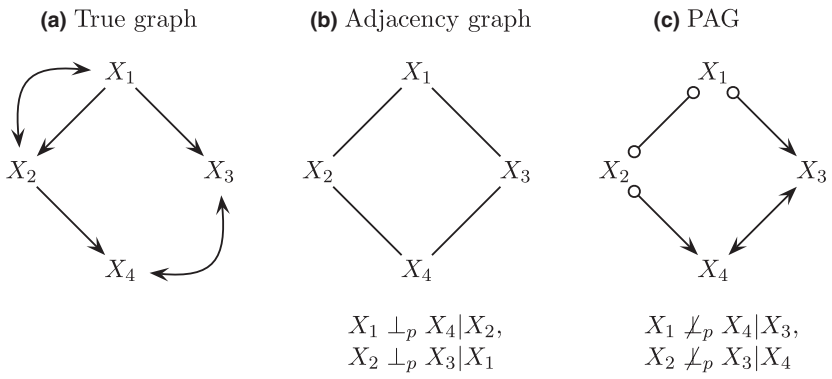


Fig. 2 Illustration of the causal discovery algorithm with the (in) dependence relationships used in each step of inference. (a) True causal structure with confounding (curved bidirected edges). (b) Adjacency graph obtained after Step 1. (c) PAG obtained after Step 2. Open circles indicate where the algorithm could not determine the direction of causal influence.

decisions of independence. Moreover, the mixed model is inapplicable to observational field studies that lack pedigree information or cannot dismiss the possibility of environmental confounding. The FCI algorithm described above will prove useful in such observational studies or even in designed experiments where some form or another of nongenetic confounding is suspected or cannot be discarded.

Applications

Inferring a developmental structure

This section applies the causal search algorithm described above to biological data reported in Cheverud *et al.* (1983). Based on a longitudinal growth study of 561 cross-fostered rats (Atchley & Rutledge, 1980), Cheverud and colleagues estimated additive genetic, maternal and residual covariance matrices for the log-transformed body weight measured at ages 14, 28, 42, 56, 70 and 189 days. The causal connections among the six ontogenetic stages were explored using the FCI algorithm in the TETRAD software version 4.3.10-7 (available at http://www.phil.cmu.edu/projects/tetrad/). Assuming normality as in the original study, Fisher's z was used to test independence:

$$z = \frac{\sqrt{n - |\mathbf{C}| - 3}}{2} \ln\left(\frac{1 + r_{XY.\mathbf{C}}}{1 - r_{XY.\mathbf{C}}}\right)$$

where n is the sample size and $r_{XY.\mathbf{C}}$ is the sample partial correlation of X and Y given $\mathbf{C}$. If the population correlation is zero, z is asymptotically standard normal (Anderson, 2003). In each test, the independence hypothesis was rejected at the 5% significance level; however, the overall results remained robust at the 1% level. In addition to the reported matrices, the time order of the measurement points was used to restrict possible directions of causal flow.

Figure 3(a) is the output of the FCI algorithm applied to the phenotypic correlation matrix of the six weight measurements. As expected, the result indicates a sequential causal path (straight edges), with some possible connections between distant stages (curved edges). As noted earlier, however, these curved edges in the PAG do not necessarily imply direct causal relationships; rather, they may reflect genetic or environmental confounding. For example, the curved edge $W_{56} \rightarrow W_{189}$ does not entail a direct influence of the weight at day 56 on that at day 189, but is consistent with an indirect causal path $W_{56} \rightarrow W_{70} \rightarrow W_{189}$ with some confounding in between.

To clear the remaining uncertainties, the FCI algorithm was reapplied to the residual correlation matrix, which amounts to the partial correlation matrix conditioned on additive genetic components and thus is free from additive genetic confounding (Valente *et al.*, 2010). The output PAG indicates a linear pathway (Fig. 3b) with a possible confounding effect only between W_{14} and W_{42}. The result suggests that all the curved edges in Fig. 3(a) do not represent direct causal links, but rather are artefacts of genetic confounding.

Furthermore, Fig. 3(b) allows us to conclude the absence of a significant environmental confounding for the stages later than W_{28}. To see this, suppose that the edge $W_{42} \rightarrow W_{56}$ is confounded. Then, W_{42} would be the middle point both of a collider $W_{28} \rightarrow W_{42} \leftrightarrow W_{56}$ and of a chain $W_{28} \rightarrow W_{42} \rightarrow W_{56}$, which means that no set could make W_{28} and W_{56} independent (because if a conditioning set contains W_{42}, the collider is not d-separated, whereas if it does not, the chain is not d-separated). However, if that were the case, they must be connected by Step 1 of the algorithm, contrary to the actual output. We can thus conclude that edge $W_{42} \rightarrow W_{56}$ is not confounded. As noted earlier, edges in a PAG are consistent with confounding in general; however, the possibility can be excluded if there is an additional causal input (W_{28} in this case), called *instrumental variable*, to the cause variable. In our example, all variables other than W_{189} serve as instrumental variables in the PAG (Fig. 3b) to clear the unconfoundedness of the edges later than W_{28}.

Figure 3(c) summarizes the above results. Recall that the curved edges in graph (a) suggested either direct causal links or confounding of environmental or genetic

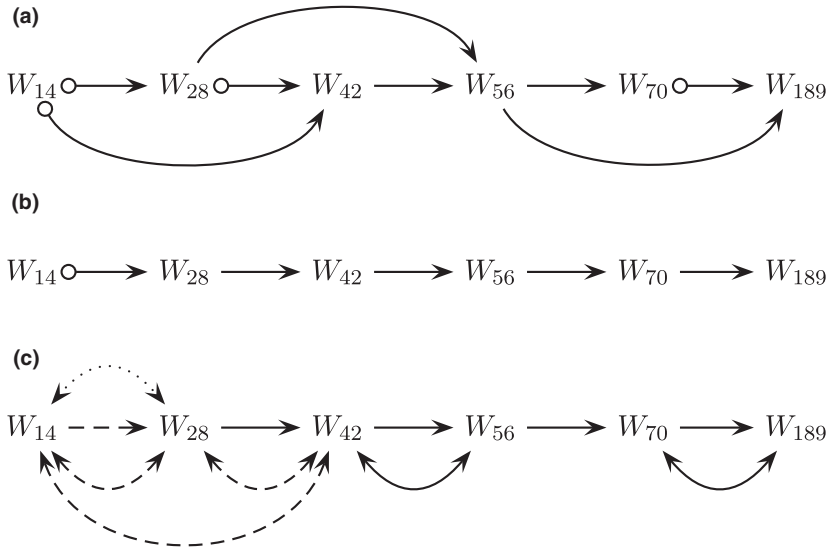


Fig. 3 PAGs obtained from the FCI algorithm applied to the rat measurement data in Cheverud *et al.* (1983). (a) Output from the phenotypic correlation matrix, which suggests linear pathways with possible causal influences over distant stages (curved edges). (b) Causal search on the residual matrix confirms that the phenotypic casual structure is sequential. (c) Results indicate that the curved edges in (a) reflect not a direct relationship, rather a genetic confounding (solid bidirected edges). Dashed edges indicate possible genetic confounding (bidirected, curved), interphenotypic causation (directed, straight) and environmental confounding (dotted, curved), respectively.

origin. The first two possibilities, however, are eliminated by the output from the residual correlation matrix (b), leaving genetic confounding to be the only consistent explanation. By reversing the above reasoning of instrumental variables, genetic confounding of W_{70} and W_{189} makes W_{70} the middle point of both a collider and a chain. This means that W_{56} and W_{189} are dependent conditioned on any phenotypic variable and thus connected in graph (a). For the same reason, we can conclude that the curved edge $W_{28} \rightarrow W_{56}$ in (a) was induced by genetic confounding between W_{42} and W_{56}. The solid bidirected edges at the bottom of graph (c) denote these genetic associations. Conversely, the dashed edges represent the remaining uncertainties due primarily to the lack of instrumental variables at the early stages. If the individual phenotypic records and pedigree information are known (which are no longer available for the present data; J. Cheverud, pers. comm), the path and genetic correlation coefficients of model (Fig. 3c) can be estimated using the Bayesian MCMC technique mentioned above (Gianola & Sorensen, 2004).

The obtained causal graph suggests that the early development of rat body weight is controlled by age-specific growth factors rather than by global pleiotropic effects. This hypothesis is consistent with the subsequent study on mice (Cheverud *et al.*, 1996) that found distinct QTLs for each age-specific body weight, with few QTLs affecting post-natal growth as a whole. In the sequential pathway, the phenotypic effect of a single QTL would be detectable for only a limited time frame owing to diminishing autocorrelation. The autocorrelation model also explains the observation made by both the original and subsequent studies that associations between age-specific weights decline as the time interval increases in all (phenotypic, genetic, maternal and residual) correlation matrices (Cheverud *et al.*, 1983, 1996). We thus conclude that the causal hypothesis obtained from the search algorithm is not only consistent with the input data but also extrapolatable to a further study.

Identifying selective pressures

The issue of confounding also appears significant in empirical studies of selection (Lande & Arnold, 1983). It is well known that the ordinary least-square estimate $\hat{\boldsymbol{\beta}}$ of selection gradients is biased in the presence of unconditioned genetic, phenotypic or environmental confounding factors (Mitchell-Olds & Shaw, 1987; Rausher, 1992; Mauricio & Mojonniner, 1997; Hadfield, 2008; Morrissey *et al.*, 2010). If an unmeasured environmental factor X affects both phenotype Z_1 and fitness W, regressing W on Z_1 yields a nonzero $\hat{\beta}_1$ even if there is no selection. Moreover, if Z_1 is causally affected by or shares genetic or environmental common causes with another trait Z_2, the estimate $\hat{\beta}_2$ of its selection gradient is also biased because the multiple regression amounts to conditioning on the colliding point of $Z_2 \rightarrow Z_1 \leftrightarrow W$ or $Z_2 \leftrightarrow Z_1 \leftrightarrow W$ (Table 1, first and second rows). Hence, a confounding factor on only one trait may impair the entire selection estimators. The bias is

detectable by comparing the regression slopes of fitness on the focal traits on the one hand and those on their breeding values on the other (Queller, 1992; Rausher, 1992; Morrissey *et al.*, 2010, 2012). This, however, requires phenotypic records of the offspring generation and therefore does not serve the purpose of predicting evolutionary responses.

Although there exists no infallible method to detect environmental confounders, under certain conditions it is possible to distinguish genuine selection from a spurious correlation using the method of instrumental variables discussed above. An instrumental variable in the present context is an auxiliary trait that causes or is genetically or environmentally confounded with the focal trait, but not so with the fitness measure (Z_2 in Table 1). The first condition is met by any trait previous in time that has a nonzero phenotypic correlation with the focal trait. Under the faithfulness assumption, a sufficient condition for the second requirement is the trait being independent of the fitness measure given some (possibly empty) set.

Let Z_2 be such a trait, and $S(Z_2, W)$ be the set that renders Z_2 independent of the fitness (i.e. $Z_2 \perp_p W \mid S(Z_2, W)$). Then, the association between fitness and Z_1 implies selection without confounding if and only if Z_1 is included in $S(Z_2, W)$. The rationale for the rule is as follows. If W and Z_1 are confounded, $Z_2 \ast\!\!\rightarrow Z_1 \leftrightarrow W$ forms an unshielded collider. Then, the set $S(Z_2, W)$ cannot contain Z_1 because Z_2 and W would be dependent given Z_1. Conversely if they are not confounded, the only possibilities are $Z_2 \leftrightarrow Z_1 \rightarrow W$ or $Z_2 \rightarrow Z_1 \rightarrow W$ because the fitness cannot be a cause of a phenotype. In either case, Z_2 and W are dependent unless conditioned on Z_1 and hence $Z_1 \in S(Z_2, W)$. Table 1 illustrates this with bivariate cases.

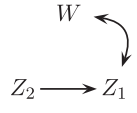
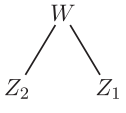
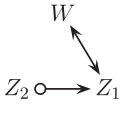
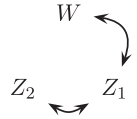
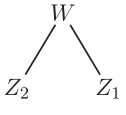
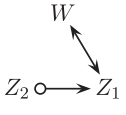
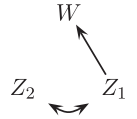
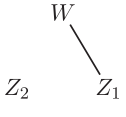
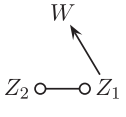
The test requires no more than the standard significance test of selection gradients; that is, a distribution must be faithful and allow independent tests. Upon determining the set of unconfounded phenotypes, the selection gradients can be estimated by regressing the fitness only on such a subset.

Discussion

The adaptive response of a population is a function of its causal structure and genetic variances (Otsuka, 2015, 2016). Owing to the inherent difficulty and limited feasibility of manipulative experiments, however, our knowledge on organismal architectures and selective regimes must often rely on statistical analyses of observational data. A common issue in nonexperimental studies is that a phenotypic correlation confounds genetic associations, environmental effects and a direct causal relationship between phenotypes (Kempthorne, 1978; Cowley & Atchley, 1992). The present article proposed a new approach to discern these three sources of statistical associations based on differential patterns of conditional independence. Unlike conventional statistical tools that try to fit *a priori* structure to data, the causal search attempts to build a path diagram based on the general conditions regarding the relationship between causality and probability.

The key assumption of the FCI algorithm is faithfulness, which claims that an active (not *d*-separated) cau-

Table 1 Comparison of multiple regression and the FCI algorithm under different selective scenarios

	Truth	Regression	FCI	Independence
Confounded				$W \perp_p Z_2 \mid \emptyset$ $(\mathbf{S}(Z_2, W) = \emptyset)$
Confounded				$W \perp_p Z_2 \mid \emptyset$ $(\mathbf{S}(Z_2, W) = \emptyset)$
Selection				$W \perp_p Z_2 \mid Z_1$ $(\mathbf{S}(Z_2, W) = \{Z_1\})$

The leftmost column depicts three possible cases where only the bottom row represents selection. Z_2 is an auxiliary phenotype *d*-separated from fitness. This means that there exists a set $S(Z_2, W)$ conditioned on which Z_2 is independent of W. The test declares selection on Z_1 if $S(Z_2, W)$ includes Z_1 (bottom). Note that multiple regression (second column) provides no information as to the possible confounding and falsely concludes selection on unselected trait Z_2 if there is confounding.

sal link yields a statistical dependence, or by contraposition, that independence entails *d*-separation. In practice, however, the algorithm requires a stronger assumption that a causal relationship must yield a *significant* association to trim edges based upon nonsignificant results of independence tests (Zhang & Spirtes, 2003). Strictly speaking, therefore, the absence of an edge in an output graph does not warrant a conclusion of causal unrelatedness; rather, it reflects the fact that the statistical test did not detect a significant total effect. The qualification is especially imminent in the causal inference of life-history traits where various types of trade-off are expected. If, for example, negatively correlated traits are positively selected, the trade-off among the fitness components may lead to an almost unfaithful distribution, which invites type II errors and consequent incorrect edge omissions. The risk can be reduced by using a more inclusive variable set or increasing the sample size. No finite sample size, however, can guarantee that the probability of overlooking an existing causal influence is less than a given threshold, because any causal search procedure for confounded data is only *pointwise consistent* but not *uniformly consistent* (Robins *et al.*, 2003). That is, we do not know how powerful our test must be to detect a causal relationship because the effect in question can be arbitrarily small (i.e. almost unfaithful). Thus, although the FCI algorithm can correctly judge the adjacency at the asymptotic limit, no finite sample can be guaranteed to approximate this result. This limitation is not peculiar to the algorithm, but rather applies to every nonexperimental method including any type of regression analysis.

Whereas type II errors result in edge omissions, type I errors – mistakenly rejecting independence hypotheses – lead to incorrect edge commissions. Given that the algorithm involves repeated tests, there is a non-negligible probability of making incorrect statistical decisions even with a reasonably small significance level. The overall error probability of the algorithm is difficult to evaluate because (i) tests are not independent of each other; (ii) an incorrect statistical decision may be offset by subsequent tests without affecting the final output; (iii) conversely, one error may have multiple consequences, for example yielding incorrect adjacency and orientation; and all of these depend on the unknown causal structure (Spirtes *et al.*, 2000, p. 96). An alternative strategy is to control the false discovery rate of the adjacency step of the algorithm to curb the expected proportion of the falsely discovered links to all those discovered under a specific value (Li & Wang, 2009).

Similar to any other statistical method, the output from the algorithm should not be accepted as a solid fact, but as a hypothesis for further experimental studies. Provided these caveats, causal modelling provides a promising approach to build, identify and examine a causal hypothesis from observational data. The emphasis on the model construction is what distinguishes causal search from the conventional statistical methods that focus on model fitting. Because these model assumptions remain basically untested, the estimated parameters frequently present a misleading picture regarding the biology of the organisms under study. The above discussed study of Cheverud *et al.* (1983), for example, identified two major principal components in the estimated **G** matrix, which they interpreted as two modules of pleiotropic genes with distinct causal roles – the first group regulating the height of the growth curve and the second its shape. Such global pleiotropic effects, however, were not confirmed by the subsequent study (Cheverud *et al.*, 1996) or the present reanalysis. Rather, the two principal components are likely to be artefacts resulting from the sequential ontogenetic pathway. Because the upstream stages mainly determine the intercept of the growth curve and the later stages regulate its slope, genetic variances at these different stages may have been detected as principal components each associated with the curve height on the one hand and its shape on the other. If this is the case, the detected 'components' do not point to distinct biological mechanisms. In general, reification of principal components as genetic modules assumes the absence of interphenotypic causal relationships, whereas the causal interpretation of path coefficients estimated by path analysis or structural equation modelling presupposes no genetic or environmental confounding. The method described in this study can be used to examine these causal assumptions and biological interpretations of the statistical models.

Another possible area of application is the study of selection. Despite its crucial importance in understanding adaptive evolution, the empirical studies of selection have paid relatively little attention to the possible bias in the estimated selection gradients due to unobserved confounders (Mitchell-Olds & Shaw, 1987; Rausher, 1992; Mauricio & Rausher, 1997; Hadfield, 2008; Morrissey *et al.*, 2010). Although selection and a spurious correlation are indistinguishable from a statistical association alone, they generate distinctive patterns of conditional independence that can be tested with the aid of an instrumental trait variable. In addition, a violation of faithfulness in this context is less likely because it only happens when an unobserved fitness contribution of the auxiliary trait cancels the statistical dependence that arises from conditioning on the collider, and there is no biological reason to expect this to occur. To repeat, this does not imply impossibility, but given the scarcity of alternatives the method provides a useful tool to detect a bias of evolutionary predictions with fewer assumptions than conventional regression analysis.

Conclusions

A major obstacle for observational studies of phenotypic integration or selection is that statistical associations among traits or fitness are presumably confounded by multiple unknown factors. The present article described a causal search algorithm that exploits the patterns of conditional independence to discern direct causal relationships between phenotypes from genetic or environmental confounding and to determine the direction of causal influence using the property of unshielded colliders. The method was illustrated with the growth data of rats reported in Cheverud *et al.* (1983) and its application to distinguish selection from spurious correlation was suggested with a hypothetical example. The causal hypothesis obtained from the rat data explains the observed patterns of pleiotropy in accord with the common understanding of the ontogenetic mechanism and a subsequent study (Cheverud *et al.*, 1996). In combination with confirmatory experiments, the causal search algorithm can provide valuable information for furthering our understanding of the generative structure of phenotype and its implication to adaptive evolution.

Acknowledgments

I am grateful to Jeremy Berg, James Cheverud, Brandon Cooper, Bruce Glymour, Clark Glymour, James Griesemer, Michael Morrissey and anonymous reviewers for helpful comments on earlier versions of the manuscript. I also thank Graham Coop, Steve Sterns and Günter Wagner for discussions. The author has no conflict of interest to declare. A part of this study was supported by the Japan Society for the Promotion of Science Grant-in-Aid for Postdoctoral Fellows (25-659).

References

Anderson, T.W. 2003. *An Introduction to Multivariate Statistical Analysis*, 3rd edn. Wiley, Hoboken, New Jersey.

Arnold, S.J., Pfrender, M.E. & Jones, A.G. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* **112–113**: 9–32.

Atchley, W.R. 1984. Ontogeny, timing of development, and genetic variance-covariances structure. *Am. Nat.* **123**: 519–540.

Atchley, W.R. & Rutledge, J.J. 1980. Genetic components of size and shape. I. Dynamics of components of phenotypic variability and covariability during ontogeny in the laboratory rat. *Evolution* **34**: 1161–1173.

Blows, M.W. 2007. A tale of two matrices: multivariate approaches in evolutionary biology. *J. Evol. Biol.* **20**: 1–8.

Blows, M.W. & Hoffmann, A.A. 2005. A reassessment of genetic limits to evolutionary change. *Ecology* **86**: 1371–1384.

Cheverud, J.M. 1996. Developmental integration and the evolution of pleiotropy. *Integr. Comp. Biol.* **36**: 44–50.

Cheverud, J.M. & Buikstra, J. 1981. Quantitative genetics of skeletal non-metric traits in the rhesus macaques on Cayo Santiago. II. Phenotypic, genetic, and environmental correlations between traits. *Am. J. Phys. Anthropol.* **54**: 51–58.

Cheverud, J.M., Rutledge, J.J. & Atchley, W.R. 1983. Quantitative genetics of development: genetic correlations among age-specific trait values and the evolution of ontogeny. *Evolution* **37**: 895–905.

Cheverud, J.M., Routman, E.J., Duarte, F.A.M., van Swinderen, B., Cothran, K. & Perel, C. 1996. Quantitative trait loci for murine growth. *Genetics* **142**: 1305–1319.

Cowley, D. & Atchley, W.R. 1992. Quantitative genetic models for development, epigenetic selection, and phenotypic evolution. *Evolution* **46**: 495–518.

Crespi, B.J. & Bookstein, F.L. 1989. A path-analytic model for the measurement of selection on morphology. *Evolution* **43**: 18–28.

Gianola, D. & Sorensen, D. 2004. Quantitative genetic models for describing simultaneous and recursive relationships between phenotypes. *Genetics* **167**: 1407–1424.

Hadfield, J.D. 2008. Estimating evolutionary parameters when viability selection is operating. *Proc. Biol. Sci.* **275**: 723–734.

Hansen, T.F. 2006. The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. Syst.* **37**: 123–157.

Hansen, T.F. & Houle, D. 2008. Measuring and comparing evolvability and constraint in multivariate characters. *J. Evol. Biol.* **21**: 1201–1219.

Kempthorne, O. 1978. Logical, epistemological and statistical aspects of nature-nurture data interpretation. *Biometrics* **34**: 1–23.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**: 402–416.

Lande, R. & Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.

Li, C. 1975. *Path Analysis-a Primer*. Boxwood Press, Pacific Grove, California.

Li, J. & Wang, Z. 2009. Controlling the false discovery rate of the association/causality structure learned with the PC algorithm. *J. Mach. Learn. Res.* **10**: 475–514.

Lynch, M. 1988. Path analysis of ontogenetic data. In: *The Dynamics of Size-Structured Populations* (L. Persson & B. Ebenman, eds), pp. 29–46. Springer-Verlag, Berlin.

Magwene, P.M. 2001. New tools for studying integration and modularity. *Evolution* **55**: 1734–1745.

Magwene, P.M. 2008. Using correlation proximity graphs to study phenotypic integration. *Evol. Biol.* **35**: 191–198.

Mauricio, R. & Mojonniner, L. 1997. Reducing bias in the measurement of selection. *Trends Ecol. Evol.* **12**: 433–436.

Mauricio, R. & Rausher, M.D. 1997. Experimental manipulation of putative selective agents provides evidence for the role of natural enemies in the evolution of plant defense. *Evolution* **51**: 1435–1444.

Mezey, J.G., Cheverud, J.M. & Wagner, G.P. 2000. Is the genotype-phenotype map modular? A statistical approach using mouse quantitative trait loci data. *Genetics* **156**: 305–311.

Mitchell, R.J. 1992. Testing evolutionary and ecological hypotheses using path analysis and structural equation modelling. *Funct. Ecol.* **6**: 123–129.

Mitchell-Olds, T. & Bergelson, J. 1990. Statistical genetics of an annual plant, Impatiens capensis. II. Natural selection. *Genetics* **124**: 416–421.

Mitchell-Olds, T. & Shaw, R.G. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* **41**: 1149–1161.

Mitteroecker, P. & Bookstein, F.L. 2007. The conceptual and statistical relationship between modularity and morphological integration. *Syst. Biol.* **56**: 818–836.

Mitteroecker, P. & Bookstein, F.L. 2008. The evolutionary role of modularity and integration in the hominoid cranium. *Evolution* **62**: 943–958.

Morrissey, M.B. 2014. Selection and evolution of causally covarying traits. *Evolution* **68**: 1748–1761.

Morrissey, M.B., Kruuk, L.E. & Wilson, A. 2010. The danger of applying the breeder's equation in observational studies of natural populations. *J. Evol. Biol.* **23**: 2277–2288.

Morrissey, M.B., Parker, D.J., Korsten, P., Pemberton, J.M., Kruuk, L.E. & Wilson, A.J. 2012. The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder's equation. *Evolution* **66**: 2399–2410.

Otsuka, J. 2015. Using causal models to integrate proximate and ultimate causation. *Biol. Philos.* **30**: 19–37.

Otsuka, J. 2016. Causal foundations of evolutionary genetics. *Brit. J. Philos. Sci.* **67**: 247–269.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, California.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY.

Phillips, P.C. & Arnold, S.J. 1999. Hierarchical comparison of genetic variance-covariance matrices. I. Using the flury hierarchy. *Evolution* **53**: 1506–1515.

Polly, P. 2008. Developmental dynamics and G-Matrices: can morphometric spaces be used to model phenotypic evolution? *Evol. Biol.* **35**: 83–96.

Queller, D.C. 1992. Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* **139**: 540–558.

Rausher, M.D. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* **46**: 616–626.

Riska, B. 1986. Some models for development, growth, and morphometric correlation. *Evolution* **40**: 1303–1311.

Robins, B.J.M., Scheines, R., Spirtes, P. & Wasserman, L. 2003. Uniform consistency in causal inference. *Biometrika* **90**: 491–515.

Rosa, G., Valente, B. & De los Campos, G. 2011. Inferring causal phenotype networks using structural equation models. *Genet. Select. Evol.* **43**: 1–13.

Shipley, B. 2000. *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference*. Cambridge University Press, New York, NY.

Shipley, B. 2010. *From Plant Traits to Vegetation Structure: Chance and Selection in the Assembly of Ecological Communities*. Cambridge University Press, Cambridge, UK.

Spirtes, P., Meek, C. & Richardson, T. 1995. Causal inference in the presence of latent variables and selection bias. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (P. Besnard & S. Hanks, eds), pp. 499–506. Morgan Kauffmann, San Francisco, California.

Spirtes, P., Glymour, C. & Scheines, R. 2000. *Causation, Prediction, and Search*, 2nd edn. The MIT Press, Cambridge, Massachusetts.

Valente, B., Rosa, G., De los Campos, G., Gianola, D. & Silva, M. A. 2010. Searching for recursive causal structures in multivariate quantitative genetics mixed models. *Genetics* **185**: 633–644.

Valente, B.D., Rosa, G.J.M., Silva, M.A., Teixeira, R.B. & Torres, R.A. 2011. Searching for phenotypic causal networks involving complex traits: an application to European quail. *Genet. Select. Evol.* **43**: 37.

Walsh, B. & Blows, M.W. 2009. Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. *Annu. Rev. Ecol. Evol. Syst.* **40**: 41–59.

Wright, S. 1920. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proc. Natl. Acad. Sci.* **6**: 320–332.

Zelditch, M. 1988. Ontogenetic variation in patterns of phenotypic integration in the laboratory rat. *Evolution* **42**: 28–41.

Zhang, J. & Spirtes, P. 2003. Strong faithfulness and uniform consistency in causal inference. In: *Proceedings of the Nineteenth Conference Annual Conference on Uncertainty in Artificial Intelligence* (C. Meek & U. Kjærulff, eds), pp. 632–639. Morgan Kauffmann, San Francisco, California.

Received 12 December 2014; revised 9 March 2016; accepted 14 March 2016