Path Analyses of Selection
Joel G. Kingsolver and Douglas W. Schemske

Identifying the targets and causal mechanisms of phenotypic selection in natural populations remains an important challenge for evolutionary biologists. Path analysis is a statistical modeling approach that may aid in meeting this challenge. We describe several types of path model that are relevant to the analysis of selection, and review some recent empirical studies that apply path models to issues in pollination biology, phenotypic integration and selection on morphometric and ontogenetic traits. Path analysis may play two roles in the analysis of selection: first, as an exploratory analysis suggesting possible targets of selection, which are then tested by direct experimentation; and second, as a means of evaluating the relative importance of different causal pathways of selection, once the likely targets of selection have been established.

Which phenotypic traits experience selection in natural populations? What are the biological mechanisms underlying such selection? Despite a half-century of field studies in evolutionary biology, these central questions remain generally unanswered. The primary obstacle in field studies is the identification of the actual targets and mechanisms of selection. These have been demonstrated convincingly in only a few cases. The difficulty lies in distinguishing traits that are the actual targets of selection from correlated traits that do not experience direct selection. The multiple regression techniques championed by Lande, Arnold and Wade have greatly stimulated the analysis and measurement of selection on multiple phenotypic traits, but these techniques have both statistical and biological limitations for the identification of the actual traits under selection and for the examination of causal mechanisms.

Several recent studies have proposed that the use of path analysis may help to address these problems. Path analysis has been used widely in population genetics, but is less familiar to most evolutionary ecologists interested in selection. When applied judiciously, path analysis holds considerable promise for the examination of mechanisms of selection on multiple phenotypic traits.

Path analysis and selection
Path analysis is a method for partitioning the correlations among variables. Sewall Wright first developed path analysis as a means of interpreting quantitatively the causal relationships among a system of correlated variables. Wright distinguished two distinct situations of special relevance to the analysis of selection: systems ordered in time, in which there may be a cause-and-effect relation be-
Path analysis begins with the construction of a path diagram showing the relationships among all variables in the system (Box I), based upon a priori knowledge of such relationships. In this sense, the path diagram represents a hypothesis for evaluation. The path coefficient indicates the direct contribution of one variable to the variation observed in another (see Box I). Path analysis assumes that the relationships among variables are linear, so it is applicable only to the analysis of directional selection.

Because selection results from the covariation of phenotypic traits with relative fitness, applying the techniques of path analysis to selection is straightforward. For example, Lande and Arnold's model of selection may be represented as a path diagram in which each phenotypic trait (X, Y, Z) contributes directly to variation in fitness (W) (Fig. 1a). Thus, in this simple case, the path coefficients directly represent selection, and these coefficients may be estimated by multiple regression analysis. An important advantage of Lande and Arnold's approach is that estimates of the selection coefficients (the selection gradient in Lande and Arnold's terminology) may be directly incorporated into models of phenotypic evolution.

One way to represent causal mechanisms of selection is to incorporate in the path diagram intermediate variables that relate phenotypic traits to fitness (Fig. 1b). In this sense, selection on some trait X results from the effects of X on some intermediate variable P that in turn affects fitness W. As Arnold noted, these relationships can be represented in a path diagram with direct paths from phenotypic traits through intermediate variables to fitness. Because of the equivalence of path coefficients and of selection coefficients in Lande and Arnold's model, \( P_{\text{unm}} \) may be interpreted as the selection coefficient for trait X via its effects on the intermediate variable P — i.e. selection via a specific causal path. Again, these path coefficients may be estimated by a series of multiple regressions (see Box I).

An important difficulty in the analysis of selection is the existence of highly correlated phenotypic traits. For example, if selection actually acts on some unmeasured aspect of phenotype that is correlated with the measured variables, the interpretation of the selection coefficients obtained from regression analyses may be misleading. In keeping with Wright's original application of path analysis, several workers have used path models in which selection may act on unmeasured phenotypic factors, for which the measured traits are imperfect indicators (Fig. 1c). Here, the covariances among the measured variables are modeled in terms of one or more unmeasured factors, with the goal of identifying possible targets of selection that are not measured directly in the system (see Box 2). For systems of morphological traits, relevant unmeasured factors might include overall size and shape allometry. The use of general factors may make the analysis more stable to the addition or omission of measured traits in the path diagram (but see below). This is particularly relevant to systems with large numbers of correlated traits. One limitation of incorporating unmeasured factors into the path diagram is that the estimated selection (path) coefficients cannot be incorporated directly into existing models of phenotypic evolution.

Case studies
Several recent studies have used path analyses of the sort described...
Fig. 1. Path diagrams for three models of phenotypic selection. See Box 1 for terminology. Here, X, Y, Z, P, and Q are measured phenotypic traits. W is fitness. S is an unmeasured phenotypic factor, and U is a residual variable. (a) Path diagram of the Lande-Arnold model of selection on traits X, Y, and Z via intermediate traits P and Q. (c) Path diagram of selection on traits X, Y, and Z incorporating an unmeasured factor S.

in Fig. 1b, c. Here, we briefly review studies applying path analysis to issues in pollination biology, phenotypic integration and selection on size, shape and ontogeny.

Pollination and plant reproductive success

Studies in pollination biology have often hypothesized a relationship between pollinator visitation and plant floral characters. Pollination biologists have recently turned to path analysis to provide a way of estimating how pollinators may contribute to plant reproductive success13. What is most desirable in studies of this kind is simultaneously to examine floral characters, pollinator visitation and plant success. Stanton and colleagues14 estimated the contributions of floral characters, pollinator visitation and postpollination success to variation in seed paternity (a measure of a plant’s contribution to reproductive success through male function) in the mustard Raphanus sativus.

Path analysis revealed substantial variation among pollinators in their effects on seed paternity and in the effects of floral characters on visitation (Fig. 2). Of the four pollinator groups, only small bees had a substantial positive contribution to variation in seed paternity, and honeybees had a large negative effect on seed paternity. The floral characters contributing most to higher visitation by honeybees were pollen-grain size and petal size, while pollen number had the largest positive effect on visitation by small bees. The direct contribution of pollen-grain number to seed paternity was small and negative, while the indirect effect, obtained by summing the products of the path coefficients for all indirect pathways, was rather large and positive. These results indicate how path analysis can be used to reveal the complex relationships between phenotypic traits and reproductive success.

Phenotypic integration and plant performance

Physiological ecologists are interested in the ways that physiological, phenological and morphological traits contribute to overall performance, i.e. phenotypic integration. Path analysis—path diagrams concerning the pattern of phenotypic integration within a population—provides a way of testing various hypotheses.

Evans (PhD thesis, University of Chicago, 1990) used path analysis to determine the contribution of physiological, phenological and allocation traits to among-plant variation in growth and reproduction of Brassica rapa grown in three distinct nutrient environments (Fig. 3). She found that the direct effects of physiology and leaf physiology varied across treatments. For example, specific leaf weight had a positive effect on leaf biomass in all three treatments, but the magnitude of this effect decreased from the low to the high nutrient treatment. Indirect effects of phenological and physiological traits on reproductive biomass were often as large as the direct effects. For example, the indirect effect of specific leaf weight on reproductive biomass in the medium treatment was nearly as large as the direct effect. Differences in the sign of direct and indirect effects were also observed. For example, the direct effect of specific leaf weight on reproductive biomass in the medium treatment was negative, whereas the indirect effect was positive. These results suggest that
the functional relationships among plant traits involve paths of causation that vary across environments.

In a similar study of functional integration, Farris and Lechowicz developed a path model to investigate the factors affecting fruit production in the annual plant *Xanthium strumarium*. Their path model was not based on previous observations or on hypotheses regarding causal paths, but was instead derived from a series of multiple-regression analyses from which only the best predictor variables were used in the path analysis. This approach assures that the final path model will account for the greatest amount of variation in all dependent variables, but is unlikely to reveal the true functional relationships among traits.

**Selection on size and shape**

Morphological traits that reflect the size and shape of organisms have been identified as targets of selection in numerous studies. Because these traits are often highly correlated due to the effects of overall size, it is often difficult to distinguish selection on a particular morphological trait from selection on size. In addition, the large correlations among size-related traits can result in estimated selection coefficients that are not robust to the addition or deletion of other traits to the analysis, especially when multiple regression is used to estimate selection. One approach to these problems is to use factor analysis to compute a general size factor from the measured morphological variables (Box 2). A path analysis is then constructed in which size and the size-adjusted morphological traits influence fitness.

Crespi and Bookstein used a path analysis incorporating size as an unmeasured factor in a reanalysis of Bumpus' data on the survival of house sparrows (*Passer domesticus*) following a winter storm. Crespi and Bookstein estimated the directional selection coefficients on both the unmeasured size factor and on size-adjusted morphological traits. Analyses were conducted on both sexes for all nine traits measured by Bumpus, and on a subset of seven traits (total length and weight were excluded). For comparison, they also conducted a multiple regression analysis following the approach of Lande and Arnold. If significant selection coefficients are only considered, the factor/path analysis and the multiple-regression analysis gave identical results for both sexes for both the seven- and nine-trait data sets. For the seven-trait data set, selection was observed on weight for females, and on total length and weight for males. For the seven-trait data set, no significant selection was detected for females, but selection was observed on wing length for males.

Crespi and Bookstein also performed multiple regression analysis and factor/path analysis on data sets of four and seven morphological traits obtained from pentatomid bugs. This allows one to evaluate whether the selection coefficients obtained from regression analysis are more sensitive (less robust) to the choice of characters. For the four-trait data set, both analyses gave a significant selection coefficient for forewing length only. In the seven-trait data set, regression analysis indicated significant selection on two scutellum traits, but no selection on forewing length. In contrast, the factor/path analysis for this data set indicated significant selection on four characters, including forewing length and two traits that were not found to be significant by this analysis in the original four-trait data set. It is clear that for these data, neither the regression nor the factor/path results are very robust to the addition of traits to the analysis.

**Selection and ontogeny**

Path analysis can also be useful for examining selection acting at different ontogenetic stages for traits that change through the life cycle. Identifying those life stages that make the greatest direct or indirect contribution to variation in fitness is of particular interest.

Maddox and Antonovics used a path analysis of determinants of fitness variation in two *Plantago* species. The number of leaves and the area of the median leaf, determined at each of five different developmental stages, were used as indicators of the unmeasured factor, total leaf area. Three reproductive factors were also estimated, and fitness was an unmeasured factor indicated by seed number and seed mass. The path diagram was based on the hypothesis that leaf area has a direct effect on fitness through seed filling, and an indirect effect through the formation of reproductive structures. They found that the model provided a marginally acceptable fit to the observed data, and that the path diagrams for the two species did not differ significantly. Stages in the middle of the life cycle made the greatest
Based on Evans (PhD thesis, University of Chicago, 1990)