

COMMENTARY

A tale of two methods: putting biology before statistics in the study of phenotypic evolution

J. K. CONNER

Kellogg Biological Station and Department of Plant Biology, Michigan State University, MI, USA

The matrix diagonalization methods advocated by Blows have a number of statistical advantages, especially for the study of natural selection, but these advantages are usually outweighed by the disadvantage that the results are not very biologically interpretable. This lack of interpretability has been the major impediment to the adoption of these methods over the past 15 years.

Blows (2007) makes a number of valuable points that deserve wider attention from evolutionary biologists. The paper advocates a more wholly multivariate approach to understanding adaptive evolution of phenotypic traits as an alternative to the now-standard selection gradient and **G**-matrix analysis. In the standard analysis, multiple regression techniques are used to measure selection on individual traits, correcting for correlations among all traits included in the analysis. With this correction, selection gradients are an extremely useful way of studying adaptive evolution, as they aim to measure direct selection after removing indirect selection caused by phenotypic correlations. More ambitious studies using the standard analysis include both linear terms to estimate the strength of directional selection, and quadratic (squared) terms to estimate the degree of curvature in the fitness function, which might indicate stabilizing and disruptive selection if the maximum or minimum fitness value occurs at intermediate phenotypic values. Still more ambitious studies include the cross-product terms in the regression to measure correlational selection, that is, whether pairs of traits interact to determine fitness in ways not described by the linear and quadratic terms alone.

This accumulation of terms in the regression model is perhaps the biggest problem with standard selection gradient analyses, and thus the biggest advantage of the multivariate methods like canonical analysis (one kind of diagonalization) advocated by Blows. To estimate all the quadratic and correlational selection gradients requires a large number of predictor variables in the analysis with only a modest number of traits. With n traits, the number of predictors in the full model (all linear, quadratic and cross-product terms) is $2n+[n(n-1)]/2$, so with only five

traits there are 20 variables in the model! With canonical analysis, this number is reduced to $2n$, because the number of new variables created is the same as the number of original variables in the analysis, and there is a linear and quadratic term for each. The second term in the formula above represents the correlational selection estimates, which drop out when new axes are created that are orthogonal (i.e. uncorrelated). The main reason that correlational terms are often not included in selection gradient analyses (Kingsolver *et al.*, 2001) is not their difficulty in interpretation (more on this below), but rather that people rarely have the large sample sizes needed to fit so many variables in their regression models. Thus, a method that allows people to explore the multivariate nature of selection with the more modest sample sizes of most studies is a real advantage.

Therefore, the multivariate methods that Blows champions will prove to be useful additions to the evolutionary biologist's toolbox in some cases, providing new insight not available from standard selection gradient analysis. However, it is not yet clear how often this will be the case, and these methods will not replace the now-familiar selection gradient and **G**-matrix analysis. This is not just due to tradition; there are good reasons why the selection gradients and **G**-matrix framework has become one of the most widely adopted in all of evolutionary biology (allowing meta-analyses like that of Kingsolver *et al.*, 2001): the approach can be undertaken by empiricists with modest quantitative skills, and more importantly it produces results that are eminently interpretable, leading to robust conclusions about the nature of the adaptive process. Selection gradients are often presented in standardized form, that is, in standard deviation units, making them readily comparable across different traits and studies. This makes it possible to make comparisons between, for example, the strength of direct selection on tarsus length in different bird species, or to compare the strength of selection on morphological vs. life-history traits (Kingsolver *et al.*, 2001). Blows performs a good service by reminding us of the potential pitfalls of selection gradient analysis; however, as long as practitioners are aware of these and exercise due caution in analysis and interpretation, they are almost never a problem.

The diagonalization techniques advocated by Blows will not replace selection gradient and **G**-matrix analyses because they have a fundamental problem, which is pointed out at the very end of Blows' article: the difficulty of biological interpretation. This is the Achilles' heel of these methods, because the central goal of selection and **G** matrix analyses has to be biological understanding. Once new axes through multivariate space are created for statistical, rather than biological reasons (for example, to maximize variance explained and to make axes orthogonal), then it is often very difficult to understand what these axes mean in terms of adaptation or organismal function (Endler, 1986, p. 192;

Correspondence: Jeffrey K. Conner, Kellogg Biological Station, 3700 East Gull Lake Drive, Hickory Corners, MI 49060, USA.
Tel.: 269-671-2269; fax: 269-671-2104; e-mail: connerj@msu.edu

Phillips & Arnold, 1989; Simms, 1990). Can we say that these new axes are adaptations, that is, traits that allow the organism to deal with some challenge in its environment, just because one of them represents the direction of maximal multivariate correlation with fitness? I think not. Sometimes the canonical axes will be similar to the original traits (e.g. Simms, 1990), which means that the original traits were likely those under selection and the canonical analysis adds little. I think this will most often be the case, because our knowledge of the organisms we study will usually be a good guide to which traits are adaptive. When it is not, and the canonical axes under strongest selection are composites of the original traits, then this is definitely worth further study, but it may be easiest to interpret this situation using traditional directional, quadratic and correlational selection gradients on the original traits.

In the example in Blows' Fig. 1 (data from Brooks *et al.*, 2005), there is significant stabilizing selection on the fourth and fifth eigenvectors (the last two columns of the \mathbf{M} matrix), which are in turn complex mixtures of the five call characteristics manipulated in the study. What does this tell us about the biology of the crickets? The power issue mentioned above was not a major problem in this study, as five of the selection gradients in the analysis of the original traits were significant (Brooks *et al.*, 2005), including two negative quadratic terms. It is difficult to fully interpret these gradients from the original study without graphical depictions of the fitness functions and surfaces, but it appears there is stabilizing selection for intermediate dominant frequency of the call, as well as directional selection for shorter intercall duration that becomes weaker with shorter duration calls (i.e. there are diminishing fitness returns, as both the linear and quadratic terms were significantly negative). There is also positive correlational selection between intercall duration and trill number, meaning that calls with short duration and few trills, as well as calls with long duration and many trills, are favoured by females over the opposite combinations. These selection gradients on the original traits are more readily interpretable than the selection on the eigenvectors. This is the main reason why selection gradient analysis has been widely adopted and canonical analysis has not, despite the latter having been introduced to evolutionary biologists over 15 years ago (Phillips & Arnold, 1989; Simms, 1990).

Thus, I disagree with Blows' assertions that estimates of selection on the new traits created by canonical analysis are more readily interpretable than traditional selection gradients, especially nonlinear or correlational gradients. The biological meaning of stabilizing and disruptive selection are clear: that organisms with intermediate phenotypic values have higher or lower fitness than either extreme respectively. It is also clear what correlational selection on a pair of traits means – that certain combinations of traits have higher fitness than

other combinations, and furthermore that the bivariate fitness function has a ridge or saddle shape, and is not just a tilted plane (i.e. there is not just *one* combination of traits that has highest fitness). Obviously correlational selection between two traits is already far more difficult to interpret and explain than selection on single traits in isolation. As I have learned from many years of teaching this material and trying to explain my floral evolution work to students and colleagues, 'simple' correlational selection, involving a ridge or a saddle-shaped bivariate fitness function, is initially hard to understand. A multidimensional surface with 'peaks, bowls and saddles' is not 'more readily interpretable' as Blows suggests; indeed, trying to visualize the biological meaning of anything that contains more than three dimensions is something that most people simply cannot do, and a complex fitness surface in only three dimensions can be very hard to interpret as well.

This discussion raises a difficult and fundamental issue, that is, how is a phenotypic trait defined? It is useful to think of a hierarchy of traits, with the lowest level being proteins (or perhaps mRNA?), which give rise to metabolic pathways and higher level physiological traits, which in turn affect morphological traits, and finally all these traits interact to produce behaviour and life history traits. As we move up this hierarchy the traits become more complex, as they are affected by more traits at lower levels in the hierarchy, which in turn means that they are affected by more gene loci and more environmental factors. The multivariate constructs created by PCA or canonical analysis are not simply higher-level traits in the hierarchy, because they are not affected or determined by the original traits in the analysis through biological processes. Nor are they defined using biological information, but rather to maximize statistical tractability. It is true that sometimes we may be misled in defining a trait by our human perceptions of what is important to an organism, but biologists are more often led into fuzzy thinking when they try to interpret any kind of multivariate statistical construct instead of the traits that they can directly measure and determine how they function. In their well-known Spandrels paper, Gould & Lewontin (1979) say the chin is not a 'thing' (trait), but rather 'a product of interaction between two growth fields'. However, I would argue that if a feature of the chin that can be quantified has a direct effect on fitness, that is, it has a function and is under direct phenotypic selection, then it certainly is an adaptive trait. It is possible that canonical analysis could lead to a better definition of a particular adaptive trait, and thus to a better understanding of how an organism functions in its environment, but I am not aware of an example of this to date.

Blows is certainly correct that more of us need a better knowledge of linear and matrix algebra. Part of the problem here is a lack of courses available (at least at many US universities) that are designed to present this

material to nonmath or stats majors. There are precious few, if any, courses in matrix algebra that are at the right level and have the right coverage for a biology graduate student. There is no doubt in my mind that a better understanding of what eigenvalues and eigenvectors represent (which does not require a full semester course) would enable evolutionary biologists to have a deeper understanding of phenotypic evolution. However, this is not nearly as important as understanding the selection acting on, and the genetics underlying, real functional traits that are defined based on a solid biological understanding of the organisms in their natural environment and not on purely statistical grounds.

Acknowledgments

I thank Frances Knapczyk, Cindy Mills, Neil Patel, and Heather Sahli for discussion, and Meghan Duffy, Buffy Silverman, and Jay Sobel for discussion and comments on an earlier draft of this paper. I was supported by the National Science Foundation under DEB 0108354 and by the Cooperative State Research, Education, and Extension Service, U.S. Department of Agriculture, under

Agreement No. 2002*35320-11538. This is KBS Contribution no. 1273.

References

- Blows, M.W. 2007. A tale of two matrices: multivariate approaches in evolutionary biology. *J. Evol. Biol.* **20**: 1–8.
- Brooks, R., Hunt, J., Blows, M.W., Smith, M.J., Bussiere, L.F. & Jennions, M.D. 2005. Experimental evidence for multivariate stabilizing sexual selection. *Evolution* **59**: 871–880.
- Endler, J.A. 1986. *Natural Selection in the Wild*. Princeton University Press, Princeton, NJ.
- Gould, S.J. & Lewontin, R.C. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**: 581–598.
- Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C.E., Hoang, A., Gibert, P. & Beerli, P. 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* **157**: 245–261.
- Phillips, P.C. & Arnold, S.J. 1989. Visualizing multivariate selection. *Evolution* **43**: 1209–1222.
- Simms, E.L. 1990. Examining selection on the multivariate phenotype: plant resistance to herbivores. *Evolution* **44**: 1177–1188.