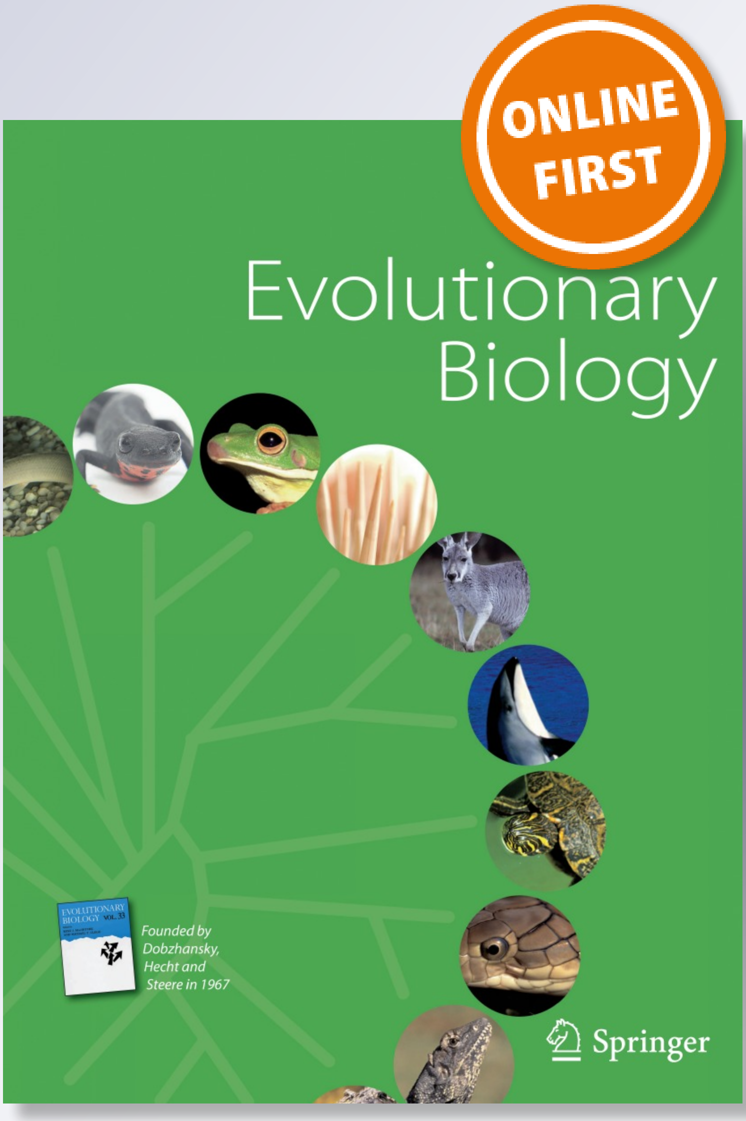
Many-to-Many Mapping of Phenotype to Performance: An Extension of the F-Matrix for Studying Functional Complexity

Philip J. Bergmann & Eric J. McElroy

Evolutionary Biology
Evolutionary Biology

ISSN 0071-3260

Evol Biol
DOI 10.1007/s11692-014-9288-1



Evol Biol
DOI 10.1007/s11692-014-9288-1

RESEARCH ARTICLE

Many-to-Many Mapping of Phenotype to Performance: An Extension of the F-Matrix for Studying Functional Complexity

Philip J. Bergmann · Eric J. McElroy

Received: 16 December 2013 / Accepted: 25 June 2014

Abstract Performance capacity influences ecology, behavior and fitness, and is determined by the underlying phenotype. The phenotype-performance relationship can influence the evolutionary trajectory of an organism. Several types of phenotype-performance relationships have been described, including one-to-one relationships between a single phenotypic trait and performance measure, trade-offs and facilitations between a phenotypic trait and multiple performance measures, and redundancies between multiple phenotypic traits and a single performance measure. The F-matrix is an intraspecific matrix of measures of statistical association between phenotype and performance that is used to quantify these relationships. We extend the F-matrix in two ways. First, we use the F-matrix to describe how the different phenotype-performance relationships occur simultaneously and interact in functional systems, a phenomenon we call many-to-many mapping. Second, we develop methods to compare F-matrices among species and compare phenotype-performance relationships at microevolutionary and macroevolutionary levels. We demonstrate the expanded F-matrix approach with a dataset of eight phrynosomatine lizard species, including six phenotypic traits and two measures of locomotor performance. Our results suggest that all types of relationships occur in this system and that phenotypic traits involved in trade-offs are more functionally constrained and tend evolve slower interspecifically than those involved in facilitations or one-to-one relationships.

Keywords Facilitation · Functional morphology · Lizard · Redundancy · Trade-offs · Constraint · Locomotion

Introduction

A central tenet of the study of natural selection is that form and function are closely coupled. Historically, correlations between organismal phenotypes and their environments have been viewed as strong evidence of adaptation (Darwin 1859). To study these correlations from an intraspecific perspective, Arnold (1983) argued that performance at a given task is an important link between phenotype and fitness. Performance can be defined as how well an organism does a certain task, and it is now widely thought that selection acts primarily on fitness-related performance and only secondarily on the underlying phenotype (e.g., Garland and Carter 1994; Calsbeek 2008; Irschick et al. 2008; Langerhans and Reznick 2010). However, the concept of fitness is not directly transferable to the macroevolutionary level, and so the phenotype → performance → fitness paradigm has been expanded for this purpose, relating phenotype to performance to habitat use, as species with different task performance levels tend to use different habitats where they perform the best (Garland and Losos 1994; Aerts et al. 2000; Van Damme

---

Electronic supplementary material The online version of this article (doi:10.1007/s11692-014-9288-1) contains supplementary material, which is available to authorized users.

P. J. Bergmann (✉)
Department of Biology, Clark University, 950 Main Street, Worcester, MA 01610, USA
e-mail: pbergmann@clarku.edu

E. J. McElroy
Department of Biology, College of Charleston, 58 Coming Street, Charleston, SC 29424, USA
e-mail: mcelroye@cofc.edu

et al. 2003; Bergmann and Irschick 2010). An implicit assumption of both approaches has been a direct, or one-to-one, relationship between phenotype and performance. For example, it has long been known that bill size correlates positively with bite force, giving the ability to crack larger seeds in Darwin's finches, and that the availability of different sized seeds exerts a selective pressure on bill size (Gibbs and Grant 1987; Herrel et al. 2005).

However, in many instances phenotype and performance are weakly or not related even though the type of performance measured is fitness-relevant and the phenotypic traits are clearly of functional importance (e.g., Wiens and Rotenberry 1980; McNally and Doolan 1986; Wiens 1991; Zaaf and Van Damme 2001; Alfaro et al. 2005). These findings suggest that either organismal phenotypes are imperfectly adapted to multiple functional demands (Lewontin 1978), or that direct relationships do not fully describe how biological systems function (e.g., Irschick and Losos 1998). To resolve these central issues to understanding trait evolution, we must study phenotype-performance relationships with a focus on complexity, where the phenotype is composed of multiple parts and must be able to carry out multiple tasks (Walker 2010). Under these circumstances, complex relationships between phenotype and performance might weaken the strength of direct, one-to-one relationships between the two.

Recent work has identified several phenomena to explain the complexity of phenotype-performance relationships. These are functional trade-offs, facilitation, and redundancy. Functional trade-offs are a form of constraint where there is an inverse relationship between two performance measures, which place conflicting demands on a phenotypic trait, resulting in a situation where the performance measures cannot be maximized simultaneously (Garland and Losos 1994; Van Damme et al. 2003; Walker 2007, 2010; Langerhans 2009; Holzman et al. 2011b; Vanhooydonck et al. 2011). In the Darwin's finch example, bill size is involved in a trade-off between song performance and bite force (Herrel et al. 2009). Smaller bills are able to produce sounds with a wider range of frequencies than larger bills, which is important in attracting a mate, but are unable to generate high bite forces (Podos 2001; Huber and Podos 2006).

Facilitation is the converse of trade-offs because it places similar demands on the phenotype (Walker 2007), whereby a given change in a phenotypic trait improves multiple aspects of performance (Table 1). Facilitation has been documented in *Anolis* lizards, where velocity and acceleration are both positively correlated with hindlimb muscle mass (Vanhooydonck et al. 2006b). Both trade-offs and facilitation are forms of one-to-many relationships (Walker 2010), where a single phenotypic trait influences multiple aspects of performance (Table 1).

Table 1 Characteristics of various phenotype-performance relationships

	# of Performance measures affected	
	1	>1
# Phenotypic Traits Involved		
1	*One-to-one relationship* A phenotypic trait will directly affect one performance measure	*One-to-many relationship (multitasking)* A phenotypic trait affects multiple performance measures *Trade-off* Performance measures are affected in opposite ways *Facilitation* Performance measures are affected in the same way
>1	*Many-to-one relationship* Multiple phenotypic traits affect a performance measure *Redundancy* increases with the number of phenotypic traits related to performance	*Many-to-many relationship* Multiple phenotypic traits variously affect multiple performance measures Some phenotypic traits affect only one measure of performance, others multiple One-to-one, redundancy, trade-offs, and facilitation all happen simultaneously

Redundancy also results in complex relationships between phenotype and performance because a range of phenotypes can yield the same level of performance (Table 1) (Alfaro et al. 2005). Such many-to-one mapping of phenotype to performance is ubiquitous, arising in any situation where a phenotype consists of multiple traits that influence a single performance measure (Alfaro et al. 2005; Wainwright et al. 2005). Redundancy can weaken a trade-off, as different aspects of a complex phenotype can be co-opted for different aspects of performance (Vanhooydonck et al. 2006a; Wainwright 2007; Walker 2007; Holzman et al. 2011b). Evidence for redundancy has been documented in a diversity of organisms (e.g., Alfaro et al. 2005; Vanhooydonck et al. 2006a; Strobbe et al. 2009). It may also be at play in the Darwin's finches because the depths and widths of both the head and bill all influence bite force (Herrel et al. 2005).

To fully understand complex phenotype-performance relationships it is imperative to integrate the study of one-to-one relationships, trade-offs, facilitation, and redundancy into a single quantitative framework because they can all be at play simultaneously in a functional system. To do this, we propose the concept of "many-to-many mapping" of phenotype to performance. Many-to-many mapping considers multiple phenotypic traits that influence multiple measures of performance, allowing for a comprehensive consideration of how the different types of phenotype-performance relationships shape the functional system (Table 1). In a complex system, some phenotypic traits may underlie trade-offs between performance measures, while others may underlie facilitation, and others still can influence one performance measure but not others. Increased complexity acts to mitigate trade-offs (Holzman et al. 2011b), and the degree to which it does so is determined by the degree of redundancy (Alfaro et al. 2005), facilitation, and one-to-one relationships between phenotype and performance. Quantifying the relationships between multiple phenotypic traits and multiple aspects of performance allows one to assess how phenotypic traits constrain performance, and how performance functionally constrains phenotypic traits (Walker 2007, 2010).

The **F**-matrix is a general model for relating multiple phenotypic traits to multiple performance measures at the intraspecific level (Ghalambor et al. 2003; Walker 2007). As such, it is ideally suited for studying many-to-many mapping of phenotype to performance. The **F**-matrix also has the strength of quantifying functional constraints in a phenotype-performance system (Walker 2007). However, the **F**-matrix approach remains largely theoretical, as it has only been applied to predation survival and climbing ability in a goby fish (Blob et al. 2010), and to feeding in sunfishes (Holzman et al. 2011b), likely because it lacks a practical implementation and the hypothetical example provided by Walker (2007) considered only two phenotypic traits. Another current limitation of the **F**-matrix is that it has not been expanded from the intraspecific to the interspecific, or macroevolutionary, level.

Here, we use a dataset of locomotor performance and morphometric data collected from eight species of Phrynosomatine lizards to (1) illustrate how the **F**-matrix can be applied to characterize many-to-many mapping of phenotype to performance, and (2) extend the **F**-matrix to the interspecific level to characterize the macroevolution of phenotype-performance relationships. We also implement all of the associated analyses into a set of easy-to-use functions in the R computing environment (R Development Core Team 2012). Locomotion in phrynosomatines is an ideal system for studying many-to-many mapping because it is complex, consisting of many integrated morphological traits, and because locomotion is ecologically-relevant (Irschick and Jayne 1998; Toro et al. 2003; Vanhooydonck et al. 2006b). Phrynosomatine lizards in particular are morphologically and ecologically diverse, and their locomotion has been studied extensively (Miles 1994; Reilly 1995; Bonine and Garland 1999; Irschick and Jayne 1999; Bonine et al. 2001; Van Damme and Vanhooydonck 2001; Scales et al. 2009; Bergmann and Irschick 2010; Higham et al. 2011b; McElroy et al. 2012).

Materials and Methods

Phrynosomatine Locomotor Dataset

Phrynosomatine lizards consist of three clades: the horned lizards (including the genus *Phrynosoma*), the sand lizards (*Callisaurus, Cophosaurus,* and *Holbrookia*), and the fence lizards (*Sceloporus, Urosaurus* and *Uta*) (Reeder and Wiens 1996). The horned lizards are slow-moving, terrestrial, broad-bodied ant specialists (Sherbrooke 2003; Bergmann et al. 2009); the sand lizards are fast, terrestrial, gracile sprinters (Irschick and Jayne 1998); and the fence lizards are semi-arboreal with an intermediate body shape (Herrel et al. 2002).

Our dataset includes eight species, some from each subclade: *Callisaurus draconoides, Cophosaurus texanus, Holbrookia maculata, Phrynosoma mcallii, P. platyrhinos, Sceloporus jarrovii, S. virgatus,* and *Uta stansburiana*. The dataset includes, six morphometric phenotypic traits [snout-vent length (SVL), and relative length of the front limb (rFLL), thigh (rThigh), crus (rCrus), foot (rFoot), and the fourth hind digit (rDigit)], and two performance measures (maximal velocity and acceleration on a horizontal surface). We also used a phylogeny relating these species to do phylogenetically-informed analyses (Online Resource 1).

We calculated the relative lengths of our morphometric traits to account for the confounding affects of body size and to decrease collinearity in a phylogenetically-informed way, following the approach of Blomberg et al. (2003) as implemented in the phytools package (Revell 2009) in R (R Development Core Team 2012). To do this for individual lizards, we used our species-level phylogeny, but with individuals as taxa, with a polytomy for each species. We used this approach on log-transformed variables, using log(SVL) as the proxy for size. Our approach ensured that data were maintained in a single morphospace and were, therefore, comparable among species, which separate regressions for each species would not have accomplished.

To obtain maximum velocity and acceleration data, we fitted quintic splines to displacement data obtained from digitized videos taken in the field. The velocity data were previously published and calculated by taking the maximum of the first derivative of the quintic splines (Bergmann and Irschick 2010). As noted in the original publication, maximal velocities for some of the sand lizards are underestimated due to the length of the racetrack used, and so this represents a caveat of the current study as well, but does not influence our ability to demonstrate many-to-many mapping and the F-array approach. The acceleration data were calculated as the maximum value of the second derivative of our quintic spline (Walker 1998), and are originally published here. In order to obtain reliable

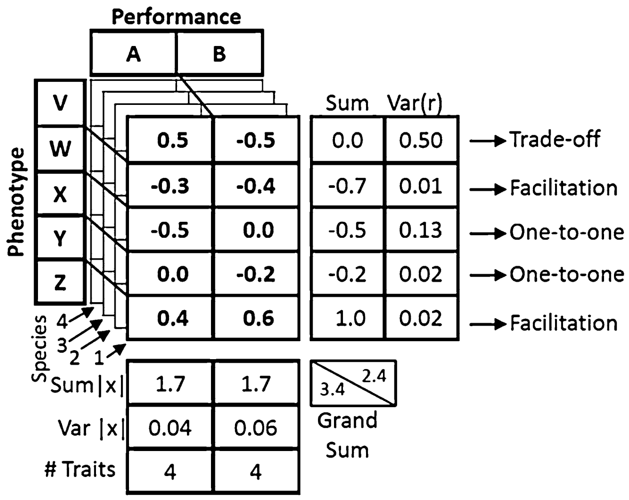


Fig. 1 A diagram of an **F**-array composed of **F**-matrices for four species, two performance measures, and five phenotypic traits. Numbers in the cells of the top **F**-matrix represent a hypothetical example to illustrate how an **F**-matrix works. Row sums indicate the functional constraint of each aspect of the phenotype. The column (*before backslash*) and row (*after backslash*) grand sums are overall measures of functional constraint for performance and phenotypic traits, respectively. Variances are measures of how different the influences of different variables are. The number of traits with a relationship to each performance measure and the column sums are measures of potential complexity and redundancy. Types of phenotype-performance relationships are mapped to the *right of the diagram*

acceleration data, we eliminated any trial in which the lizard was not stationary at the beginning of the video because peak acceleration occurs during the first two strides of movement (Vanhooydonck et al. 2006b; McElroy and McBrayer 2010). Our final dataset thus included six phenotypic traits, two performance measures, and eight species (see Table 3 for intraspecific sample sizes), and is archived at datadryad.org.

Construction and Interpretation of **F**-matrices

The **F**-matrix contains the relationships between multiple phenotypic traits (shown in rows) and multiple performance measures (shown in the columns, see hypothetical example in Fig. 1). The **F**-matrix for a species is filled with standardized partial regression coefficients from a multiple regression of phenotypic traits on each performance measure (Ghalambor et al. 2003; Walker 2007). To accomplish this for each species, we scaled all of our phenotypic and performance variables to mean zero and standard deviation of one and ran multiple regression on each performance measure (dependent variable) with log(SVL), rFLL, rThigh, rCrus, rFoot, and rDigit as predictors in R (R Development Core Team 2012). We also adopted the same approach, but using phylogenetically independent contrasts (Felsenstein 1985), calculated in Mesquite v2.75 (Maddison and Maddison 2011), to construct an interspecific F-matrix. We positivized contrasts for each performance measure, adjusting the signs for the phenotypic contrasts appropriately, and ensured that contrasts were adequately standardized by the phylogenetic branch lengths using correlation (Garland et al. 1992). We used effect sizes to determine which values to consider interpretable, using effects of 0.5 (medium to large) as the cut-off (Cohen 1988). This was necessary due to decreased power of the multiple regression due to collinearity, which acts to increase parameter variance but not bias the parameter estimates themselves (Freckleton 2011; Dormann et al. 2013). Collinearity was present in our dataset despite adjustment of phenotypic traits for body size and phylogeny, with tolerances for some phenotypic traits <0.1, and correlations among several phenotypic traits >0.7 (Online Resource 3). Use of effect sizes is also desirable because in interpreting an F-matrix, it the magnitude rather than the significance associated with null hypothesis tests (Moran 2003) that is important.

For each **F**-matrix, we also calculated the sum and variance of elements for each row, the sum and variance of the absolute values of elements for each column, the grand sums of absolute column and absolute row sums, as well as $\mathbf{FF}^{\mathrm{T}}$ and $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ matrices, using basic algebra. These metrics are important for a full interpretation of the **F**-matrix, and allow us to determine whether each phenotypic trait is involved in a trade-off, facilitation, or a one-to-one relationship (see Fig. 1 for a hypothetical example). **F**-matrix row sums give a measure of the degree of functional constraint placed on each phenotypic trait by the performance measures (Walker 2007). For example, traits involved in trade-offs tend to have low sums, indicating high degrees of constraint (trait V in Fig. 1). Those traits involved in one-to-one relationships have weaker functional constraints (traits X and Y in Fig. 1), and those involved in facilitation have the weakest functional constraint acting on them (traits W and Z in Fig. 1). Row variances are estimates of how different the effects of the performance measures are on each phenotypic trait, and trade-offs generally result in the greatest row variance (Fig. 1). The row sums and variances have also been interpreted as measures of phenotype evolvability and variability in selection coefficient, respectively (Walker 2007), but this requires consideration of the **G**-matrix because phenotypic traits with low underlying genetic variability will not be able to evolve readily (Schluter 1996; Holzman et al. 2011b), irrespective of the functional constraints on the phenotype.

F-matrix columns provide an understanding of how performance measures are influenced by the suite of phenotypic traits considered. The number of phenotypic traits influencing a performance measure and the sum of absolute column values provide measures of the complexity of the system and the potential for redundancy. The variance of absolute column values is a measure of how different the effects of the phenotypic traits are on a given performance measure. We calculated column sums and variances using absolute values because having relatively longer or shorter body parts is not in itself dis/advantageous to the animal, but rather depends on how a trait influences performance. Hence, performance A and B in Fig. 1 are parts of equally complex systems. By combining column sums and variances with information on trade-offs, facilitations and one-to-one relationships, one can predict which traits will be most able to evolve in response to selection for an increase in a performance measure, given functional constraints. The grand sums of absolute row and column sums, provide two additional numbers that give overall measures constraint on phenotypic traits from performance, and on performance from phenotypic traits, respectively. These measures of functional constraint in the system are important because they consider all phenotypic traits and performance measures simultaneously (Fig. 1). The higher the magnitude of these values, the less constrained the system is.

The $\mathbf{FF}^{\mathbf{T}}$ and $\mathbf{F}^{\mathbf{T}}\mathbf{F}$ matrices are useful for quantifying functional integration among q phenotypic traits or p performance measures, respectively (Walker 2007). The $\mathbf{FF}^{\mathbf{T}}$ is a symmetrical $q \times q$ matrix, where the diagonal values are proportional to the square of row sums of the **F**-matrix, providing a measure of functional constraint for each trait, and the off-diagonal values indicate the functional integration among phenotypic traits, given performance constraints on the phenotype. The $\mathbf{F}^{\mathbf{T}}\mathbf{F}$ matrix is a symmetrical $p \times p$ matrix, where the diagonal gives a measure of phenotypic constraint on performance measures, and the off-diagonal values indicate whether performance measures trade-off (if negative) or facilitate one another (if positive), given phenotypic constraints on performance.

Construction and Interpretation of the Interspecific **F**-array

One limitation of the **F**-matrix approach is that it is designed for intraspecific analyses, and therefore, does not consider macroevolutionary changes. To address this, we suggest first compiling an **F**-matrix for each species, and then assembling multiple **F**-matrices into a 3-D **F**-array with phenotypic traits, performance measures, and species as the three axes (Fig. 1). An **F**-array can be analyzed in conjunction with a phylogeny using comparative statistics (Felsenstein 1985; Revell and Collar 2009) to study how the **F**-matrix has evolved, and to compare intraspecific and interspecific patterns. This is a powerful approach because intraspecific microevolutionary patterns can be compared directly to interspecific macroevolutionary patterns.

We assembled our species **F**-matrices into an **F**-array, re-divided it into a species by phenotype matrix for each of our two performance measures, and calculated one phenotypic evolutionary VCV (**eVCV**) matrix for each performance measure using our phylogeny and the ratematrix function in the R package Geiger (Harmon et al. 2003). We then used a Mantel test that includes the diagonal to compare **eVCV** matrices for maximal velocity and acceleration to test whether the influence of phenotypic traits on the two performance measures has evolved a similar structure in the clade. Including the diagonals of the **eVCV**s was necessary because the values in the diagonal of these matrices are informative, unlike the diagonal of a dissimilarity matrix, which is the usual input for a Mantel test (Bonnett and Van de Peer 2002). Since the diagonal contains different information from the off-diagonals (rates of trait evolution as opposed to strength of trait co-evolution), we randomized the diagonal elements separately from the off-diagonals (Mantel 1967).

We also used the **F**-array to compare phenotypic trait evolution under functional constraint between intraspecific and interspecific levels. We did this using Mantel tests, modified to compare the entire matrix (Mantel 1967). We used these tests to compare intraspecific **F**-matrices in a pairwise manner, and to compare each intraspecific **F**-matrix to the interspecific **F**-matrix that we calculated using phylogenetically independent contrasts (similar to Holzman et al. 2011b). For these tests, we compared all elements of the **F**-matrices because they are all informative, containing standardized partial regression coefficients between phenotypic traits and performance measures. Here, a significant Mantel test indicates that phenotype-performance relationships are similar between species, and that species-level relationships coincide with clade-level relationships.

For all of these Mantel tests, we converted the included cells of the two matrices being compared to vectors, and randomized one of the vectors 10,000 times, calculating the Pearson correlation between vectorized matrices for the unrandomized data and each randomized dataset. We then calculated the p value as the twice the number of randomized correlation coefficients that were more extreme than the unrandomized coefficient, divided by 10,000, for a two-tailed test. We implemented these tests in an R function that we wrote (Online Resources 5 and 6). We used the Benjamini-Hochberg (BH) (Benjamini and Hochberg 1995) technique for correcting for multiple comparisons. The BH technique is superior at controlling type I error rates to the Bonferroni and sequential Bonferroni approaches because it takes false discovery rate into account (Williams et al. 1999).

Table 2 Intraspecific **F**-matrix for *Cophosaurus texanus* (A), and the interspecific **F**-matrix of phylogenetically independent contrasts with cells filled with standardized partial multiple regression coefficients

A. *Cophosaurus texanus*

	Vel	Accel	Sum	Var
SVL	**0.65**	0.07	0.72	0.17
rFLL	-0.32	**-1.07**	-1.38	0.28
rThigh	0.20	**-1.99**	-1.79	2.40
rCrus	0.46	**3.73**	4.19	5.37
rFoot	**-0.76**	**-1.59**	-2.36	0.34
rDigit	**-0.57**	**1.19**	0.63	1.55
Sum|x|	2.95	9.65	**12.57\11.06**	
var|x|	0.04	1.50		
# traits	3	5		

B. Interspecific F-matrix

	Vel	Accel	Sum	Var	Sigma
SVL	**1.29**	**1.60**	2.89	0.05	0.05
rFLL	**-1.34**	**-1.79**	-3.13	0.10	0.01
rThigh	**-2.91**	**1.82**	-1.09	11.14	0.01
rCrus	**2.47**	-0.39	2.07	4.10	0.02
rFoot	**1.24**	**3.09**	4.33	1.70	0.02
rDigit	0.15	**-2.82**	-2.67	4.42	0.08
Sum|x|	9.39	11.51	**20.90\16.18**		
var|x|	0.97	0.93			
# traits	5	5			

For each **F**-matrix, the row sums and variances, column sums and variances of the absolute values, and grand sums of absolute values for columns (before backslash) and rows (after backslash) are provided. **F**-matrix cells with values of medium to large effect size (>0.5) are indicated by shading, with positive values in grey and negative values in black. Brownian motion model sigma parameters for each phenotypic trait are also presented for the interspecific **F**-matrix

Results

Many-to-Many Mapping in Lizard Locomotion: An examination of Two F-Matrices

We show that for the phrynosomatine locomotor system, species differ in their phenotype-performance relationships. Our analyses show that one-to-one relationships, trade-offs, facilitations, and redundancy are common and occur simultaneously in many of the species. Therefore, the **F**-matrix can be used for studying many-to-many mapping. Although we analyzed eight species, we provide a detailed interpretation of **F**-matrices for one of them and the interspecific **F**-matrix here so as to demonstrate the approach on real data. **F**-matrices for all species are available in Online Resource 2.

Table 2 presents the **F**-matrix for *Cophosaurus texanus* and the interspecific **F**-matrix. Both **F**-matrices document a combination of trade-offs, facilitations, and one-to-one relationships between phenotype and performance. In *Co. texanus* rFoot is involved in a facilitation, where individuals with shorter feet have higher velocities and accelerations, while rDigit is involved in a trade-off where relatively longer digits result in higher accelerations but lower velocities. All other phenotypic traits are involved in one-to-one relationships with either velocity or acceleration.

Interspecifically, SVL, rFLL, and rFoot are involved in facilitations, where species that evolve longer bodies, relatively shorter front limbs, and relatively longer feet evolve higher velocities and accelerations. Additionally, species that evolve relatively longer thighs evolve lower velocities but higher accelerations. Again, the other phenotypic traits have one-to-one relationships with one of the performance measures.

Phenotypic traits show different amounts of functional constraint, as represented by row sums, with SVL and the relative length of the digit being more constrained than other traits in *Co. texanus*, and the relative length of the thigh being most constrained interspecifically. The grand row sum is high for both *Co. texanus* and the interspecific **F**-matrix (Table 2), when compared to other species, except *Callisaurus draconoides* (Table 3), suggesting lower overall levels of constraint in *Co. texanus* and the clade in general.

Focusing on the performance measures velocity has lower phenotypic redundancy than acceleration, being influenced by fewer traits and having a lower grand column sum in *Co. texanus* (Table 2). Therefore, selection for higher acceleration could act more effectively and on more phenotypic traits than selection on velocity due to weaker functional constraints. Both aspects of performance are influenced by five phenotypic traits and have similar column sums interspecifically, resulting in a complex system

Table 3 **F**-matrix row sums for eight species of phrynosomatine lizards

Species	*C.dra.*	*C.tex.*	*H.mac.*	*P.mca.*	*P.pla.*	*S.jar.*	*S.vir.*	*U.sta.*
n	8	8	13	14	12	11	14	14
SVL	**-2.94**	0.72	-0.02	-0.77	0.23	0.25	0.44	0.14
rFLL	**-1.66**	-1.38	0.89	0.65	-0.74	-0.13	-0.56	**1.77**
rThigh	1.16	**-1.79**	-0.26	0.19	0.59	-0.02	-0.10	-0.11
rCrus	**4.10**	**4.19**	0.43	-0.21	-0.43	-0.28	-0.69	-1.18
rFoot	**-3.12**	**-2.36**	-0.65	-0.60	0.02	0.22	-0.66	-0.70
rDigit	-0.56	0.63	**-1.89**	-0.29	1.23	-0.84	0.63	-0.39
Grand	13.53	11.06	4.14	2.70	3.23	1.73	3.08	4.29

Darkness of shading corresponds to magnitude of each sum—darker values have weaker functional constraints on their evolution

The grand sum of absolute values of row sums is also provided for each species

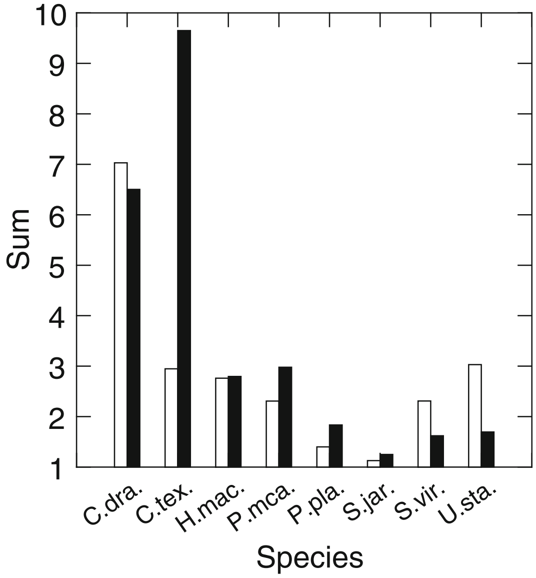


Fig. 2 F-matrix column sums of absolute values of the cells of the **F**-matrix for velocity (*white bars*) and acceleration (*black bars*), indicating the amount of redundancy governing each performance measure

with high redundancy, and therefore, weak functional constraints relative to what is seen in *Co. texanus* (Table 2). Overall functional constraints on the performance measures are weaker in the fast-running sand lizards *Co. texanus* and *Ca. draconoides* than in the other species that we sampled (Fig. 2, Online Resource 2).

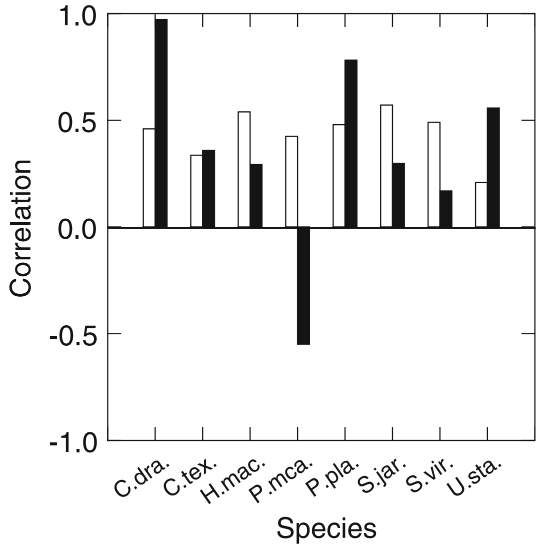


Fig. 3 Correlations between velocity and acceleration for each species. *White bars* represent correlations between raw velocity and acceleration for each species. *Black bars* represent correlations obtained from $\mathbf{F}^{T}\mathbf{F}$ matrices for each species; these take phenotypic constraints on performance into account. *Positive values* indicate facilitations, *negative values* indicate trade-offs. The greater the magnitude, the stronger the facilitation or trade-off

Accounting for functional relationships between phenotype and performance can change the relationship between performance measures depending on what relationships different phenotypic traits are involved in. Although standard Pearson correlations between velocity and acceleration are positive in all species (Fig. 3, white bars), the relationships change when the constraints the phenotype are taken into account (as calculated from $\mathbf{F}^{T}$-$\mathbf{F}$ matrices: Fig. 3, black bars). Phenotypic traits involved in facilitations in *Ca. draconoides*, *Phrynosoma platyrhinos*, and *Uta stansburiana* strengthen the relationships between velocity and acceleration. However, when trade-offs play a more important role in the functional system (as in *Holbrookia maculata, Sceloporus jarrovii,* and *S. virgatus*—Online Resource 2), the strength of the relationship between performance measures decreases. This is taken to an extreme in *P. mcallii,* where the direction of relationship between performance measures is reversed on account of a strong trade-off associated with rCrus (Fig. 3, Online Resource 2).

Both the *Co. texanus* and interspecific **F**-matrices illustrate the concept of many-to-many mapping of phenotype on performance. Each performance measure is affected by multiple phenotypic traits, and various phenotypic traits are involved in trade-offs, facilitations, and one-to-one relationships (Table 2). The phenotypic traits that trade-off (rDigit in *Co. texanus* and rThigh in the interspecific example) are most constrained functionally. Each performance measure can nevertheless be selected for, either via the evolution of phenotypic traits involved in facilitations or one-to-one relationships. These facilitations and one-to-one relationships mitigate the effects of the trade-offs on the functional system. Indeed, row sums tend to be greatest for facilitating and "one-to-one" traits, and lowest for traits that trade-off. The interspecific **F**-matrix allows us to test whether phenotypic traits that are least constrained functionally also have the high rates of evolution, as estimated by fitting a Brownian motion model to each phenotypic trait (Table 2B). Interestingly, the trait that trades-off (rThigh) has the highest functional constraint (row sum) and lowest rate of evolution (sigma). However, there is only a weak relationship between functional constraint and rate of evolution in general ($R = 0.142$). There is a stronger but inverse relationship ($R = -0.228$) between sigma and the row variance, the latter being a measure of variation in the constraining effects of the performance measures on phenotypic traits. This suggests that the concept of functional constraint is at least partially related to rates of evolution on phylogenies (Holzman et al. 2012).

Interspecific **F**-Array Analysis

There are some phylogenetic trends in species **F**-matrices. A table of row sums for all species is a map of how functional constraint on phenotypic traits varies among species (Table 3). In our dataset, the most specialized sprinters, *Ca. draconoides* and *Co. texanus* have the weakest functional constraints, and the front limb along with the rDigit of the hind limb are least constrained across species (Table 3). Both species of *Phrynosoma* and *Sceloporus jarrovii* have phenotypes that are most constrained functionally and, interestingly, are the slowest running and most ecologically specialized (Table 3, also see discussion).

The **eVCV** matrix for each performance measure describes the rate (diagonal) and functional co-evolution (off-diagonal) of the influence exerted by phenotypic traits on performance at a macroevolutionary level (Table 4). In our **eVCVs**, the influence of rCrus on acceleration evolves most quickly in this group of species, while the influence of body length on acceleration evolves most slowly (Table 4, upper triangle). The influence of rCrus on velocity has co-

Table 4 Evolutionary VCV matrix for velocity (lower triangle) and acceleration (upper triangle) for eight species of phrynosomatine lizards, given the phylogeny presented in Online Resource 1

	SVL	rFLL	rThigh	rCrus	rFoot	rDigit
SVL	5.7 / 5.2	-0.8	-4.3	2.1	1.0	4.6
rFLL	1.3	8.3 / 8.6	5.3	-13.9	3.8	-8.1
rThigh	-1.1	-1.7	2.2 /12.6	-17.9	5.2	-10.1
rCrus	-6.2	-6.2	2.5	11.2/35.4	-11.9	17.8
rFoot	3.0	3.8	-1.2	-6.4	4.6 / 5.5	-5.1
rDigit	0.6	-3.9	-0.2	0.1	-1.1	4.9 /14.3

Diagonals contain evolutionary variances for both aspects of performance. Negative covariances are shaded black and positive ones are shaded grey

Table 5 Mantel tests comparing intraspecific **F**-matrices to the clade-level **F**-matrix

Species	R	*p*
Ca.dra.	−0.196	0.5560
Co.tex.	−0.292	0.4090
H.mac.	−0.092	0.7498
P.mca.	−0.218	0.5262
P.pla.	0.090	0.7850
S.jar.	0.305	0.3242
S.vir.	0.311	0.3396
U.sta.	−0.338	0.2804

Each Mantel test is based on 10,000 data randomizations

evolved in an opposite direction to SVL, rFLL, and rFoot interspecifically (Table 4, lower triangle). In contrast, the influences of rCrus and rDigit on acceleration are positively related. A Mantel test comparing **eVCV** matrices for velocity and acceleration approaches significance ($R = 0.653$, $P = 0.089$), indicating that phenotypic traits influence both maximum velocity and acceleration in somewhat similar ways in phrynosomatine lizards.

The interspecific $\mathbf{F}^{\mathbf{T}}\mathbf{F}$ matrix shows a weak relationship between velocity and acceleration ($R = 0.071$), in contrast to higher positive correlations for individual species (Fig. 3). This is likely due to rThigh being involved in a strong trade-off (Table 2B). Mantel tests indicate that none of the species-level **F**-matrices are significantly related to the clade-level **F**-matrix (Table 5), suggesting that macroevolutionary phenotype-performance relationships are different from intraspecific ones. Such functional divergence among species is further demonstrated with pairwise Mantel tests on intraspecific **F**-matrices, where very few species have similar **F**-matrices (Table 6). Interestingly, the sprinting specialist sand lizards (*Callisaurus, Cophosaurus*, and *Holbrookia*) have among the most similar **F**-matrices (Table 6), suggesting common phenotypic-performance relationships in sprinters.

Discussion

Many-to-Many Mapping of Phenotype on Performance

That there are complex relationships between phenotype and performance is well accepted (Wainwright 2007). Studies of redundancy (Toro et al. 2004; Alfaro et al. 2005) and trade-offs (Gould and Lewontin 1979; Losos 1993; Garland and Else 1994; Westneat 1994; Vanhooydonck et al. 2001; Wilson et al. 2002; Angilletta et al. 2008; Holzman et al. 2011b, 2012) have given us important insights into these complex relationships. Together, these studies suggest that redundancy, trade-offs, facilitations, and one-to-one relationships between phenotype and performance interact with one another. For example, increased redundancy and complexity in a phenotype can mitigate trade-offs between aspects of performance (Holzman et al. 2011b). However, most often, these phenotype-performance relationships are studied independently, and this hampers our understanding of their complexity and integrated nature.

The **F**-matrix approach provides a powerful quantitative tool for studying the complexity of phenotype-performance relationships, allowing for the consideration of functional constraint and evolvability of phenotype and performance (Ghalambor et al. 2003; Walker 2007). The **F**-matrix is perfect for studying many-to-many mapping, which is broadly applicable to any functional system by allowing for

Table 6 Mantel tests comparing intraspecific **F**-matrices in a pairwise manner

	Ca.dra.	***Co.tex.***	***H.mac.***	***P.mca.***	***P.pla.***	***S.jar.***	***S.vir.***	***U.sta.***
Ca.dra.		**0.015**	**0.027**	0.216	0.484	0.305	0.706	0.104
Co.tex.	**0.775**		0.080	**0.011**	0.246	0.738	0.481	0.115
H.mac.	**0.735**	0.699		0.377	**0.003**	0.955	0.270	0.114
P.mca.	-0.374	**-0.639**	-0.264		0.953	0.244	0.274	**0.006**
P.pla.	-0.232	-0.357	**-0.781**	0.009		0.390	**0.007**	0.195
S.jar.	-0.341	-0.066	0.029	-0.411	-0.268		0.585	0.894
S.vir.	0.107	-0.165	-0.370	-0.347	**0.710**	0.167		0.756
U.sta.	-0.520	-0.566	-0.575	**0.705**	0.392	-0.076	0.054	

The lower triangle provides the correlations and the upper triangle provides p-values. Values in bold are significant after correction for multiple comparisons using the Benjamini-Hochberg method. Boxes near the diagonal contain within-clade comparisons. Each Mantel test is based on 10,000 data randomizations one-to-one relationships, trade-offs, facilitation, and redundancy to be considered simultaneously so their interactions can be studied.

The many-to-many concept is predictive of how phenotypes may evolve along "functional lines of least resistance" when under selection for performance, similar to the concept of evolving along genetic lines of least resistance (Schluter 1996). Indeed both functional and genetic lines of least resistance are important to consider in predicting how phenotypic evolution is likely to occur, where the former addresses the effects of functional constraints and integration, while the latter addresses genetic constraints resulting from heritable variation and covariation. Many-to-many mapping further helps to explain why some traits evolve more readily than others. Holzman et al. (2011a) showed that redundancy and complexity mitigate trade-offs. Our findings agree with those of Holzman et al. (2011a), but go further, suggesting that trade-offs are most mitigated by phenotypic traits that are involved in facilitation or one-to-one relationships with performance. The simulated data of Holzman et al. (2011b), also support this because, while the strength of the initial trade-off is reduced with the addition of more phenotypic traits, some of the added traits are involved in trade-offs and reverse this trend (for example, traits 8 and 10 in their Fig. 2).

We would expect phenotypic traits involved in functional trade-offs to be most constrained in their evolution, traits that uniquely affect a single performance measure to evolve more readily, and traits involved in facilitations to evolve most readily because they can enhance multiple aspects of performance simultaneously. Our results lend some support to this, as rThigh was involved in a trade-off in the interspecific **F**-matrix and also had the lowest rate of evolution, but the general pattern is weak (Table 2B). This may be because Brownian rates of evolution do not take functional constraints into account (while **F**-matrix row sums and variances do), or because our **F**-matrices do not take genetic variation into account (see below for further discussion). Interestingly, not all functional trade-offs constrain evolution: evolutionary rates in the feeding apparatus of fishes are larger for traits strongly involved in functional trade-offs (Holzman et al. 2012). One possible explanation for this is that the constraining nature of trade-offs depends on the relative impacts of a phenotypic trait on the performance measures—if one performance measure is influenced far more than the other, then this can result in a trade-off associated with a trait that evolves quickly. Another explanation for this is that not all performance measures are equally important to fitness, so performance measures less important to fitness will also help to mitigate the effect of trade-offs. For example, relative and absolute velocity are differentially important to evading a predator (Van Damme and Van Dooren 1999), and the importance of biting and suction feeding differ based on diet in fishes (Price et al. 2011). To better understand the interacting effects of trade-offs, redundancies, and facilitations on evolutionary rates and trends, more empirical studies are needed using the **F**-matrix and considering many-to-many mapping.

Other studies of lizard locomotion have also shown that traits with a one-to-one relationship with a performance measure, trade-offs, facilitation, and redundancy all occur, but generally have not considered these phenomena simultaneously. For example, longer hind limbs typically allow for longer strides and higher sprint speed (Losos 1990; Bonine and Garland 1999; Irschick and Jayne 1999; Losos et al. 2002; Toro et al. 2003), but acceleration is determined by muscle force, which is driven by muscle mass and fiber type (Irschick et al. 2003; Vanhooydonck et al. 2006b; Higham et al. 2011a). Hence, velocity and acceleration should both be able to be maximized because they depend upon different aspects of the phenotype. Indeed, velocity and acceleration are positively related in *Anolis* lizards (Vanhooydonck et al. 2006a) and phrynosomatines (this study), despite general expectations of force–velocity trade-offs in functional systems (e.g., Westneat 1994; Levinton and Allen 2005).

Our findings also suggest that the degree of specialization influences phenotype-performance relationships in phrynosomatine lizards. Specialization can be viewed as facilitating the invasion of new niches, which could lead to further evolution of diverse phenotypes, as articulated in a key innovation hypothesis (Simpson 1944; Mayr 1963; Nitecki 1990), or it could result in an evolutionary dead-end whereby a specialized phenotype is only good at doing one thing, leading to decreased phenotypic evolution (Collar et al. 2009). We found that the species with the weakest functional constraints (*Ca. draconoides* and *Co. texanus*) were the most specialized for sprinting, possessing long, gracile hind limbs and frequently using bipedal locomotion (Irschick and Jayne 1999; Stebbins 2003). In contrast, the slow moving ant-eating specialists (*Phrynosoma* spp.) and *Sceloporus jarrovii* had the strongest functional constraints (Table 3), lending credence to the hypothesis that some specialists are functionally constrained in their evolution. A similar finding was described for piscivorous centrarchid fishes (Collar et al. 2009). Although a strong functional constraints and generalized body shape in *S. jarrovii* appears to contradict this hypothesis, the ecology of this species is highly specialized in that it inhabits medium-to-high elevation talus slopes and boulder fields (Brennan and Holycross 2006), darts under rocks to escape predators, and so runs only very short distances (pers. obs.).

Some previous studies of complex phenotype-function relationships have developed explicit biomechanical models relating phenotypic traits to aspects of function, and explored the properties of their models (e.g., Westneat 1994; Alfaro et al. 2005; Wainwright et al. 2005; Holzman et al. 2011b). Such a model precisely relates phenotype to function and facilitates the study of predictive trait evolution through simulation (Alfaro et al. 2005; Holzman et al. 2011b). A biomechanical model of lizard locomotion does not yet exist, so should be a future area of research, as it will further enhance our understanding of the system. Lack of explicit biomechanical models is the case with many functional morphology studies, which predominantly rely on relationships between phenotypic traits and performance. Using the latter approach is noisier, but is an important step to ultimately creating a biomechanical model that can complement the **F**-matrix.

The Interspecific **F**-array

We extended the **F**-matrix approach to accommodate interspecific datasets by assembling a 3-D **F**-array, with phenotypic traits, performance measures, and species as the three axes (Fig. 1). The **F**-array has the important strength of allowing for the direct comparison of phenotype-performance relationships among multiple species, while not discarding intraspecific patterns of variation (e.g., Table 2). How micro- and macroevolution are related to one another is a fundamental question in evolutionary biology, and our approach is one that allows for the quantitative comparison of these scales of evolution from a functional perspective (Tables 5, 6).

The **F**-array approach assumes that relationships between phenotype and performance can evolve, although these relationships themselves may not be directly determined by specific genes. This is not a trivial assumption, but is supported by both current practice in macroevolutionary analyses and evolutionary theory. In practice, studies of the macroevolution of phenotype-performance relationships typically use phylogenetically independent contrasts of various performance measures (e.g., Garland et al. 1992; Vanhooydonck et al. 2001; Herrel et al. 2005; Vanhooydonck et al. 2006b; Holzman et al. 2011b), and this makes the same assumption as the **F**-array approach. Recent theoretical studies have also suggested that phenomena without a direct genetic basis such as phenotypic plasticity (Pfennig et al. 2010; Pigliucci 2010), niche construction (Odling-Smee 2010), and epigenetics (Jablonka and Lamb 2010) can be heritable and evolve. The concept of exaptation, where a phenotypic trait evolves for one function only to be co-opted for another function (Gould and Vrba 1982), also implies changes to phenotype-performance relationships through time. Finally, experimental evidence from *Anolis* lizards shows that natural selection can act on correlations between phenotype and performance, as opposed to acting directly on the phenotypic traits or performance measures (Calsbeek 2008).

Our study and others (Vanhooydonck et al. 2006b; Holzman et al. 2011b) have also assembled interspecific **F**-matrices using phylogenetically independent contrasts (Felsenstein 1985), which may allow for smaller intraspecific sample sizes (Harmon and Losos 2005). However, this approach necessarily ignores species-specific relationships between phenotype and performance (Hansen and Martins 1996; Ives et al. 2007). Both our **F**-array and the **F**-matrix with phylogenetically independent contrasts approaches are perfectly valid, but have different goals. The **F**-array allows one to compare phenotype-performance relationships among species and can be compared to intraspecific patterns, while the latter approach is focused on studying clade-level phenotype-performance relationships. Our results (Table 5) show that the two approaches provide different information and that intraspecific and interspecific phenotype-performance relationships often do not coincide. The comparison of micro- to macroevolutionary patterns has been a persistent challenge in evolutionary biology, and this is an important strength of our approach.

Implementation and Technical Issues

Frequently in biology, mathematical and theoretical advances precede either data availability or wide-spread knowledge of how to apply them (e.g., Fisher 1930; Felsenstein 1973; Lande 1979a; Arnold 1983), and applying such techniques can be daunting. Despite the utility of the **F**-matrix/**F**-array approach, it has only rarely been applied to real data (e.g., Blob et al. 2010; Holzman et al. 2011b; this study). To facilitate further use of these techniques, we have implemented the entire workflow in a series of functions in R (R Development Core Team 2012). These functions and a usage manual are available in Online Resources 4 and 5.

There are two issues that must be considered that influence the power of an **F**-matrix or **F**-array. The more obvious of these is sample size. Sample sizes must be sufficiently high to give acceptable power to detect phenotype-performance relationships in each species. Our study included data for 8–14 individuals for each of eight species, a substantial dataset for a comparative functional morphology study, but really the bare minimum for an **F**-matrix for any given species. Although these sample sizes are likely sufficient to account for intraspecific variation (Harmon and Losos 2005; Adolph and Hardin 2007), sample size must also exceed the number of phenotypic traits so as not to run out of degrees of freedom (Sokal and Rohlf 1995). This affects the choice of phenotypic traits for inclusion in the analysis. Restricting the number of phenotypic traits in an analysis means that estimates of system complexity should be made in the context of the traits that are included. However, since closely related species have the same suite of traits, choosing a subset of traits may not be problematic. Nevertheless, an additional consequence of including only a subset of traits is that the omitted traits may also affect performance, and this can influence partial regression coefficients (Walker 2014). Hence, it is important to carefully consider which traits to include in an analysis.

The second issue is collinearity, which occurs when multiple explanatory variables (phenotypic traits) are highly correlated with one another, explaining non-independent components of variation in the response (Quinn and Keough 2002; Murray and Conner 2009). Collinearity acts to inflate the standard error of parameter estimates in various techniques, including multiple regression, and in some cases can make parameter estimates unstable (Murray and Conner 2009; Dormann et al. 2013). Collinearity is common in morphological and ecological data (Fisher and Hogan 2007; Yakubu 2010; Dormann et al. 2013). For example, even though we adjusted our phenotypic variables for size and phylogeny, 21 % had tolerances <0.1 and between one and three pairwise correlations between phenotypic traits in each species were >0.7 (Online Resource 3)—these are the suggested cutoffs for when collinearity becomes a problem (Freckleton 2011; Dormann et al. 2013). Fortunately, collinearity often does not bias parameter estimates, but does decrease the power of significance tests (Freckleton 2011). Other methods such as principal component analysis and canonical phylogenetic ordination can eliminate collinearity (Quinn and Keough 2002; Giannini 2003), but confound relationships between performance and specific phenotypic traits (Holzman et al. 2011b). The technique of hierarchical partitioning explicitly addresses collinearity (Chevan and Sutherland 1991; Murray and Conner 2009), but provides numbers equivalent to correlations, which are non-linearly related to standardized partial regression coefficients, so may confound **F**-matrix interpretation (Phillips and Arnold 1999). Nevertheless, hierarchical partitioning represents a promising approach that needs to be explored in more detail in the context of the **F**-matrix.

We adopted the use of effect sizes (Cohen 1988; Steidl and Thomas 2001) to help interpret our **F**-matrices due to the low power of significance tests, which may be unreliable due to collinearity (Dormann et al. 2013). While this may be a caveat, the issue of how to correct partial regression coefficients for collinearity remains unresolved. Our approach allowed for interpretation of our results and draws attention to the continuous nature of the values in the **F**-matrix. One may rely on effect sizes or significance tests to identify one-to-one relationships, but it is important to remember that some of these are actually weak/non-significant facilitations or trade-offs, and still affect **F**-matrix calculations such as the row and column sums and variances.

Extensions of the **F**-matrix and **F**-array

The **F**-matrix and **F**-array provide an innovative approach to studying phenotype-performance relationships and can be applied to any functional system, or really any system in which multiple explanatory variables are related to multiple response variables. For example, kinematics is a behavioral mechanism that influences relationships between phenotype and performance, making a functional system even more complex (Lauder and Reilly 1996; Irschick and Jayne 1998, 1999). Kinematics allows morphologically similar species to attain different levels of performance or morphologically dissimilar species to attain similar levels of performance, resulting in many-to-one mapping (Wainwright et al. 2005; Kohlsdorf and Navas 2007). Kinematics can be incorporated into our approach by producing one **F**-matrix relating phenotypic traits to kinematic traits and a second **F**-matrix relating kinematics to performance, effectively expanding the form-function paradigm (Arnold 1983) to Phenotype → Kinematics → Performance → Fitness. Our framework is not limited to kinematics, locomotion, or gross morphological traits. Other aspects of the phenotype (such as muscle physiology and force generation), and other functional systems (such as feeding or reproductive performance) can and should be analyzed in the context of many-to-many mapping and the **F**-matrix.

Another area that requires further development is integration of the **F**-matrix with the additive genetic variance–covariance, or **G**-matrix (Lande 1979b; Steppan et al. 2002). These matrices have been theoretically integrated into a single framework (by (Walker 2010), but rarely empirically, although Holzman et al. (2011b) did this in a simulation study, using a matrix of functional trait variances and covariances (their equation 1). Integration of **G** and **F**-matrices is important because both have been shown to evolve (Steppan et al. 2002; Cano et al. 2004; Revell 2007); this study). Furthermore, the $\mathbf{FF^T}$ provides an estimate of functional integration among phenotypic traits, while the **G**-matrix provides a similar estimate, but of genetic integration among phenotypic traits (Schluter 1996). Therefore, patterns of functional and genetic integration among traits can be compared.

Acknowledgments The original morphometric and performance data were collected as part of the PhD project of PJB, and are published in Bergmann and Irschick (2010). We thank R. Bergmann, who suggested the idea of using grand sums as overall measures of functional constraint. We also thank C. Martinez del Rio, C. Murren, the Bergmann lab, and two anonymous reviewers for insightful comments that greatly improved this article.

References

Adolph, S. C., & Hardin, J. S. (2007). Estimationg phenotypic correlations: Correcting for bias due to individual variability. *Functional Ecology, 21*, 178–184.

Aerts, P., Vanhooydonck, B., Zaaf, A., & Herrel, A. (2000). Lizard locomotion: How morphology meets ecology. *Netherlands Journal of Zoology, 50*, 261–277.

Alfaro, M. E., Bolnick, D. I., & Wainwright, P. C. (2005). Evolutionary consequences of many-to-one mapping of jaw morphology to mechanics of labrid fishes. *American Naturalist, 165*, E140–E154.

Angilletta, M. J., Wilson, R. S., Niehaus, A. C., & Ribiero, P. (2008). The fast and the fractalous: Tradeoffs between running speed and manoeuvrability in leaf-cutter ants. *Functional Ecology, 22*, 78–83.

Arnold, S. J. (1983). Morphology, performance and fitness. *American Zoologist, 23*, 347–361.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57*, 289–300.

Bergmann, P. J., & Irschick, D. J. (2010). Alternate pathways of body shape evolution translate into common patterns of locomotor evolution in two clades of lizards. *Evolution, 64*, 1569–1582.

Bergmann, P. J., Meyers, J. J., & Irschick, D. J. (2009). Directional evolution of stockiness coevolves with ecology and locomotion in lizards. *Evolution, 63*, 215–227.

Blob, R. W., Kawano, S. M., Moody, K. N., Bridges, W. C., Maie, T., Ptacek, M. B., et al. (2010). Morphological selection and the evaluation of potential tradeoffs between escape from predators and the climbing of waterfalls in the Hawaiian stream goby *Sicyopterus stimpsoni*. *Integrative and Comparative Biology, 50*, 1185–1199.

Blomberg, S. P., Garland, T., Jr, & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution, 57*, 717–745.

Bonine, K. E., & Garland, T. (1999). Sprint performance of phrynosomatid lizards, measured on a high-speed treadmill, correlates with hindlimb length. *Journal of Zoology, London, 248*, 255–265.

Bonine, K. E., Gleeson, T. T., & Garland, T., Jr. (2001). Comparative analysis of fiber-type composition in the iliofibularis muscle of phrynosomatid lizards (Squamata). *Journal of Morphology, 250*, 265–280.

Bonnett, E., & Van de Peer, Y. (2002). zt: A software tool for simple and partial Mantel tests. *Journal of Statistical Software, 7*, 1–12.

Brennan, T. C., & Holycross, A. T. (2006). *A fieldguide to amphibians and reptiles in Arizona*. Phoenix, AZ: Arizona Game & Fish Department.

Calsbeek, R. (2008). An ecological twist on the morphology-performance-fitness axis. *Evolutionary Ecology Research, 10*, 197–212.

Cano, J. M., Laurila, A., Palo, J., & Merila, J. (2004). Population differentiation in G matrix structure due to natural selection in *Rana temporaria*. *Evolution, 58*, 2013–2020.

Chevan, A., & Sutherland, M. (1991). Hierarchical partitioning. *American Statistician, 45*, 90–96.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, USA: Lawrence Erlbaum Associates Inc.

Collar, D. C., O'Meara, B. C., Wainwright, P. C., & Near, T. J. (2009). Piscivory limits diversification of feeding morphology in centrarchid fishes. *Evolution, 63*, 1557–1573.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London, England: John Murray.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*, 27–46.

Felsenstein, J. (1973). Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology, 22*, 240–249.

Felsenstein, J. (1985). Phylogenies and the comparative method. *American Naturalist, 125*, 1–15.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford, UK: Clarendon Press.

Fisher, R., & Hogan, J. D. (2007). Morphological predictors of swimming speed: A case study of pre-settlement juvenile coral reef fishes. *Journal of Experimental Biology, 210*, 2436–2443.

Freckleton, R. P. (2011). Dealing with collinearity in behavioral and ecological data: Model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology, 65*, 91–101.

Garland, T., & Carter, P. A. (1994). Evolutionary physiology. *Annual Review of Physiology, 56*, 579–621.

Garland, T., & Else, P. L. (1994). Quantitative genetics of locomotor behavior and physiology in a garter snake. In C. R. B. Boake (Ed.), *Quantitative genetic studies of behavioral evolution* (pp. 251–277). Chicago, USA: University of Chicago Press.

Garland, T., Harvey, P. H., & Ives, A. R. (1992). Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Systematic Biology, 41*, 18–32.

Garland, T., & Losos, J. B. (1994). Ecological morphology of locomotor performance in squamate reptiles. In P. C. Wainwright & S. M. Reilly (Eds.), *Ecological morphology: Integrative organismal biology* (pp. 240–302). Chicago, USA: University of Chicago Press.

Ghalambor, C. K., Walker, J. A., & Reznick, D. N. (2003). Multi-trait selection, adaptation, and constraints on the evolution of burst swimming performance. *Integrative and Comparative Biology, 43*, 431–438.

Giannini, N. P. (2003). Canonical phylogenetic ordination. *Systematic Biology, 52*, 684–695.

Gibbs, L. H., & Grant, P. R. (1987). Oscillating selection on darwin's finches. *Nature, 327*, 511–513.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London B, 205*, 581–598.

Gould, S. J., & Vrba, E. S. (1982). Exaptation—A missing term in the science of form. *Paleobiology, 8*, 4–15.

Hansen, T. F., & Martins, E. P. (1996). Translating between microevoluionary process and macroevolutionary patterns: The correlational strucutre of interspecific data. *Evolution, 50*, 1404–1417.

Harmon, L. J., & Losos, J. B. (2005). The effects of intraspecific sample size on type I and type II error rates in comparative studies. *Evolution, 59*, 2705–2710.

Harmon, L. J., Schulte, J. A., Larson, A., & Losos, J. B. (2003). Tempo and mode of evolutionary radiation in iguanian lizards. *Science, 301*, 961–964.

Herrel, A., Meyers, J. J., & Vanhooydonck, B. (2002). Relations between microhabitat use and limb shape in phrynosomatid lizards. *Biological Journal of the Linnean Society, 77*, 149–163.

Herrel, A., Podos, J., Huber, S. K., & Hendry, A. P. (2005). Evolution of bite force in Darwin's finches: a key role for head width. *Journal of Evolutionary Biology, 18*, 669–675.

Herrel, A., Podos, J., Vanhooydonck, B., & Hendry, A. P. (2009). Force-velocity trade-off in Darwin's finch jaw function: A biomechanical basis for ecological speciation? *Functional Ecology, 23*, 119–125.

Higham, T. E., Korchari, P. G., & McBrayer, L. D. (2011a). How muscles define maximum running performance in lizards: An analysis using swing- and stance-phase muscles. *Journal of Experimental Biology, 214*, 1685–1691.

Higham, T. E., Korchari, P. G., & McBrayer, L. D. (2011b). How muscles define maximum running performance in lizards: An analysis using swing and stance phase muscles. *Journal of Experimental Biology, 214*, 1685–1691.

Holzman, R., Collar, D. C., Mehta, R. S., & Wainwright, P. C. (2011a). An integrative approach to elucidate suction feeding performance. *Journal of Experimental Biology, 215*, 1–13.

Holzman, R. S., Collar, D. C., Mehta, R. S., & Wainwright, P. C. (2011b). Mechanical complexity can mitigate evolutionary trade-offs. *American Naturalist, 177*, E69–E83.

Holzman, R., Collar, D. C., Price, S. A., Hulsey, C. D., Thomson, R. C., & Wainwright, P. C. (2012). Biomechanical trade-offs bias rates of evolution in the feeding apparatus of fishes. *Proceedings of the Royal Society of London B, 279*, 1287–1292.

Huber, S. K., & Podos, J. (2006). Beak morphology and song features covary in a population of Darwin's finches (Geospiza fortis). *Biological Journal of the Linnean Society, 88*, 489–498.

Irschick, D. J., & Jayne, B. C. (1998). Effects of incline on speed, acceleration, body posture and hindlimb kinematics in two species of lizard Callisaurus draconoides and Uma scoparia. *Journal of Experimental Biology, 201*, 273–287.

Irschick, D. J., & Jayne, B. C. (1999). Comparative three-dimensional kinematics of the hindlimb for high-speed bipedal and quadrupedal locomotion of lizards. *Journal of Experimental Biology, 202*, 1047–1065.

Irschick, D. J., & Losos, J. B. (1998). A comparative analysis of the ecological significance of maximal locomotor performance in Caribbean Anolis lizards. *Evolution, 52*, 219–226.

Irschick, D. J., Meyer, A., Husak, J. F., & Le Galliard, J.-F. (2008). How does selection operate on whole-organism functional performance capacities? A review and synthesis. *Evolutionary Ecology Research, 10*, 177–196.

Irschick, D. J., Vanhooydonck, B., Hertel, A., & Andronescu, A. (2003). Effects of loading and size on maximum power output and gait characteristics in geckos. *Journal of Experimental Biology, 206*, 3923–3934.

Ives, A. R., Midford, P. E., & Garland, T., Jr. (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology, 56*, 252–270.

Jablonka, E., & Lamb, M. J. (2010). Transgenerational epigenetic inheritance. In M. Pigliucii & G. Muller (Eds.), *Evolution: The extended synthesis* (pp. 137–174). Cambridge, USA: MIT Press.

Kohlsdorf, T., & Navas, C. A. (2007). Evolution of jumping capacity in Tropidurinae lizards: Does habitat complexity influence obstacle-crossing ability? *Biological Journal of the Linnean Society, 91*, 393–402.

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

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

Langerhans, R. B. (2009). Trade-off between steady and unsteady swimming underlies predator-driven divergence in *Gambusia affinis*. *Journal of Evolutionary Biology, 22*, 1057–1075.

Langerhans, R. B., & Reznick, D. N. (2010). Ecology and evolution of swimming performance in fishes: Predicting evolution with biomechanics. In P. Domenici & B. G. Kapoor (Eds.), *Fish locomotion: An etho-ecological perspective* (pp. 200–248). Enfield, USA: Science Publishers.

Lauder, G. V., & Reilly, S. M. (1996). The mechanistic basis of behavioral evolution: Comparative analysis of muscoskeletal function. In E. P. Martins (Ed.), *Phylogenies and the comparative method in animal behaviour* (pp. 105–137). Oxford, England: Oxford University Press.

Levinton, J. S., & Allen, B. J. (2005). The paradox of the weakening combatant: Trade-off between closing force and gripping speed in a sexually selected combat structure. *Functional Ecology, 19*, 159–165.

Lewontin, R. C. (1978). Adaptation. *Scientific American, 239*, 156–169.

Losos, J. B. (1990). The evolution of form and function: morphology and locomotor performance in west indian *Anolis* lizards. *Evolution, 44*, 1189–1203.

Losos, J. B. (1993). Trade-offs between sprinting and clinging ability in Kenyan chameleons. *Functional Ecology, 7*, 281–286.

Losos, J. B., Mouton, L. N., Bickel, R., Cornelius, I., & Ruddock, L. (2002). The effect of body armature on escape behavior in cordylid lizards. *Animal Behaviour, 64*, 313–321.

Maddison, W. P., & Maddison, D. R. (2011). Mesquite: A modular system for evolutionary analysis v.2.75. http://mesquiteproject.org.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research, 27*, 209–220.

Mayr, E. (1963). *Animal species and evolution*. Cambridge, USA: Harvard University Press.

McElroy, E. J., Archembeau, K. L., & McBrayer, L. D. (2012). The correlation between locomotor performance and hindlimb kinematics during burst locomotion in the Florida scrub lizard, *Sceloporus woodi*. *Journal of Experimental Biology, 215*, 442–453.

McElroy, E. J., & McBrayer, L. D. (2010). Getting up to speed: Acceleration strategies in the Florida scrub lizard, *Sceloporus woodi*. *Physiological and Biochemcial Zoology, 83*, 643–653.

McNally, R. C., & Doolan, J. M. (1986). Patterns of morphology and behavior in a cicada guild: A neutral model hypothesis. *Australian Journal of Zoology, 11*, 279–294.

Miles, D. B. (1994). Covariation between morphology and locomotor performance in sceloporine lizards. In L. J. Vitt & E. R. Pianka (Eds.), *Lizard ecology: Historical and experimental perspectives* (pp. 207–235). Princeton, USA: Princeton University Press.

Moran, M. D. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos, 100*, 403–405.

Murray, K., & Conner, M. M. (2009). Methods to quantify variable importance: Implications for the analysis of noisy ecological data. *Ecology, 90*, 348–355.

Nitecki, M. (1990). The plurality of evolutionary innovations. In M. Nitecki (Ed.), *Evolutionary innovations* (pp. 3–18). Chicago, USA: University of Chicago Press.

Odling-Smee, J. (2010). Niche inheritance. In M. Pigliucii & G. Muller (Eds.), *Evolution: The extended synthesis* (pp. 175–208). Cambridge, USA: MIT Press.

Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C. D., & Moczek, A. P. (2010). Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution, 25*, 459–467.

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

Pigliucci, M. (2010). Phenotypic plasticity. In M. Pigliucii & G. Muller (Eds.), *Evolution: The extended synthesis* (pp. 355–378). Cambridge, USA: MIT Press.

Podos, J. (2001). Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature, 409*, 185–188.

Price, S. A., Holzman, R., Near, T. J., & Wainwright, P. C. (2011). Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes. *Ecology Letters, 14*, 462–469.

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge, UK: Cambridge University Press.

Reeder, T. W., & Wiens, J. J. (1996). Evolution of the lizards family Phrynosomatidae as inferred from diverse types of data. *Herpetological Monographs, 10*, 43–84.

Reilly, S. M. (1995). Quantitative electromyography and muscle function of the hind limb during quadrupedal running in the lizard *Sceloporus clarki*. *Zoology, 98*, 263–277.

Revell, L. J. (2007). The G matrix under fluctuating correlational mutation and selection. *Evolution, 61*, 1857–1872.

Revell, L. J. (2009). Size-correction and principal components for interspecific comparative studies. *Evolution, 63*, 3258–3268.

Revell, L. J., & Collar, D. C. (2009). Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution, 63*, 1090–1100.

Scales, J. A., King, A. A., & Butler, M. A. (2009). Running for your life or running for your dinner: what drives fiber-type evolution in lizard locomotor muscles? *American Naturalist, 173*, 543–553.

Schluter, D. (1996). Adaptive radiation along genetic lines of least resistance. *Evolution, 50*, 1768–1774.

Sherbrooke, W. C. (2003). *Introduction to Horned Lizards of North America*. Berkeley, USA: California University Press.

Simpson, G. G. (1944). *Tempo and mode in evolution*. New York, USA: Columbia University Press.

Sokal, R. R., & Rohlf, F. J. (1995). *Biometry*. New York: W.H. Freeman and Co.

Stebbins, R. C. (2003). *Peterson field guide to the western reptiles and amphibians*. Boston, USA: Houghton Mifflin.

Steidl, R. J., & Thomas, L. (2001). Power analysis and experimental design. In S. M. Scheiner & J. Gurevitch (Eds.), *Design and analysis of ecological experiments* (pp. 14–36). Oxford, UK: Oxford University Press.

Steppan, S. J., Phillips, P. C., & Houle, D. (2002). Comparative quantitative genetics: Evolution of the G matrix. *Trends in Ecology & Evolution, 17*, 320–327.

Strobbe, F., McPeek, M. A., De Block, M., De Meester, L., & Stoks, R. (2009). Survival selection on escape performance and its underlying phenotypic traits: A case of many-to-one mapping. *Journal of Evolutionary Biology, 22*, 1172–1182.

Team, R. D. C. (2012). *R: A language and environment for statistical computing*. Vienna, Austria.

Toro, E., Herrel, A., & Irschick, D. J. (2004). The evolution of jumping performance in Caribbean Anolis lizards: Resolution of a biomechanical trade-off? *American Naturalist, 163*, 844–856.

Toro, E., Herrel, A., Vanhooydonck, B., & Irschick, D. J. (2003). A biomechanical analysis of intra- and interspecific scaling of jumping and morphology in Caribbean Anolis lizards. *Journal of Experimental Biology, 206*, 2641–2652.

Van Damme, R., & Van Dooren, T. J. M. (1999). Absolute versus per unit body length speed of prey as an estimator of vulnerability to predation. *Animal Behaviour, 57*, 347–352.

Van Damme, R., & Vanhooydonck, B. (2001). Origins of interspecific variation in lizard sprint capacity. *Functional Ecology, 15*, 186–202.

Van Damme, R., Vanhooydonck, B., Aerts, P., & De Vree, F. (2003). *Evolution of lizard locomotion: Context and constraint* (pp. 267-282). Vertebrate Biomechanics and Evolution. BIOS Scientific Publishers Ltd., Oxford, England.

Vanhooydonck, B., Boistel, R., Fernandez, V., & Herrel, A. (2011). Push and bite: trade-offs between burrowing and biting in a burrowing skink (*Acontias percivali*). *Biological Journal of the Linnean Society, 102*, 91–99.

Vanhooydonck, B., Herrel, A., & Irschick, D. J. (2006a). Out on a limb: The differential effect of substrate diameter on acceleration capacity in *Anolis* lizards. *Journal of Experimental Biology, 209*, 4515–4523.

Vanhooydonck, B., Herrel, A., Van Damme, R., & Irschick, D. J. (2006b). The quick and the fast: The evolution of acceleration capacity in *Anolis* lizards. *Evolution, 60*, 2137–2147.

Vanhooydonck, B., Van Damme, R., & Aerts, P. (2001). Speed and stamina trade-off in lacertid lizards. *Evolution, 55*, 1040–1048.

Wainwright, P. C. (2007). Functional versus morphological diversity in macroevolution. *Annual Review of Ecology and Systematics, 38*, 381–401.

Wainwright, P. C., Alfaro, M. E., Bolnick, D. I., & Hulsey, D. (2005). Many-to-one mapping of form to function: a general principle in organismal design? *Integrative and Comparative Biology, 45*, 256–262.

Walker, J. A. (1998). Estimating velocities and accelerations of animal locomotion: A simulation experiment comparing numerical differentiation algorithms. *Journal of Experimental Biology, 201*, 981–995.

Walker, J. A. (2007). A general model of functional constraints on phenotypic evolution. *American Naturalist, 170*, 681–689.

Walker, J. A. (2010). An integrative model of evolutionary covariance: A symposium on body shape in fishes. *Integrative and Comparative Biology, 50*, 1051–1056.

Walker, J. A. (2014). The effect of unmeasured confounders on the ability to estimate a true performance or selection gradient (and other partial regression coefficients). *Evolution,*. doi:10.1111/evo.12406.

Westneat, M. W. (1994). Transmission of force and velocity in the feeding mechanisms of labrid fishes (Teleostei, Perciformes). *Zoomorphology, 114*, 103–118.

Wiens, J. A. (1991). Ecomorphological comparisons of the shrub-desert avifaunas of Australia and North America. *Oikos, 60*, 55–63.

Wiens, J. A., & Rotenberry, J. T. (1980). Patterns of morphology and ecology in grassland and shrubsteppe bird populations. *Ecological Monographs, 50*, 287–308.

Williams, V. S. L., Jones, L. V., & Tukey, J. W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics, 24*, 42–69.

Wilson, R. S., James, R. S., & Van Damme, R. (2002). Trade-offs between speed and endurance in the frog *Xenopus laevis*: A multi-level approach. *Journal of Experimental Biology, 205*, 1145–1152.

Yakubu, A. (2010). Fixing multicollinearity instability in the prediction of body weight from morphometric traits of White Fulani Cows. *Journal of Central European Agriculture, 11*, 487–492.

Zaaf, A., & Van Damme, R. (2001). Limb proportions in climbing and ground-dwelling geckos (Lepidosauria, Gekkonidae): A phylogenetically informed analysis. *Zoomorphology, 121*, 45–53.