The Company of
Biologists

RESEARCH ARTICLE

Performance trade-offs and individual quality in decathletes

Jeffrey A. Walker* and Sean P. Caddigan

ABSTRACT

Many constraints of organismal design at the cell and organ level, including muscle fiber types, musculoskeletal gearing and control-surface geometry, are believed to cause performance trade-offs at the whole-organism level. Contrary to this expectation, positive correlations between diverse athletic performances are frequently found in vertebrates. Recently, it has been proposed that trade-offs between athletic performances in humans are masked by variation in individual quality and that underlying trade-offs are revealed by adjusting the correlations to 'control' quality. We argue that quality is made up of both intrinsic components, due to the causal mapping between morpho-physiological traits and performance, and extrinsic components, due to variation in training intensity, diet and pathogens. Only the extrinsic component should be controlled. We also show that previous methods to estimate 'quality-free' correlations perform poorly. We show that Wright's factor analysis recovers the correct quality-free correlation matrix and use this method to estimate quality-free correlations among the 10 events of the decathlon using a dataset of male college athletes. We found positive correlations between all decathlon events, which supports an axis that segregates 'good athletes' from 'bad athletes'. Estimates of quality-free correlations are mostly very small (<0.1), suggesting large, quality-free independence between events. Because quality must include both intrinsic and extrinsic components, the physiological significance of these adjusted correlations remains obscure. Regardless, the underlying architecture of the functional systems and the physiological explanation of both the un-adjusted and adjusted correlations remain to be discovered.

KEY WORDS: Locomotion, Whole-organism performance, Physiological traits, Form–function mapping

INTRODUCTION

Vertebrate locomotor behaviors are powered by functional components with well-known trade-offs: musculoskeletal systems are geared with a specific ratio, and output displacement and force cannot be simultaneously optimized, muscle fibers develop phenotypes that increase either fatigue resistance or power at the cost of the other, and body shapes that augment acceleration have low mechanical efficiency. Physiologists typically expect these trade-offs at lower levels of organization (sub-cellular to organ-system) to scale up to whole-organism performance. Nevertheless, unexpected positive correlations between performance traits are frequently observed (Marras et al., 2013; Vanhooydonck et al., 2014).

Reidy et al. (2000), working with swimming cod, were the first to comment on these positive correlations and proposed that some cod are simply 'good athletes' and others 'bad athletes', which is a

Department of Biological Sciences, University of Southern Maine, 96 Falmouth St, Portland, ME 04103, USA

*Author for correspondence (walker@maine.edu)

useful way to describe but not explain the pattern. Van Damme et al. (2002) and Wilson et al. (2014b) observed only positive correlations among human athletic performances and suggested that trade-offs at the whole-organism level are masked by individual quality, which Wilson et al. (2014b) explain as: 'Because individuals vary in health, physical fitness, nutrition, development or genetics, which is the underlying basis of individual quality, some individuals perform better or worse across all types of motor tasks than others. This means that when researchers try to understand intra-individual functional trade-offs using inter-individual variation in performance, then trade-offs that do occur within individuals can be masked'. We try to clarify this meaning of individual quality with Fig. 1.

Both Van Damme et al. (2002) and Wilson et al. (2014b) argue that the intra-individual trade-offs can be recovered by statistically adjusting for quality. And, both found that the expected negative correlations emerged only after this adjustment. Importantly, both Van Damme et al. (2002) and Wilson et al. (2014b) are cited in the evolutionary and human performance literature as evidence of performance trade-offs without acknowledging that the measured correlations were positive (MacArthur and North, 2005; Flueck, 2009; Ruiz et al., 2010; Eynon et al., 2013; Lailvaux and Husak, 2014; Wilson et al., 2014a; Servedio et al., 2014). We emphasize this because all four methods used to infer trade-offs in Van Damme et al. (2002) and Wilson et al. (2014b) are poor estimators of quality-free correlations. These methods are: (1) culling all but the top performers, which is guaranteed to produce a negative correlation, even if no underlying trade-off exists (Garland, 1994), (2) the correlation between the residuals of performance traits regressed on first principal component scores, which is guaranteed to produce strongly negatively biased correlations (Aitchison, 2003), (3) the interpretation of principal component (PC) loadings of opposite signs as indicating an underlying trade-off, which is not a valid interpretation of loadings, and (4) the partial correlation between two performances conditional on all other performances, which removes too much of the shared correlation (Mitteroecker and Bookstein, 2009). We find similar misuses of multivariate methods common in the performance literature and strongly encourage reading our detailed criticism of all four methods (Walker, 2015a).

Despite these methodological issues, individual quality is a compelling hypothesis to explain a common phenomenon in both human and non-human performance data. In order to explore the concept of quality in performance correlations, we compiled a dataset of decathlon performance data for US collegiate athletes, and used Sewell Wright's (1932) path-analytic factor analysis to estimate 'quality-free' correlations among these 10 events. In this paper, we use a model of functional trade-offs (Ghalambor et al., 2003) to (1) show how functional trade-offs at the cell, organ or system level contribute to performance trade-offs at the whole-animal level, (2) decompose quality into intrinsic and extrinsic components and show how the extrinsic component can mask the underlying architecture of the form–function mapping, (3) show why 'bottom up' approaches to predict performance trade-offs at the
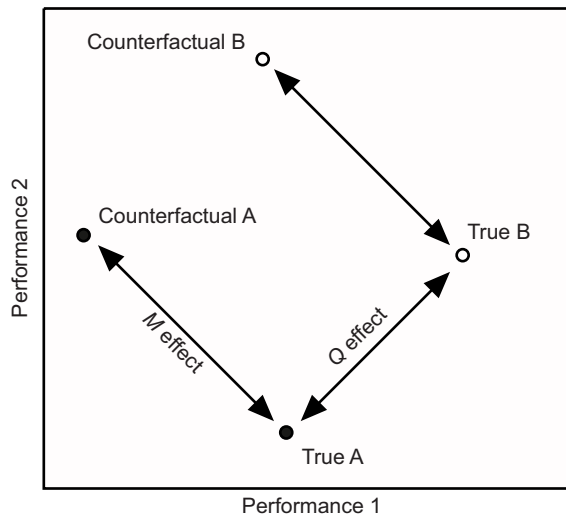
Fig. 1. How individual quality masks functional trade-offs. The double-headed arrow connecting a true and counterfactual individual represents the 'intra-individual functional trade-off' (Wilson et al., 2014b). The double headed arrow between true individual A and true individual B represents the 'inter-individual variation' in quality (Q effect) that masks the intra-individual trade-off. For simplicity, the trade-off is determined by a single underlying morpho-physiological (M-P) trait that has opposite effects on the two performances (M effect). A counterfactual individual is one that is like the real individual in every way except for a change in the underlying M-P trait and the consequent changes in both performances. Individuals A and B do not differ in their M-P trait but individual B is better at both performances because of a difference in exposure to some extrinsic quality variable. This quality might result from something like differences in training or health status (for example, individual A might have a respiratory infection that both narrows respiratory tubes and decreases muscle contractility or motor unit recruitment).

whole-animal level based on limited knowledge of trade-offs at the cell or organ level are likely to fail, and (4) show how to estimate a quality-free correlation matrix. Through analysis of the National Collegiate Athletic Association (NCAA) decathlon data, we then show that, compared with the results of Van Damme et al. (2002), the pattern of measured correlations is similar but our estimates of quality-free correlations differ in key respects. We also re-analyze the performance data from sub-elite male soccer players (Wilson et al., 2014b).

A model of extrinsic and intrinsic components of performance correlations

The concept of intra-individual variation (Wilson et al., 2014b) is similar to a counterfactual conditional statement such as, 'were the mechanical advantage 0.3 and not its real value 0.2, the force output would be 0.6 N and not its real value 0.4 N'. But there is nothing special about applying this concept to morpho-physiological traits as opposed to individual quality factors ('were my training 750 h per year and not its real value 500 h, my marathon time would be 2 h:18 min and not its real value 2 h:28 min'). We suggest that the masking problem is not one of inter- versus intra-individual variation but of intrinsic versus extrinsic variation, where extrinsic variation results from differences in exposure to extrinsic factors such as training intensity or style, diet, recovery, stressful life events, pathogens, etc. All intrinsic factors should be left unadjusted.

We developed a model of how functional trade-offs, which arise at the sub-whole-animal level, combine with extrinsic quality factors to contribute to performance correlations that we measured at

the whole-animal level. This model uses the graphical algebra of path models, which were specifically developed by Sewell Wright (1918, 1932, 1934) to model the underlying factors ('causes') generating correlations among measured traits. While elegant, path models are not necessary for any conclusion that we illustrate. The basic algebra of path models is available in some biostatistics textbooks (e.g. Sokal and Rohlf, 2012). Shipley (2002) is a more thorough introduction to path models in functional biology and ecology. Pearl (2009) formalizes many of the concepts of causal graphs. Our path models are models of simplified functional systems and used only to develop a general theory; we do not attempt to test detailed causal models of performance variation in humans. Our models ignore some complexities of real systems in order to focus on the fundamental principles. But ignoring the complexities of real systems does not make our simplified models irrelevant; on the contrary, the complexities make the goals of discovering underlying trade-offs that much more difficult. The scripts for generating simulated data using all of the path models introduced below are available elsewhere (Walker, 2015b).

Functional trade-offs are sometimes measured as a correlation between performance traits at the individual or evolutionary (among species) level, but these correlations are also influenced by non-functional factors (see below). We define a functional trade-off as a pattern of form–function mapping in which a morpho-physiological (M-P) trait causally affects two performance traits in opposite directions (Ghalambor et al., 2003). Our definition can be more precisely defined using the graphical algebra of the path diagram in which a single M-P trait M (say, percentage type I fibers in the biceps femoris muscle) has opposite causal effects (β) on the performance traits P_1 and P_2 (say 100 m and 1500 m speeds):

$$M$$
$$\beta_1 \swarrow \qquad \searrow \beta_2$$
$$U_1 \longrightarrow P_1 \qquad\qquad P_2 \longleftarrow U_2 \qquad (1)$$

where β_1 and β_2 have opposite signs. The U represent 'noise' or additional variance that is uncorrelated with all other effects. In all later path models, the U are implied but not written out.

We refer to this kind of path diagram as a form–function map. The pattern of causal arrows from M-P traits to performance traits is one component of the functional architecture of an organism. The single-headed arrow indicates a causal effect and the path coefficient represents the sign and magnitude of the effect. Here and elsewhere in this paper, all variables in a path model are standardized to unit variance, which makes the path coefficients standardized. The consequence of this standardization gives the path diagram its most elegant feature: the expected correlation between any two traits can be quickly computed as the sum of the products of the coefficients along all paths connecting the two variables. The expected performance correlation between P_1 and P_2 is $\beta_1\beta_2$. The performances are correlated because they share the common cause (M). The U do not contribute to the correlation because neither is a common cause of P_1 and P_2. We call the expected correlation due to the mapping of M-P traits to performance the 'functional correlation'. In this simple model, but not in more complex models (see below), the functional correlation is the expected performance correlation. A negative functional correlation is a functional trade-off and occurs if β_1 and β_2 have opposite signs. A positive functional correlation is a functional facilitation (Ghalambor et al., 2003).

We introduce a second causal variable Q, which represents sources of variation extrinsic to the organism (e.g. variation in training hours). Furthermore, let Q have positive effects (α) on both performances (e.g. the more one trains, the better one gets at both the 100 m and 1500 m events) that are independent of the effects of M:

$$M \quad\quad\quad\quad Q$$
$$\beta_1 \quad \beta_2 \quad \alpha_1 \quad \alpha_2$$
$$P_1 \quad\quad P_2 \tag{2}$$
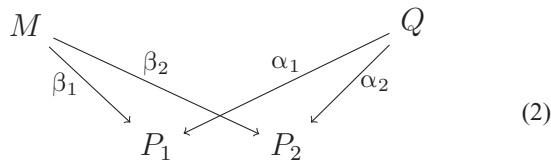
While it appears that Q affects performance by some mechanism other than through M-P traits, our model is a mathematical simplification of, but precisely equivalent to, a full model with Q acting through the M-P traits (Walker, 2015b).

The functional correlation in Eqn 2 is $\beta_1\beta_2$ but the expected performance correlation is $\alpha_1\alpha_2+\beta_1\beta_2$. The component $\alpha_1\alpha_2$ is positive; adding Q to the model shifts the expected performance correlation in a positive direction. Regardless of the magnitude of the α, the effect of the functional trade-off is masked. If the α are small relative to the β, only the magnitude will be masked (i.e. a smaller, negative performance correlation than would occur if there were no variation in training). But if the α are large relative to the β, the sign will be masked too.

If the α are anything but trivially small, the performance correlation fails to represent the expected correlation due to functional architecture. To estimate the functional correlation, one needs a dataset in which Q is constant or has very small variance. Ideally, these data would come from an experimental design in which the individuals were raised in a common environment. If Q contaminates a dataset, and the goal is to estimate a Q-free correlation, the path model in Eqn 2 suggests a simple and elegant solution:

$$r_{\text{quality-free}} = r_{\text{raw}} - \alpha_1\alpha_2, \tag{3}$$

where r_{raw} is the measured correlation. Below, we show that $r_{\text{quality-free}}$ is equivalent to the functional correlation only in the special case of no correlation among the M-P traits.

Now, we replace Q with a second M-P trait that causally affects both performances in the same direction; shorter heels, for example, increase both sprinting and endurance performance in humans (Scholz et al., 2008; Lee and Piazza, 2009).

$$M_1 \quad\quad\quad\quad M_2$$
$$\beta_{11} \quad \beta_{12} \quad \beta_{21} \quad \beta_{22}$$
$$P_1 \quad\quad P_2 \tag{4}$$
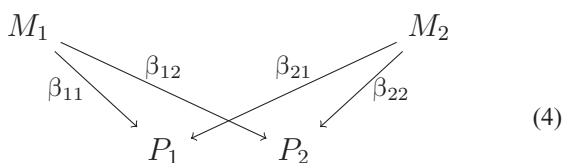
Note that we have changed the coefficient symbols to reflect the fact that our second causal variable represents an underlying M-P trait and not an extrinsic factor. The expected performance correlation ($\beta_{11}\beta_{12}+\beta_{21}\beta_{22}$) is the sum of two functional correlations, each due to its own common cause. We call the sum of the functional correlations the 'net' functional correlation. A net functional trade-off is a negative net functional correlation. It is a pattern of form–function mapping where a set of M-P traits causally affects two performance traits in opposite directions.

In Eqn 2, we have a functional system in which the underlying form–function mapping is being masked by Q. We quite reasonably consider Q a nuisance factor and want to adjust the raw correlation to get rid of its effect. By contrast, in Eqn 4, we have a functional system with two underlying M-P traits. Trait M_1 causes a trade-off (its path coefficients are of opposite sign). Trait M_2 causes a facilitation. Holzman et al. (2011), among others, emphasized that the M_2 facilitation mitigates the M_1 trade-off but it makes equal sense to say the M_1 trade-off attenuates the M_2 facilitation. If the facilitation is larger, the net functional correlation is positive. A positive performance correlation faithfully represents the underlying functional architecture; that is, the pattern of how M-P traits map to performance. We most definitely do not want to adjust the correlation to control for M_2. Instead, we have high- and low-quality athletes because of the functional architecture. Quality is determined by the intrinsic properties of the causal mapping from M-P traits to performance.

The contrast between the path diagrams in Eqns 2 and 4 raises the concern, what kinds of traits do we call Q (extrinsic quality) and consider a nuisance variable that should be statistically adjusted and what do we consider M (intrinsic quality) and part of the functional architecture? For example, recent work on the genetic predictors of the individual response to training in humans suggests the presence of networks of muscle-plasticity genes that affect the ability of muscle to remodel. These networks are activated in the same way in both resistance and endurance training (Timmons, 2011; Phillips et al., 2013). If these plasticity networks have large magnitude effects on performance, we would expect some individuals to excel in both high-power and high-endurance events and some to perform poorly in both types of events, even if all athletes had precisely the same training. This variation in the response to training is an intrinsic component of phenotypic design, as opposed to the extrinsic variation in different training plans. Do we measure this response to training, score it as Q and adjust for its effects on the correlations, or do we score it as another M-P trait M and allow it to contribute to the net functional correlations? Other traits, especially physical traits (pain and stress signaling systems related to tolerance, including CNS feedback limiting muscle strain and, consequently, performance) that contribute to psychological factors like 'mental toughness' or 'competitiveness' may be even more ambiguous as to how these should be considered. We believe all these intrinsic traits should be scored as M.

This concern is augmented if Q is a latent variable, as in Wilson et al. (2014b) and Van Damme et al. (2002). Latent variables are mathematical constructs (such as the first principal component) but are generally interpreted to be something meaningful, for example 'general intelligence', 'general size' or 'individual quality'. With a latent Q, how do we differentiate the system in Eqn 2 from that in Eqn 4, both of which result in a first PC with all positive loadings? Even worse, the first PC can be a mixture of intrinsic and extrinsic contributions to quality. In Walker (2015b), we generated simulated data representing 10 performance traits affected by two M-P traits and a single, extrinsic quality trait. The M-P traits had a modular mapping to performance with one having moderate effects on performance traits 1–5 and the other having moderate effects on performance traits 6–10. The quality trait had a small effect on six of the 10 performance traits and zero effect on the other four performance traits. The loadings on PC1 were all the same sign, suggestive of a single, global factor even though none existed. While a PC1 with all positive loadings is an axis representing athletic quality, a pattern of all-positive loadings cannot justify interpreting the axis as representing extrinsic sources of variation that need to be adjusted away.

Why performance correlations may be poorly predicted

Predicting a correlation between a pair of performance traits based on a qualitative relationship between one or a few underlying M-P traits and the performance traits is nearly ubiquitous in the performance literature (e.g. increased upper limb muscle mass should increase shot put but cost high jump performance). But a net functional correlation is the sum of all the expected component correlations (Ghalambor et al., 2003). For example, in the path diagram in Eqn 4, the net functional correlation is $\beta_{11}\beta_{12}+\beta_{21}\beta_{22}$. This raises a major concern with bottom-up approaches to predicting correlations between performance variables. Many of the performance traits that we care about, especially at the whole-animal level, have many tens, hundreds or even thousands of underlying common causal factors, many of which we are ignorant of, and for most of which we have little information on the magnitude of the effect on each performance, and thus the expected component correlation. Given our knowledge of only a fraction of the common causes, and quantitative estimate of effect size for only a fraction of these, it would seem our ability to predict a performance correlation is extremely limited. Vanhooydonck et al. (2014) make a similar argument but without the benefit of a graphical or mathematical model.

In addition to our general ignorance of the form–function map, performance correlations are confounded by correlations among M-P traits. In Eqn 4, the pattern of trade-offs and facilitations in the form–function map is the only factor contributing to correlations among performance traits and, as a consequence, the expected correlation between the performance traits (the performance correlation) is the net functional correlation. This expectation is a reason that functional trade-offs are frequently measured as a correlation between performance traits at the individual or evolutionary (among species) level. A performance trade-off, however, is determined by both the pattern of form–function mapping and the correlations between M-P traits:

$$(5)$$
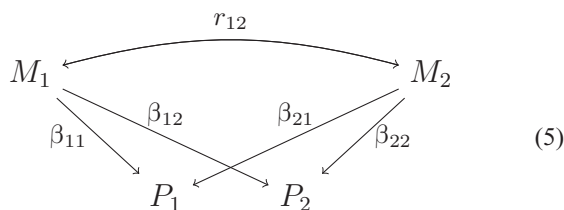
where r_{12} is the correlation between the M-P traits. The expected performance correlation is $\beta_{11}\beta_{12}+\beta_{21}\beta_{22}+r_{12}\beta_{11}\beta_{22}+r_{12}\beta_{12}\beta_{21}$. This correlation is not only due to both the sign and magnitude of the net functional correlation but also to the sign and magnitude of the phenotypic correlation between the M-P traits. A phenotypic correlation can cause a performance correlation to be larger, smaller or even of opposite sign relative to the net functional correlation (Holzman et al., 2011; Walker, 2015b). Indeed, if M_1 maps only to P_1, M_2 maps only to P_2 and r_{12} is negative, the expected performance correlation is negative despite the lack of any underlying M-P trait that has opposite effects on performance.

A brief introduction to estimating a quality-free correlation matrix

When we ask, 'is there a functional trade-off between two performances', we want to know whether the net functional correlation is negative. A net functional correlation is the expected correlation between two performance variables due only to the pattern of causal mapping from M-P traits to performance. It will generally not be equivalent to the expected performance correlation, which includes additional components due to correlations among the M-P traits and to extrinsic effects. Above, we introduced both bottom-up and top-down methods for estimating a net functional correlation. The bottom-up approach can be generalized to a $m \times m$ matrix of net functional correlations among m performance variables using (Ghalambor et al., 2003):

$$\mathbf{R}_{\text{func}} = \mathbf{F}^{\top}\mathbf{F}, \tag{6}$$

where $\mathbf{F}$ is the $p \times m$ matrix containing the path coefficients of the causal effects of p M-P traits on m performances. The matrix contains the expected trade-offs and facilitations among the performance variables at the whole-animal level given only the functional mapping. The problem with this bottom-up approach, as discussed above, is that we would need to know all causal effects to compute the net functional correlations with any accuracy.

The top-down approach is the matrix formulation of Eqn 3:

$$\mathbf{R}_{\text{quality-free}} = \mathbf{R}_{\text{raw}} - \hat{\boldsymbol{\alpha}}_Q\hat{\boldsymbol{\alpha}}_Q^{\top}, \tag{7}$$

where $\hat{\boldsymbol{\alpha}}_Q$ is the vector of path coefficients from some measure of Q to each performance variable. If $\hat{\boldsymbol{\alpha}}_Q$ is estimated as the loadings on PC1, then Eqn 7 is equivalent to the covariance matrix of the residuals of the regression of each performance trait on PC1. The matrix $\hat{\boldsymbol{\alpha}}_Q\hat{\boldsymbol{\alpha}}_Q^{\top}$ contains the component correlations due to Q only. As an estimate of the net functional correlation, Eqn 6 assumes (1) Q contains only extrinsic variables, (2) Q contains all extrinsic variables, (3) $\hat{\boldsymbol{\alpha}}_Q$ is an unbiased estimate of the effects of Q on performance, and (4) there is no phenotypic correlation among the morphological traits. We introduced assumptions 1, 2 and 4 above. Assumption 3 is violated using standard measures of a latent Q and we offer a solution here. Because assumption 4 is violated, we use the term 'quality-free' and not 'net functional correlation' to refer to the matrix $\mathbf{R}_{\text{quality-free}}$.

A bias-correction for Eqn 7 was provided by Wright (1932) in order to estimate 'size-free' correlations among a set of morphometric traits in which the underlying pattern of correlations is masked by 'general size' variation among individuals. Wright used the loadings of the traits on PC1 ($\hat{\boldsymbol{\alpha}}_{\text{PC1}}$) (although he did not use the term 'principal component') as estimates of $\boldsymbol{\alpha}_Q$ (Mitteroecker and Bookstein, 2009). Given the matrix of expected correlations due only to the individual quality (or general size) effects estimated by PC1 ($\mathbf{R}_{\text{PC1}} = \hat{\boldsymbol{\alpha}}_{\text{PC1}}\hat{\boldsymbol{\alpha}}_{\text{PC1}}^{\top}$) the bias-corrected quality (or size)-free correlations are:

$$\mathbf{R}_{\text{bc}} = \mathbf{R}_{\text{raw}} - \mathbf{R}_{\text{PC1}}\frac{\bar{r}_{\text{raw}}}{\bar{r}_{\text{PC1}}}, \tag{8}$$

where $\bar{r}_{\text{raw}}$ and $\bar{r}_{\text{PC1}}$ are the mean off-diagonal elements in $\mathbf{R}_{\text{raw}}$ and $\mathbf{R}_{\text{PC1}}$, respectively. We note that Wright (1932) used only the subset of 'among-module' correlations in the computation of $\bar{r}_{\text{raw}}$ and $\bar{r}_{\text{PC1}}$, where 'among-module' refers to a pair of morphometric traits occurring in different development modules defined *a priori*. Here, we relax the necessity of an *a priori* factor structure and use the means of all off-diagonal elements in $\mathbf{R}_{\text{raw}}$ and $\mathbf{R}_{\text{PC1}}$. We refer to the uncorrected (Eqn 7) and bias-corrected (Eqn 8) residuals as Wright's uncorrected (WUC) and Wright's bias-corrected (WBC) correlations.

MATERIALS AND METHODS

Decathlon data from male, college athletes (NCAA), generally aged 18–24 years, were collected from Track and Field Results and Reporting System (http://www.tfrrs.org/). Following the 2014 outdoor season, we collected personal-best values from each of the decathlon events for each athlete across all seasons recorded for that athlete. We allowed recorded personal bests for different events to occur in different seasons for the same athlete. While there are cogent arguments for analyzing sample mean and not maximum performance (Head et al., 2011), we used the maxima because these were accessible without mining data from individual meets. Data were collected for divisions I, II and III and combined into a single dataset of personal bests. To standardize performance direction such that larger values indicate better performance, times for the 100, 400 and 1500 m and the 110 m hurdles were converted to speed in m s^{-1}. All athletes with any missing event were removed from the dataset, leaving N=611 athletes with complete data.

We computed the partial correlations conditional on all other performance variables (Van Damme et al., 2002), regression residual correlations (Wilson et al., 2014b), and WUC and WBC residual correlations of the decathlon data. Standard statistical packages do not give the correct error of these correlations as estimates of quality-free correlations. For example, one can get the error statistics (standard errors, P-values, confidence intervals) for the regression residual correlations from any statistics package and these error statistics are correct for these correlations as estimates of the correlation of regression residuals but as a regression residual correlation is a biased estimate of the true quality-free (or PC1-free) correlation they do not give the correct error for the quality-free correlation.

To estimate the error of the quality-free correlations, we implemented a Monte Carlo simulation. Specifically, we modeled the process generating the NCAA data by simulating the causal mapping from m=19 functional 'factors' to p=10 decathlon performance traits for N=611 individuals using:

$$\mathbf{P} = \mathbf{q}\boldsymbol{\alpha}^{\top} + \mathbf{MB} + \mathbf{U}, \qquad (9)$$

where $\mathbf{P}$ is the $n{\times}p$ matrix of centered and variance-standardized simulated performance traits, $\mathbf{q}$ is the $n{\times}1$ matrix of random, normal variables that represents athletic quality for the n individuals, $\boldsymbol{\alpha}$ is the $p{\times}1$ matrix of variance-standardized causal effects (path coefficients) of quality on each performance trait, $\mathbf{M}$ is the $n{\times}m$ matrix of centered and variance-standardized random, normal variables that represent the M-P traits, $\mathbf{B}$ is the $m{\times}p$ matrix of variance-standardized causal effects of the m M-P traits on each performance trait, and $\mathbf{U}$ is an $n{\times}p$ 'unexplained error' matrix of random, normal error with column standard deviations scaled so that the columns of $\mathbf{P}$ have unit variance.

We parameterized Eqn 9 by substituting $\hat{\boldsymbol{\alpha}}_{bc}$ for $\boldsymbol{\alpha}$, where:

$$\hat{\boldsymbol{\alpha}}_{bc} = \hat{\boldsymbol{\alpha}}_{PC1}\sqrt{\frac{\bar{r}_{raw}}{\bar{r}_{PC1}}}. \qquad (10)$$

Again, $\hat{\boldsymbol{\alpha}}_{PC1}$ is the vector of loadings on PC1, $\bar{r}_{raw}$ is the mean off-diagonal correlation in the Pearson correlation matrix, and $\bar{r}_{PC1}$ is the mean off-diagonal correlation in the matrix of expected correlations due to PC1 (see Eqn 8). The β coefficients in $\mathbf{B}$ are random, exponential with rate (λ) equal to 1 and rescaled so that each element of the diagonal of $\boldsymbol{\alpha}\boldsymbol{\alpha}^{\top} + \mathbf{B}^{\top}\mathbf{B}$ is equal to R^2=0.95, the percentage of the total variance explained by the simulated quality factor and M-P traits that causally effect performance. The number of M-P traits (m) was set to 19 because this number generated a distribution (mean and s.d.) of the correlations in the size-free correlation matrix similar to that estimated by WBC correlations of the decathlon data (Fig. S1).

The true matrix of quality-free correlations of the modeled data $\mathbf{P}$ is:

$$\mathbf{R}_{qf} = \mathbf{B}^{\top}\mathbf{B}. \qquad (11)$$

We generated 2500 simulated datasets and computed for each the partial correlations, regression residual correlations, and WUC and WBC correlations. We then computed the error from the true correlations, $\hat{\mathbf{R}} - \mathbf{R}_{qf}$ for each simulated dataset, where $\hat{\mathbf{R}}$ is the estimated matrix using any of the four methods.

The complete set of scripts to process the data were written in R v3.1.2 (R Core Team, 2014) and are available as on GitHub at https://github.com/middleprofessor/NCAA_decathlon.

RESULTS

We computed raw and quality-adjusted correlations among the 10 decathlon events for 611 college athletes. To facilitate interpretation, the events are organized into blocks containing functionally similar events (throws, jumps, runs). All 45 of the Pearson product-moment correlations among the 10 events are positive and are generally of moderate to large magnitude (Fig. 2, Table 1) with a range of 0.2 to 0.84. The Pearson correlation heat map (Fig. 2) shows slightly higher correlations within than among the functional blocks (cells closer to the diagonal tend to be darker). The smallest correlations are between 1500 m speed and all other events except 400 m speed. The pole vault also has noticeably smaller correlations with other events, including other jumps.

Quality-adjusted correlations are shown in Fig. 3. In general, the partial correlations are small and positive with only one moderately negative estimate, that between 1500 m and 100 m speeds. This result is consistent with the results of Van Damme et al. (2002) although their table 1 reports only the sign if statistically significant or 'NS' if not. By contrast, the regression-residual correlations are almost all negative among functional blocks. Within functional blocks, the regression-residual correlations are positive within the throws block, mostly positive within the runs block, and mostly negative within the jumps block. The overall trend in correlations is nearly opposite between partial and regression residual correlations. The two sets of correlations, however, are orthogonal. That is, the large, positive estimates using partial correlation tend to be near zero using regression residual correlation, while the estimates near zero
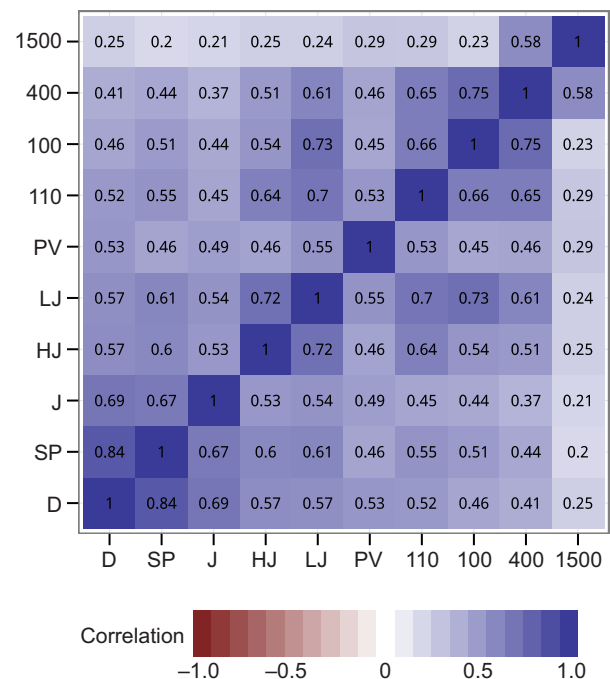
Fig. 2. Heat map depicting the sign and magnitude of the Pearson product-moment correlations among the 10 decathlon events for the 611 NCAA athletes. The map shows that all correlations in this matrix are positive and of moderate to large value. Events: 1500 m speed, 400 m speed, 100 m speed, 110 m hurdles speed, pole vault (PV), long jump (LJ), high jump (HJ), javelin (J), shot put (SP) and discus (D).

Table 1. Correlations for all pairwise combinations of events using the method of Pearson (Pearson), partial correlations conditional on all other events (PCOR), regression residual correlation (RR), Wright's uncorrected correlations (WUC) and Wright's bias-corrected correlations (WBC)

Event 1	Event 2	Pearson	PCOR	RR	WUC	WBC	Lower WBC	Upper WBC
Shot put	Discus	0.84	0.66	0.58	0.21	0.26	0.15	0.37
Javelin	Discus	0.69	0.24	0.28	0.12	0.16	0.06	0.27
Javelin	Shot put	0.67	0.18	0.22	0.09	0.14	0.03	0.24
High jump	Discus	0.57	0.07	−0.13	−0.05	0	−0.09	0.1
High jump	Shot put	0.6	0.08	−0.09	−0.03	0.02	−0.08	0.12
High jump	Javelin	0.53	0.08	−0.12	−0.05	0	−0.09	0.1
Long jump	Discus	0.57	−0.03	−0.34	−0.11	−0.05	−0.14	0.05
Long jump	Shot put	0.61	0.09	−0.24	−0.07	−0.02	−0.11	0.08
Long jump	Javelin	0.54	0.07	−0.22	−0.08	−0.03	−0.12	0.07
Long jump	High jump	0.72	0.39	0.14	0.05	0.1	0	0.21
Pole vault	Discus	0.53	0.2	−0.05	−0.02	0.03	−0.07	0.13
Pole vault	Shot put	0.46	−0.11	−0.22	−0.09	−0.05	−0.14	0.05
Pole vault	Javelin	0.49	0.15	−0.03	−0.01	0.03	−0.07	0.13
Pole vault	High jump	0.46	−0.02	−0.19	−0.08	−0.04	−0.13	0.06
Pole vault	Long jump	0.55	0.15	−0.13	−0.05	0	−0.09	0.1
Speed 110 h	Discus	0.52	0.02	−0.34	−0.12	−0.07	−0.16	0.03
Speed 110 h	Shot put	0.55	0.06	−0.29	−0.1	−0.05	−0.14	0.05
Speed 110 h	Javelin	0.45	−0.04	−0.34	−0.14	−0.09	−0.17	0.01
Speed 110 h	High jump	0.64	0.21	0	0	0.05	−0.05	0.15
Speed 110 h	Long jump	0.7	0.16	0.01	0	0.06	−0.04	0.16
Speed 110 h	Pole vault	0.53	0.16	−0.08	−0.03	0.01	−0.08	0.11
Speed 100 m	Discus	0.46	0	−0.43	−0.16	−0.11	−0.19	−0.02
Speed 100 m	Shot put	0.51	0.03	−0.32	−0.12	−0.07	−0.15	0.03
Speed 100 m	Javelin	0.44	0.07	−0.3	−0.13	−0.08	−0.17	0.02
Speed 100 m	High jump	0.54	−0.09	−0.22	−0.08	−0.03	−0.12	0.07
Speed 100 m	Long jump	0.73	0.37	0.19	0.06	0.12	0.01	0.22
Speed 100 m	Pole vault	0.45	−0.02	−0.21	−0.09	−0.05	−0.14	0.05
Speed 100 m	Speed 110 h	0.66	0.1	0.06	0.02	0.08	−0.02	0.18
Speed 400 m	Discus	0.41	−0.06	−0.48	−0.19	−0.14	−0.22	−0.04
Speed 400 m	Shot put	0.44	0.05	−0.43	−0.17	−0.12	−0.2	−0.02
Speed 400 m	Javelin	0.37	−0.07	−0.41	−0.18	−0.14	−0.22	−0.04
Speed 400 m	High jump	0.51	0.04	−0.24	−0.09	−0.04	−0.13	0.05
Speed 400 m	Long jump	0.61	−0.01	−0.12	−0.04	0.01	−0.08	0.11
Speed 400 m	Pole vault	0.46	0.04	−0.15	−0.07	−0.03	−0.12	0.07
Speed 400 m	Speed 110 h	0.65	0.23	0.09	0.03	0.09	−0.01	0.19
Speed 400 m	Speed 100 m	0.75	0.58	0.38	0.15	0.2	0.09	0.31
Speed 1500 m	Discus	0.25	0.11	−0.15	−0.08	−0.06	−0.14	0.04
Speed 1500 m	Shot put	0.2	−0.09	−0.26	−0.14	−0.11	−0.2	−0.02
Speed 1500 m	Javelin	0.21	0.05	−0.15	−0.09	−0.07	−0.15	0.03
Speed 1500 m	High jump	0.25	0.01	−0.16	−0.09	−0.06	−0.15	0.04
Speed 1500 m	Long jump	0.24	−0.01	−0.25	−0.12	−0.09	−0.17	0.01
Speed 1500 m	Pole vault	0.29	0.07	0	0	0.02	−0.07	0.12
Speed 1500 m	Speed 110 h	0.29	−0.07	−0.1	−0.06	−0.03	−0.12	0.07
Speed 1500 m	Speed 100 m	0.23	−0.33	−0.18	−0.1	−0.07	−0.16	0.02
Speed 1500 m	Speed 400 m	0.58	0.61	0.44	0.26	0.28	0.16	0.39

In addition, the lower and upper bounds of the 95% error intervals of the WBC estimates are given.

using partial correlation tend to be large and negative using regression residual correlation.

The WUC correlations show a pattern similar to the regression-residual correlations except that all values are closer to zero. This must be the case as the uncorrected residual correlations are equivalent to the covariance of the regression residuals and the variances of the residuals are all less than one.

The WBC correlations are slightly shifted in a positive direction relative to the WUC correlations. Again, this is expected because the algorithm corrects the negative bias in the uncorrected estimates. The consequence is that the negative correlations are slightly less negative while the positive correlations are slightly more positive. The correlations within the throws and the runs functional blocks are small to moderate and positive, while the correlations within the jumps functional block are very small and positive. One exception in this pattern is the small, negative correlation ($r = −0.07$) between 1500 m and 100 m speeds. There are small trade-offs among the

throws and runs functional blocks. The largest trade-offs (r from −0.12 to −0.14) are between 400 m speed and the throws. Otherwise, among-block correlations are very small (in both positive and negative directions). The biased (PC1 loadings) and bias-corrected α to compute WUC and WBC correlations are given in Table 2.

We used Monte Carlo simulation to estimate the error distribution under the assumption that the data are log-normally distributed and with a pattern of correlations similar to that of the NCAA data (Fig. 4A). There is a positive bias and large variance in the error for the partial correlations (mean±s.d., 0.10±0.22) and a negative bias and moderate variance for the regression residual error (−0.11±0.14). There is a small negative bias and small variance in the WUC errors (−0.05±0.05). There is effectively no bias and small variance in the WBC errors (<0.01±0.05). In Fig. 4B, we show the true correlations as a function of the WBC estimate. The 2.5%, 50% and 97.5% quantile regression lines are also shown. We used the lower and upper quantile regression functions to
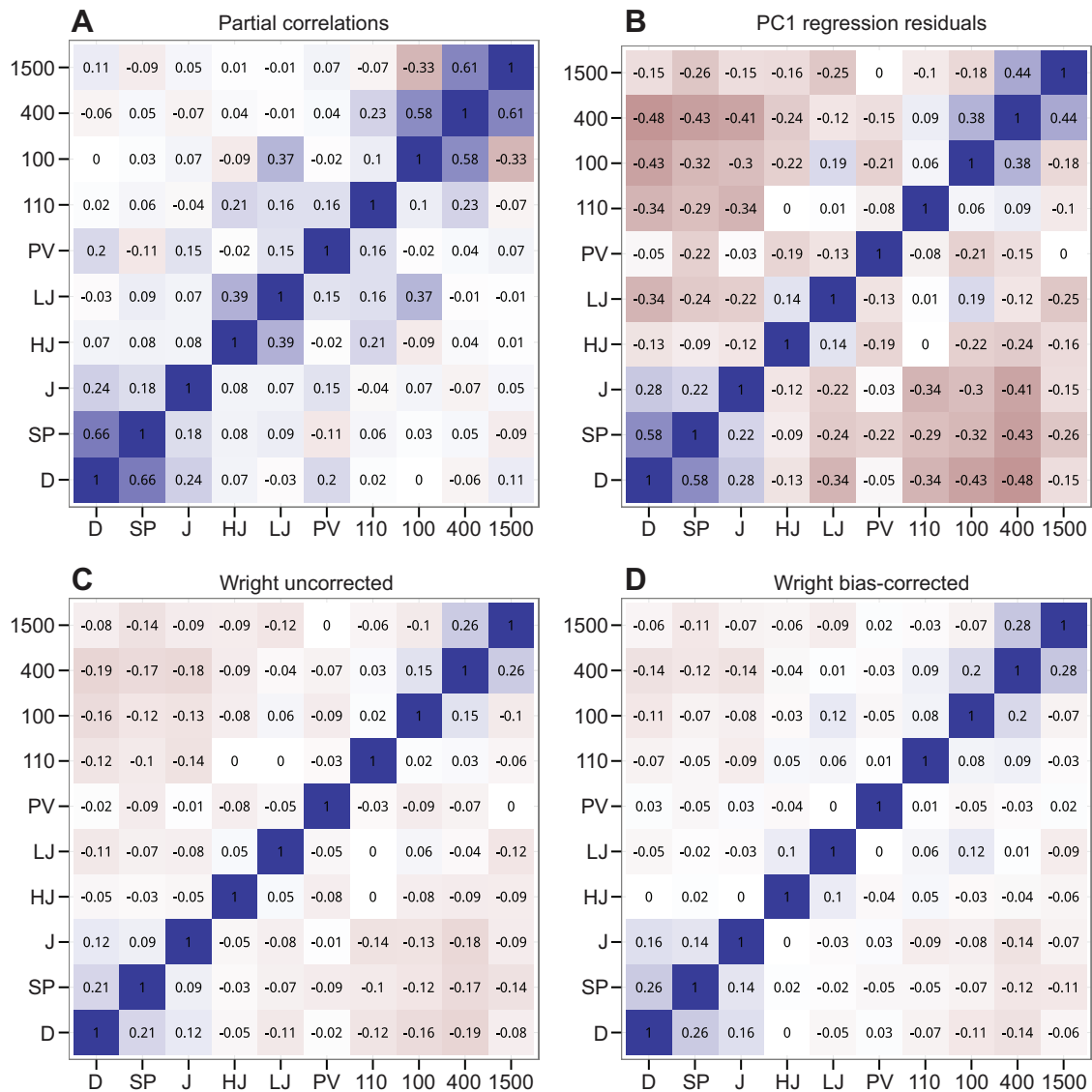
Fig. 3. Heat maps depicting the sign and magnitude of the quality-adjusted correlations among the 10 decathlon events for the 611 NCAA athletes. Positive correlations are blue, negative correlations are red, color intensity reflects the magnitude of the correlation (see legend in Fig. 2). (A) Partial correlations conditional on all additional performance variables. (B) Regression residual correlations computed using the residuals of each performance trait regressed on the first principal component (PC1) scores. (C) Wright's uncorrected correlations. The values are equivalent to the covariances of the regression residuals used in B. (D) Wright's bias-corrected correlations. Note that the running performances are in units of speed and not duration. Event abbreviations as in Fig. 2.

compute the 95% error intervals of the WBC estimate in Table 1. These error intervals are not the long-run probability intervals of the correlation coefficient but the 95% bounds of true correlation coefficients that 'generate' the observed coefficient (Fig. 4B).

Table 2. Estimate of the quality (Q) path-coefficient α

Event	$\hat{\alpha}$	$\hat{\alpha}_{bc}$
Speed 100 m	0.78	0.75
Speed 400 m	0.76	0.73
Speed 1500 m	0.42	0.4
Speed 110 h	0.81	0.78
High jump	0.79	0.76
Long jump	0.86	0.82
Pole vault	0.69	0.66
Shot put	0.8	0.77
Discus	0.79	0.76
Javelin	0.73	0.7

The consequences of the effect sizes of the WBC quality-free correlations can be explored by computing what-if scenarios. For example, if we could intervene and shift M-P trait values in the direction that would cause 1500 m speeds to increase by two standard deviations, equivalent to running 42.7 s faster (i.e. from middle of the pack among all NCAA decathletes to top 2.5%), 100 m times would slow by only 0.064 s. In the 2014 division I championship, this intervention would drop the 100 m placing by an average of 1.5 places. This effect is not trivial but is quite small given the huge intervention in 1500 m time.

In addition to analyzing the NCAA decathlon data, we re-analyzed the soccer data of Wilson et al. (2014b). As shown in the original study, all five performances are positively correlated (Fig. 5A; Table S1). As with the NCAA data, the partial correlations and regression residual correlations are largely orthogonal, and result in radically different interpretations (Fig. 5B,C). The WBC estimates are small and of both signs (Fig. 5D). Notably, the WBC
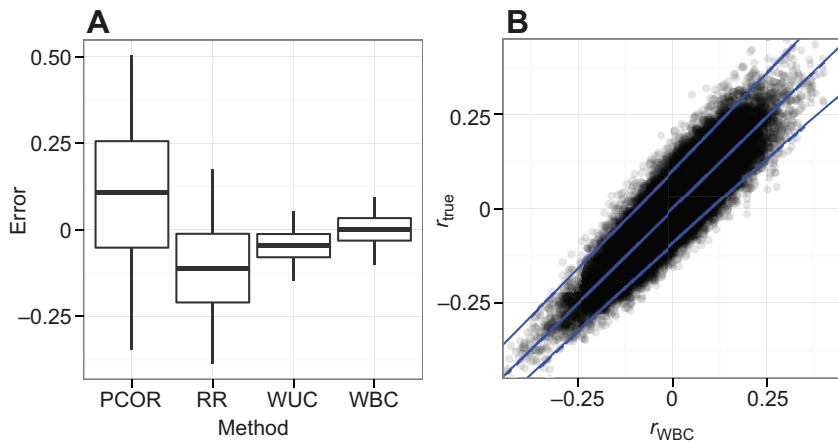
Fig. 4. The distribution of the errors estimated by Monte Carlo simulation of the NCAA decathlon data. (A) Box plots with the median, and 50% and 95% intervals represented by the horizontal line, the box and the tips of the whiskers. PCOR partial correlations conditional on all other events; RR, regression residual correlation; WUC, Wright's uncorrected correlations; and WBC, Wright's bias-corrected correlations. (B) True quality-free correlation as a function of the Wright's bias-corrected estimated correlation. The blue lines are the 2.5%, 50% and 97.5% quantile regressions used to compute the 95% confidence intervals in Table 1.

estimate (0.11) (Table S1) of the WBC correlation between 1500 m and 40 m sprint is in the direction opposite to the expectation of a trade-off between high-endurance and high-power performances. That said, the error associated with the WBC estimates is too large to have much confidence in their sign (Fig. S2).

DISCUSSION

We found only positive correlations among the 10 events of the decathlon. As a consequence, the PC1 of these data describes an axis of athletic quality; athletes with high PC1 scores perform better at all events. These results are consistent with prior work on both humans (Van Damme et al., 2002; Wilson et al., 2014b) and non-human vertebrates (Reidy et al., 2000; Marras et al., 2013; Vanhooydonck et al., 2014). We used a slight modification of

Wright-style factor analysis with bias-correction (Wright, 1932) to estimate quality-free correlations. Elsewhere (Walker, 2015a), we have shown mathematically and with simulation that all four methods used by Van Damme et al. (2002) and Wilson et al. (2014b) result in poor inferences of underlying functional trade-offs. This can also be shown empirically, as a simple comparison of the partial and regression correlations shows starkly different patterns (Figs 3, 5). The partial correlation and regression residual correlation were the major methods used by Van Damme et al. (2002) and Wilson et al. (2014b), respectively. Clearly, both cannot be correct. Indeed, both are incorrect; the partial correlations are shifted in a positive direction while the regression residual correlations are shifted in a negative direction (Fig. 4).
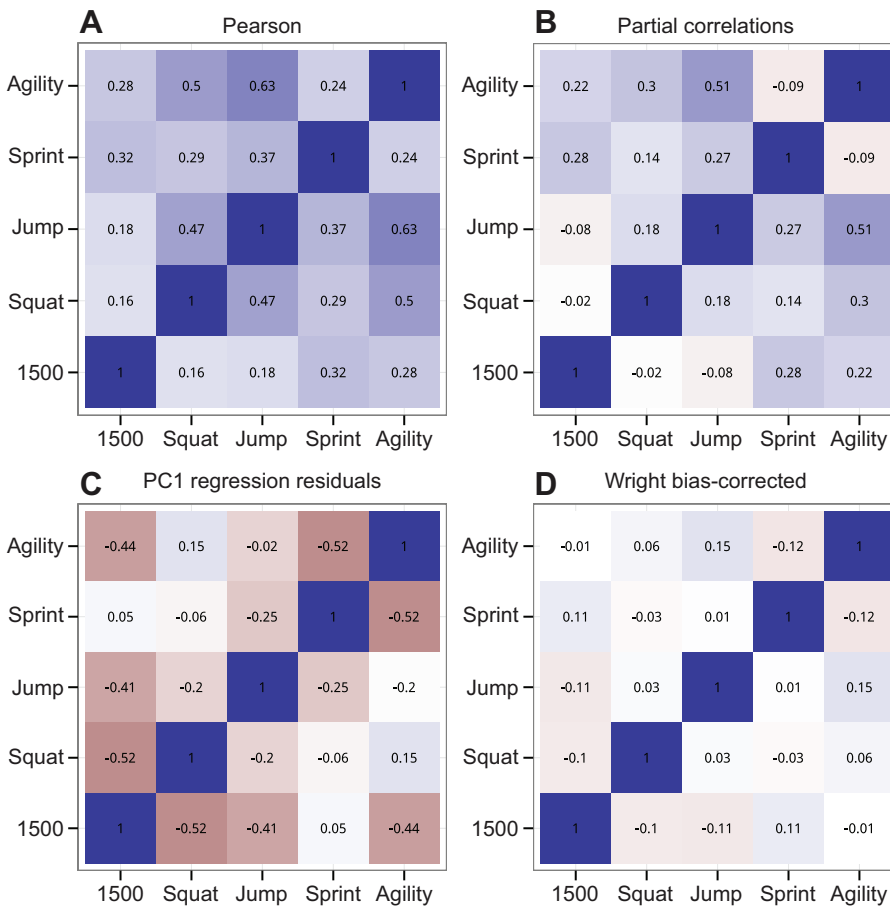
Fig. 5. Heat maps depicting the sign and magnitude of the quality-adjusted correlations among the five performances measured for 28 soccer players. Positive correlations are blue, negative correlations are red, color intensity reflects the magnitude of the correlation (see legend in Fig. 2). (A) Pearson product-moment correlations. (B) Partial correlations conditional on all additional performance variables. (C) Regression residual correlations computed using the residuals of each performance trait regressed on the PC1 scores. (D) Wright's bias-corrected correlations.

An important but unresolved question is: what is the effect of measurement error on our estimates of the quality-adjusted correlations? Correlations between variables measured with error are biased toward zero and the magnitude of this bias is a function of the repeatability and number of replicates per individual (Adolph and Hardin, 2007; Adolph and Pickering, 2008). Repeatability in our data varies between 0.81 and 0.86 for all events except 1500 m speed (0.77), which is relatively high but also computed from the best performances for each outdoor track season. Because our data are sample maxima (and not sample means), standard corrections for correlation attenuation would also likely contain some unknown error (Head et al., 2011). Even if we had estimated sample mean performances for each individual, a standard correction may not be applicable to the quality-free correlations that we attempt to interpret as these are missing the component correlation due to a global (quality) factor and we do not know how the bias is decomposed and distributed across multiple components. If the bias affects only the global (quality) component, then our quality-free correlations are not attenuated by measurement-error bias.

The predictability of performance correlations

'Physiological and biomechanical theory predicts that there should be trade-offs between certain pairs [of decathlon event performances] – for example, speed depends on the athlete having a high proportion of fast, fatigue-sensitive muscle fibres, whereas endurance relies on a higher proportion of slower fibres that are more resistant to fatigue' (Van Damme et al., 2002). In the Introduction, we showed why we will often fail to predict performance correlations given our limited knowledge of the form–function map. But the general strategy of using a P-value between performance correlations as a test of qualitative models of form–function mapping suffers from two additional flaws. First, any test of the hypothesis that a correlation is equal to zero can be rejected without statistics. All performance traits are jointly affected by multiple underlying causal factors and must co-vary. A P-value tells us nothing more than the adequacy of our sample size to reject the null. When the null is rejected, the magnitude of the correlation coefficient is rarely compared to some prediction based on a quantitative model; in fact the magnitude is typically ignored and frequently not even reported. Instead, the sign is compared to an expectation based on well-known trade-offs at the cell or organ level, such as muscle fiber type or gearing ratio. But predicting the sign of a correlation is hardly a severe test; the flip of a fair coin will predict the sign of a correlation between any two performances 50% of the time.

Second, a rigorous model of the causal basis of performance correlations requires accurate and precise estimates of causal effects. Almost all causal effects of M-P traits on performance are estimated using regression coefficients from observational data. Many biases, such as omitted confounders, necessarily infect observational designs (Shalizi, 2013) but the most commonly used biostatistics textbooks in comparative physiology fail to describe these biases. Consequently, few physiologists recognize how sensitive regression coefficients are to missing confounders or that standard errors from statistical packages are not the proper standard error if interpreting regression coefficients causally. The sensitivity of regression estimates to missing confounders using realistic causal models was not systematically explored until recently (Walker, 2014). Because standard errors from statistical packages model only sampling error, and error variance due to missing confounders does not decrease with sample size, increasing sample size simply gives the investigator increasingly false P-values (Walker, 2014). Any

rigorous estimate of causal effects will require some combination of computational modeling, direct phenotypic manipulation, and indirect manipulation via regulation of gene expression (Wang et al., 2004). It is hard to over-emphasize these two points because they are contrary to statistical training in much of biology.

What facilitations or quality do not mask

If individual quality masks functional trade-offs, these trade-offs can still emerge at the level of among-individual performance. For example, if training specifically targets M_1, which has opposite effects on performances P_1 and P_2, one performance will increase at the cost of the other, even if there is a positive correlation among P_1 and P_2 arising from a large quality effect. To show this, we parameterize Eqn 2 with a small functional trade-off and a large functional facilitation (Walker, 2015b).

$$
\begin{array}{ccc}
M_1 & & Q \\
-0.4 \;\; 0.4 & \;\; 0.7 \;\; 0.7 & \\
P_1 & & P_2
\end{array}
\tag{12}
$$
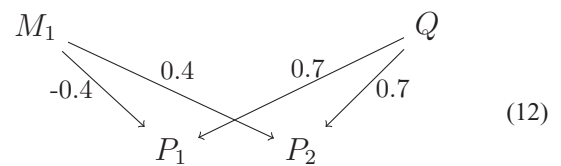
The expected performance correlation is 0.33. If we intervene and increase the mean value of M_1 by 0.5, then we expect P_2 to increase by 0.2 units but P_1 to decrease by 0.2 units. The expected response is $\Delta M_i \beta_{ij}$, where ΔM_i is the effect of the intervention on M_i and β_{ij} is the causal effect of M_i on P_j, which shows that the magnitude of the trade-off in the response is independent of the mapping of Q to performance. This result is general; that is, the masking trait Q could be a M-P trait M_2 (Eqn 2), and only assumes that the intervention is specific to M_1. The intervention could be experimental (including artificial selection) or arise from different reaction norms (strong in M_1, weak in M_2) to an environmental stimulus (including specific training in humans). Variation in the stimulus generating ΔM_i is an extrinsic factor while variation in ΔM_i due to the same stimulus is an intrinsic factor.

Intrinsic and extrinsic quality

Above, we presented the intrinsic and extrinsic quality models of variation in athletic quality, both of which will generate positive correlations among performances and a PC1 with all positive loadings. The essence of each model is captured in Eqns 2 and 4. In the extrinsic quality model, Q represents extrinsic features such as training history or health status. In the intrinsic quality model, the causal variables represent underlying features of neuromuscular and musculoskeletal design that have net agonistic rather than antagonistic effects on the performances. These models are the extremes of a continuum. Walker (2015b) developed a model in which PC1 had all positive loadings but there was no general factor generating the data; instead, PC1 was causally generated by both the extrinsic Q and two intrinsic M-P traits. With this kind of generating model, WBC correlations are the wrong answer to the question 'what are the correlations adjusted for extrinsic quality?' Or, the right answer (a quality-free correlation matrix) to a question that we do not care about, 'what is the residual correlation after factoring out both intrinsic and extrinsic quality?' Even more likely is an interaction between intrinsic and extrinsic factors. For example, there could be underlying M-P traits that increase both high-power and high-endurance performances, which, in turn, gives the athlete the

confidence to train harder, recover better, eat better and avoid behavior that might increase infection. Or, the intrinsically 'good athletes' with underlying M-P traits that make them better at both high-power and high-endurance performances are recruited to schools and coaches that have better training programs.

A necessary consequence of the above argument is our inability to use either a matrix of positive correlations or its decomposition by PCA to distinguish between the intrinsic and extrinsic quality models. This is the challenge introduced earlier. Extrinsic quality we want to control; intrinsic quality we do not – it is what we want to discover! But if we cannot know what causally generates PC1, we cannot know what we are controlling by removing the effect of PC1. Ideally, we would experimentally control extrinsic quality by raising individuals in a common-garden design where variation in the level of training, diet, temperature, parasites and pathogens is the same across individuals. Such a design is really only possible with laboratory animals.

Limiting performance measures to elite athletes must greatly reduce the contribution of extrinsic quality components to PC1. If the extrinsic quality contributions to the variance on the PC1 are indeed small relative to intrinsic quality contributions, then the positive correlations measured between the decathlon events should largely be faithful estimates of the functional organization of the causal mapping from M-P traits to performance. These positive correlations might reflect correlated, intrinsic variation in the ability of athletes to predict maximum, safe and sustainable strain and strain rates for an activity, unconscious CNS control of maximum strain and strain rates, and CNS coordination of complex motor behaviors required for skilled movement (Tucker, 2009; Martens and Collier, 2011).

If, however, extrinsic quality contributions to PC1 are large, then we have some evidence of a few, small trade-offs in the decathlon data, including a small trade-off between high endurance (1500 m speed) and high power (sprint speed). Small effects are potentially important in both evolutionary and ecological dynamics, so we do not trivialize the small effect size (-0.07) of the quality-free correlation between 1500 m and 100 m speeds (indeed, we give an example of how it affects placing in the 100 m finals of the NCAA championships). Nevertheless, we do not believe the small, negative correlations in the quality-adjusted decathlon data are good evidence for an underlying trade-off generated by the architecture of the form–function map. The issue is not so much the magnitude of these correlations but their meaning. The quality-free correlations are adjusted for quality, but this quality likely contains both intrinsic and extrinsic components and a PCA cannot quantify their relative magnitudes. Consequently, we strongly caution against citing our results as evidence of net functional trade-offs free of extrinsic quality. The high frequency of positive correlations among both human and non-human performance traits that are putatively optimized by opposing musculoskeletal designs, such as muscle fiber type, musculoskeletal gearing ratios or body shape, is surprising, and individual quality, with its intrinsic and extrinsic components, is a compelling model to explain these correlations. The physiological explanation of quality, and of the positive correlations more generally, remains to be discovered.

References
Adolph, S. C. and Hardin, J. S. (2007). Estimating phenotypic correlations: correcting for bias due to intraindividual variability. *Funct. Ecol.* **21**, 178-184.

Adolph, S. C. and Pickering, T. (2008). Estimating maximum performance: effects of intraindividual variation. *J. Exp. Biol.* **211**, 1336-1343.

Aitchison, J. (2003). A concise guide to compositional data analysis. http://www.leg.ufpr.br/lib/exe/fetch.php/pessoais:abtmartins:a_concise_guide_to_compositional_data_analysis.pdf

Eynon, N., Hanson, E. D., Lucia, A., Houweling, P. J., Garton, F., North, K. N. and Bishop, D. J. (2013). Genes for elite power and sprint performance: ACTN3 leads the way. *Sports Med.* **43**, 803-817.

Flueck, M. (2009). Tuning of mitochondrial pathways by muscle work: from triggers to sensors and expression signatures. *Appl. Physiol. Nutr. Metab.* **34**, 447-453.

Garland, T.Jr. (1994). Quantitative genetics of locomotor behavior and physiology in a garter snake. In *Quantitative Genetic Studies of Behavioral Evolution* (ed. C. R. B. Boake), pp. 251-277. Chicago: University of Chicago Press.

Ghalambor, C. K., Walker, J. A. and Reznick, D. N. (2003). Multi-trait selection, adaptation, and constraints on the evolution of burst swimming erformance. *Am. Zool.* **43**, 431-438.

Head, A. W., Hardin, J. S. and Adolph, S. C. (2011). Methods for estimating peak physiological performance and correlating performance measures. *Environ. Ecol. Stat.* **19**, 127-137.

Holzman, R., Collar, D. C., Mehta, R. S. and Wainwright, P. C. (2011). Functional complexity can mitigate performance trade-offs. *Am. Nat.* **177**, E69-E83.

Lailvaux, S. P. and Husak, J. F. (2014). The life history of whole-organism performance. *Q. Rev. Biol.* **89**, 285-318.

Lee, S. S. M. and Piazza, S. J. (2009). Built for speed: musculoskeletal structure and sprinting ability. *J. Exp. Biol.* **212**, 3700-3707.

MacArthur, D. G. and North, K. N. (2005). Genes and human elite athletic performance. *Hum. Genet.* **116**, 331-339.

Marras, S., Killen, S. S., Domenici, P., Claireaux, G. and McKenzie, D. J. (2013). Relationships among traits of aerobic and anaerobic swimming performance in individual European sea bass *Dicentrarchus labrax*. *PLoS ONE* **8**, e72815.

Martens, B. K. and Collier, S. R. (2011). Developing fluent, efficient, and automatic repertoires of athletic performance. In *Behavioral Sport Psychology* (ed. J. K. Luiselli and D. D. Reed), pp. 159-176. New York: Springer.

Mitteroecker, P. and Bookstein, F. L. (2009). Examining modularity via partial correlations: a rejoinder to a comment by Paul Magwene. *Syst. Biol.* **58**, 346-348.

Pearl, J. (2009). *Causality*. Cambridge: Cambridge University Press.

Phillips, B. E., Williams, J. P., Gustafsson, T., Bouchard, C., Rankinen, T., Knudsen, S., Smith, K., Timmons, J. A. and Atherton, P. J. (2013). Molecular networks of human muscle adaptation to exercise and age. *PLoS Genet.* **9**, e1003389.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reidy, S. P., Kerr, S. R. and Nelson, J. A. (2000). Aerobic and anaerobic swimming performance of individual Atlantic cod. *J. Exp. Biol.* **203**, 347-357.

Ruiz, J. R., Arteta, D., Buxens, A., Artieda, M., Gomez-Gallego, F., Santiago, C., Yvert, T., Moran, M. and Lucia, A. (2010). Can we identify a power-oriented polygenic profile? *J. Appl. Physiol.* **108**, 561-566.

Scholz, M. N., Bobbert, M. F., van Soest, A. J., Clark, J. R. and van Heerden, J. (2008). Running biomechanics: shorter heels, better economy. *J. Exp. Biol.* **211**, 3266-3271.

Servedio, M. R., Brandvain, Y., Dhole, S., Fitzpatrick, C. L., Goldberg, E. E., Stern, C. A., Van Cleve, J. and Yeh, D. J. (2014). Not just a theory—the utility of mathematical models in evolutionary biology. *PLoS Biol.* **12**, e1002017.

Shalizi, C. R. (2013). Advanced data analysis from an elementary point of view. http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/.

Shipley, B. (2002). *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference*. Cambridge: Cambridge University Press.

Sokal, R. R. and Rohlf, F. J. (2012). *Biometry*. New York: W H Freeman & Company.

Timmons, J. A. (2011). Variability in training-induced skeletal muscle adaptation. *J. Appl. Physiol.* **110**, 846-853.

Tucker, R. (2009). The anticipatory regulation of performance: the physiological basis for pacing strategies and the development of a perception-based model for exercise performance. *Br. J. Sports Med.* **43**, 392-400.

Van Damme, R., Wilson, R. S., Vanhooydonck, B. and Aerts, P. (2002). Performance constraints in decathletes. *Nature* **415**, 755-756.

Vanhooydonck, B., James, R. S., Tallis, J., Aerts, P., Tadic, Z., Tolley, K. A., Measey, G. J. and Herrel, A. (2014). Is the whole more than the sum of its parts? Evolutionary trade-offs between burst and sustained locomotion in lacertid lizards. *Proc. R. Soc. Lond. B Biol. Sci.* **281**, 20132677.

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* **68**, 2128-2136.

Walker, J. A. (2015a). A critique of statistical methods used to adjust for Individual Quality and un-mask underlying trade-offs. https://github.com/middleprofessor/NCAA_decathlon/blob/master/Critique_of_van_Damme_and_Wilson.pdf

Walker, J. A. (2015b). R scripts used in Performance trade-offs and Individual Quality in NCAA decathletes". https://github.com/middleprofessor/NCAA_decathlon/blob/master/NCAA_decathlon.pdf

Wang, Y.-X., Zhang, C.-L., Yu, R. T., Cho, H. K., Nelson, M. C., Bayuga-Ocampo, C. R., Ham, J., Kang, H. and Evans, R. M. (2004). Regulation of muscle fiber type and running endurance by PPARδ. *PLoS Biol.* **2**, e294.

Wilson, D. R., Bitton, P.-P., Podos, J. and Mennill, D. J. (2014a). Uneven sampling and the analysis of vocal performance constraints. *Am. Nat.* **183**, 214-228.

Wilson, R. S., Niehaus, A. C., David, G., Hunter, A. and Smith, M. (2014b). Does individual quality mask the detection of performance trade-offs? A test using analyses of human physical performance. *J. Exp. Biol.* **217**, 545-551.

Wright, S. (1918). On the nature of size factors. *Genetics* **3**, 367-374.

Wright, S. (1932). General, group and special size factors. *Genetics* **17**, 603-619.

Wright, S. (1934). The method of path coefficients. *Ann. Math. Statist.* **5**, 161-215.
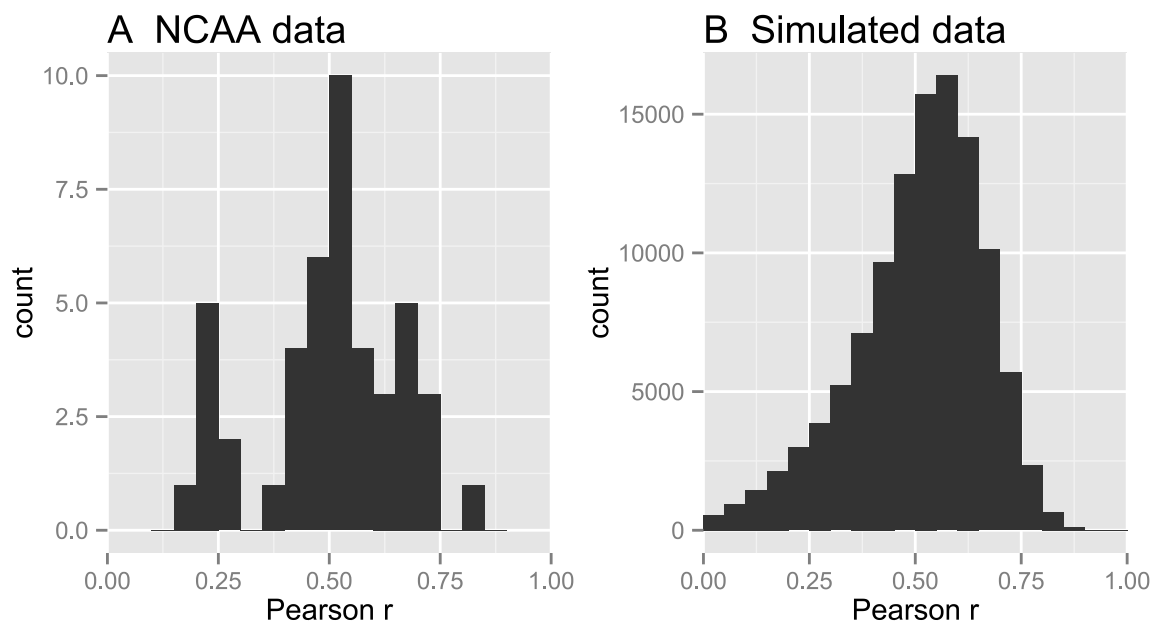
Fig. S1: The distribution of the raw, Pearson product-moment correlations of the A) NCAA data and B) simulated data.
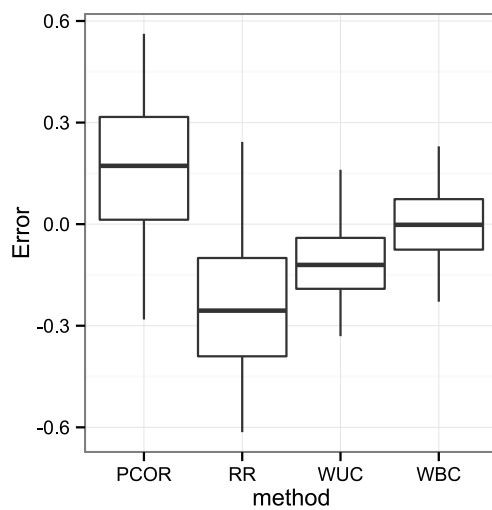
Fig. S2: The distribution of the errors for the four adjustment methods estimated by Monte Carlo simulation of the soccer data. Box plots with the median, 50% and 95% intervals represented by the horizontal line, the box, and the tips of the whiskers. Correlations are partial (PCOR), regression residual (RR), Wright's uncorrected (WUC) and bias corrected (WBC).

Table S1. Pearson and quality-adjusted correlations for soccer data. Pear- son, partial (PCOR), regression residual (RR), and Wright's bias corrected (WBC) correlations are given in Supplement Table 1. The error intervals associated with PCOR, RR, and WBC method are illustrated in Supplement Fig. 2 and were computed using Monte Carlo simulation of 1000 data sets with a distribution of correlations similar to that of the soccer data (see Methods) The Pearson correlations are from Table 1 of Wilson et al. (2014). The regression residual correlations are from Table 2 of Wilson et al. (2014). The WBC correlations were computed using the PC1 loadings in Table 3 of Wilson et al. (2014) (these loadings are estimates of the Quality path coefficients α).

	Event_A	Event_B	Pearson	PCOR	RR	WBC
1	Squat	X1500	0.16	-0.02	-0.52	-0.10
2	Jump	X1500	0.18	-0.08	-0.41	-0.11
3	Jump	Squat	0.47	0.18	-0.20	0.03
4	Sprint	X1500	0.32	0.28	0.05	0.11
5	Sprint	Squat	0.29	0.14	-0.06	-0.03
6	Sprint	Jump	0.37	0.27	-0.25	0.01
7	Agility	X1500	0.28	0.22	-0.44	-0.01
8	Agility	Squat	0.50	0.30	0.15	0.06
9	Agility	Jump	0.63	0.51	-0.02	0.15
10	Agility	Sprint	0.24	-0.09	-0.52	-0.12