

Performance trade-offs in wild white-footed mice

(Peromyscus leucopus)

Ilias Berberi

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
M.Sc. degree in the
Ottawa-Carleton Institute of Biology

Department of Biology
Faculty of Science
University of Ottawa

© Ilias Berberi, Ottawa, Canada 2018

Abstract

Various aspects of performance (e.g., sprint speed, grip strength) are thought to be important determinants of the success of animals in natural activities such as foraging, mating, and escaping from predators. However, it is generally known that morphological properties enhancing one type of performance (e.g., speed) can lead to a reduction in another (e.g., strength). Such performance trade-offs have been quantified at the inter-specific level, but evidence at the inter-individual level remains equivocal. To test for the presence of a performance trade-off, I initiated a study on wild white-footed mice (*Peromyscus leucopus*). In summer 2016, I captured a total of 186 individuals, 87 of which were repeatedly phenotyped for grip strength and sprint speed. A significant positive relationship was found between body mass and grip strength (but not for sprint speed). Individual differences in performance were repeatable through time for both grip strength and sprint speed. Using a bivariate mixed model, I detected a significant negative correlation between grip strength and sprint speed at the *among*-individual level. By contrast, the *within*-individual correlation between grip strength and sprint speed tended to be positive, suggesting that some unquantified aspects of the mouse phenotype (e.g., body condition, age) may have a positive effect on both performance traits. Given the relatively low repeatability of grip strength and sprint speed, a failure to properly partition the correlation at the among- and within-individual level generates a counter-intuitive, positive correlation. This study is one the first to detect a performance trade-off at the among-individual level in a wild animal population.

Résumé

Plusieurs aspects de la performance (p. ex. force de préhension, vitesse de course) peuvent contribuer au succès des animaux lorsqu'ils recherchent de la nourriture, s'accouplent et se sauvent des prédateurs. Cependant, il est généralement connu que les propriétés morphologiques améliorant un type de performance (p. ex. la force) peuvent conduire à une réduction d'un autre type de performance (p. ex. la vitesse). Ces compromis de performance ont été quantifiés au niveau inter-spécifique, mais l'évidence au niveau inter-individuel demeure équivoque. Pour tester la présence d'un compromis de performance, j'ai initié une étude sur des souris à pattes blanches (*Peromyscus leucopus*). Durant l'été 2016, j'ai capturé un total de 186 individus, dont 87 ont été phénotypés à plusieurs reprises pour des mesures de force de préhension et de vitesse de course. Une relation positive significative a été trouvée entre la masse corporelle et la force de préhension (mais pas pour la vitesse de course). Les différences individuelles en force de préhension et vitesse de course étaient consistantes dans le temps, tel qu'indiqué par des valeurs de répétabilité significatives. À l'aide d'un modèle mixte bivarié, j'ai détecté une corrélation négative significative entre la force de préhension et la vitesse de course au niveau *inter-individuel*. Au contraire, la corrélation *intra-individuelle* entre la force de préhension et la vitesse de course était positive, suggérant que certains aspects non quantifiés (p. ex. la condition corporelle, l'âge) peuvent avoir un effet positif sur les deux traits de performance. Compte tenu de la répétabilité relativement faible de la force de préhension et de la vitesse de course, il est important de partitionner correctement la corrélation au niveau inter- vs. intra-individuelle, autrement une corrélation positive contre-intuitive est obtenue. Cette étude est la première à détecter un compromis de performance au niveau inter-individuel chez une population d'animaux sauvages.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Vincent Careau for his incredible mentorship throughout my research. From the ‘unique’ field experiences we’ve shared (thanks again for the socks) to the countless hours of discussions we’ve had, your unwavering support and guidance during these past two years has helped me grow as both a scientist and a person. Working alongside you, I have seen the passion of someone that truly loves their work and it has continuously motivated me to give my all in everything I do. Along with Vincent, I am thankful for the time, assistance, and suggestions provided by my committee members Dr. Gabriel Blouin-Demers, Dr. Sue Bertram, and Dr. Julie Morand-Ferron. You have given me insight and constructive criticisms on my research that has been greatly valued. I extend my thanks to Ian Myers from the Electronics Shop for creating the mouse race track and teaching me its intricacies.

A tremendous thank you goes to the researchers and staff at the Queen’s University Biological Station. You have given me some of the most memorable moments of my life. From the work done in the field to the help received in the office, words cannot explain how thankful I am for my fantastic lab mates. Mathieu Videlier, Jasmine Veitch, Sasha Newar, Amy Villerreal, Alyssa Fiedler, and Sharla Foster, we have formed some magnificent friendships that will never wither. I also greatly appreciate the members of the BGSA for making Ottawa feel like a home away from home. It has been an honour being your president!

Finally, I thank my family for being my support since the beginning. Ana, you have given me such important encouragement throughout my academic career – Brosis for life. Babi and mami, thank you for believing in me and providing everything I have needed to succeed. Faleminderit nena për ushqimin e shijshëm që keni bërë dhe për mesazhet e bukura që më keni dhënë. And to my lovely Maria; you’ve made my dark days bright and my bright days brighter. Thank you.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iv
1. Introduction.....	1
1.1 Importance of trade-offs in biology.....	1
1.2 Performance trade-offs.....	2
1.3 Difficulties in quantifying performance trade-offs.....	3
1.3.1 ‘Personal best’ performance values.....	4
1.3.2 Use of phenotypic correlations.....	5
1.4 Objectives.....	6
2. Materials and Methods.....	7
2.1 Study Species.....	7
2.2 Study Design.....	7
2.2.1 Grip strength.....	8
2.2.2 Sprint speed.....	9
2.3 Statistical Analysis.....	11
2.3.1 Univariate models considering all trials.....	11
2.3.2 Univariate models considering short- vs. long-term repeatability.....	12
2.3.3 Bivariate mixed model considering only maximum performance.....	13
2.3.4 Phenotypic correlations using individual maxima.....	14
3. Results.....	15
3.1 Descriptive results.....	15
3.2 Univariate mixed models: All trials.....	15

3.3 Univariate mixed models: Short- vs. long-term repeatability.....	16
3.4 Bivariate mixed model: Performance trade-off	16
3.5 Phenotypic correlation using individual maxima	17
4. Discussion	18
4.1 Overview	18
4.2 Covariates of grip strength and sprint speed.....	18
4.3 Short-term vs. long-term repeatability.....	20
4.3 Trade-off detection by covariance partitioning	22
4.4 Conclusion	25
References	27
Tables and Figures	35
Appendix.....	42

1. Introduction

1.1 Importance of trade-offs in biology

Trade-offs – when one trait cannot increase without a decrease in another – occur in multiple aspects of biology. A simple example in birds is the behavioural trade-off between the time and energy invested in paternal care (within-pair siring success) vs. extra-pair copulation (Magrath & Elgar, 1997; Petrie & Kempenaers, 1998; Sardell, Arcese, Keller, & Reid, 2011). In addition, birds face a trade-off between total number of eggs produced vs. egg size, given a fixed amount of resources. Another classic example is the life-history trade-off between the energy invested in current vs. future reproduction (van Noordwijk & de Jong, 1986). Since organisms have access to a finite amount of resources throughout their life, resource limitation is a common aspect to many trade-offs (Stearns, 1989, 1992). Resources acquired by an organism must be allocated across different traits to maximize fitness, yet traits that share an opposing mechanism (i.e., the parts and processes responsible for a given outcome) will conflict during allocation.

Consequently, increased allocation in one trait (e.g., current reproduction) will inevitably result in a decrease in another (e.g., survival and future reproduction). Therefore, the presence of a trade-off suggests the traits involved share an opposing underlying mechanisms and relationship to fitness, and as such the quantification of trade-offs leads to a better understanding of the nature of the multivariate relationships among traits.

Trade-off literature has highlighted two contrasting scenarios that model the presence of trade-offs at either the within- or among-individual level. A trade-off occurring within individuals can be modeled using the ‘big houses, big cars’ scenario, where two traits are negatively correlated within individuals, yet are positively correlated among individuals (Reznick, Nunney, & Tessier, 2000). For example, we experience a within-individual trade-off when we purchase a new house and car where our resources (in this case, income) can be

allocated into either a big house, or an expensive car. However, people with more resources will be able to spend more on both their house and car relative to those with fewer resources, thus generating a positive among-individual correlation. On the other hand, a trade-off occurring among individuals can be modeled with the ‘sink or swim’ scenario, where two traits are negatively correlated among individuals, yet are positively correlated within individuals (Careau & Wilson, 2017a). For example, krakens are mythical, ocean-dwelling creatures that experience an among-individual trade-off between their ship-sinking ability and swimming performance due to body shape, where broad-bodied individuals successfully sink more ships yet are slower swimmers in comparison to slimmer individuals. However, a kraken’s indeterminate body growth with age positively influences both their sinking and swimming traits, thus generating a positive within-individual correlation. The ‘big houses, big cars’ and ‘sink or swim’ scenarios can be used to further understand underlying relationships among traits in the presence of a trade-off.

1.2 Performance trade-offs

Trade-offs are particularly important with respect to whole-organism performance, defined as the ability of an individual to conduct an ecologically relevant task when maximally motivated (Careau & Garland Jr., 2012). Performance traits such as bite force, sprint speed, endurance, and grip strength can be key determinants of an individual’s ability to escape predation, outcompete conspecifics, and acquire mating partners (Husak, Fox, Lovern, & Van Den Bussche, 2006; Lailvaux & Husak, 2014; Lappin & Husak, 2005; Le Galliard & Ferrière, 2008; Márquez & Cejudo, 2000; O’Steen, Eby, & Bunce, 2010; Sneddon et al., 2000). To the extent that performance can directly influence Darwinian fitness, simultaneously enhancing multiple aspects of performance would yield a selective advantage. However, it may not always be possible to

increase speed, strength, and endurance due to physiological constraints arising from the well-known biochemical and morphological components of performance (Bennett, Huey, & John-Alder, 1984; Feder, Bennett, & Huey, 2000; Garland Jr. & Carter, 1994). For example, fast oxidative muscle fibres promote quick burst performance (i.e., sprint speed), but are hindrances in comparison to slow-oxidative fibres in stamina-based performance (i.e., endurance running) (Karlsson, Sjödin, Jacobs, & Kaiser, 1981). Variation in morphology can also generate performance trade-offs, as for example the elongated limbs and flexible spine allowing a cheetah to run at an incredibly fast speed also make the cheetah weaker than a large-limbed, broad-shouldered African lion (Eaton, 1974; Schaller, 1972). Similarly, a large lower vs. upper scapular fossa are respectively associated with greater climbing vs. running performance in laboratory mice (Green, Richmond, & Miran, 2012).

1.3 Difficulties in quantifying performance trade-offs

Performance trade-offs have been regularly studied at both the inter- and intra-specific levels. Although performance trade-offs have been clearly detected at the inter-specific level (Stephens & Wiens, 2008; Vanhooydonck, Van Damme, & Aerts, 2001), empirical evidence at the individual level remains equivocal (Friedman, Garland Jr., & Dohm, 1992; Garland Jr., Else, Hulbert, & Tap, 1987; Herrel & Bonneaud, 2012; Oufiero, Walsh, Reznick, & Garland Jr., 2011). In fact, positive correlations are sometimes detected among performance traits that are biologically expected to be involved in a trade-off (Huey, Dunham, Overall, & Newman, 1990; Jayne & Bennett, 1990). One of the reasons put forward for the non-significant or counter-intuitive correlations is that individuals vary in their “quality”, including differences in health, nutrition, and genetics (Van Damme, Wilson, Vanhooydonck, & Aerts, 2002; Wilson, Niehaus, David, Hunter, & Smith, 2014). If “good” individuals can perform better than “bad” individuals

in all tasks, then among-individual variation in quality should mask any performance trade-off occurring within individuals (Van Damme et al., 2002; Wilson et al., 2014). Quality differences may be intrinsic (some individuals may be morphologically or physiologically better suited to perform than others) or extrinsic (some environments or diets may enhance overall performance). Ideally, scientists should test for performance trade-offs while only controlling for the extrinsic part of quality, yet separating the intrinsic and extrinsic components of quality is challenging (Walker & Caddigan, 2015).

1.3.1 'Personal best' performance values

Recently it has been suggested that performance trade-offs can be better studied by fully recognizing the hierarchical structure of performance data and, accordingly, a proper partitioning of the (co)variance in performance traits (Careau & Wilson, 2017a, 2017b). Performance data is typically organized in a hierarchical structure, as measurements usually involve repeated tests of the same individuals on subsequent days and repeated trials within each test. Researchers interested in performance will be naturally inclined to select only the maximum performance score across all tests for every individual (i.e., 'personal best' values). Using maximum values, however, may introduce a bias in which sampled values are not accurately representative of the true performance of an individual (Careau & Wilson, 2017b; Head, Hardin, & Adolph, 2012). For example, if individuals randomly vary in their motivation to perform at their maximum across successive tests, then more heavily sampled individuals are more likely to "show" their true maximum performance compared to individuals with fewer repeated tests. In this case, selecting the 'personal best' values will generate a downward bias in the performance scores of weakly sampled individuals, a bias that may be so severe to generate positive correlations among performance traits (Careau & Wilson, 2017a, 2017b). Furthermore, performance may be

influenced by environmental variables that are specific to a given test, and more heavily sampled individuals are more likely to have performed at least once under favourable conditions.

Importantly, considering all repeated tests and trials within a test allows the identification of potentially important covariates such as time of day, test sequence, and age.

1.3.2 Use of phenotypic correlations

A simple solution to reduce the bias associated with ‘personal best’ scores would be to take the average or a percentile of the distribution of the performance measures on each individual (Head et al., 2012). However, aggregating performance scores at the individual level has the effect of blending two levels of variation into one. Phenotypic variation observed in any complex trait can be partitioned – in its simplest form – into variation occurring among vs. within individuals.

Consequently, the phenotypic covariance between two performance traits results from the among- and within-individual covariances. Indeed, the “masking effect” of quality (see above) is thought to occur because quality varies *among* individuals while performance trade-offs occur *within* individuals, a scenario akin to the ‘big houses, big cars’ model (Careau & Wilson, 2017a; Reznick et al., 2000). However, there is no a priori reason to expect such a scenario vs. an alternative ‘sink or swim’ scenario where trade-offs occur at the among-individual level while quality generates positive correlations at the within-individual level (Careau & Wilson, 2017a). Therefore, it is crucial to partition trait covariance at the among- and within-individual levels, which requires repeated pairs of measurements on many individuals (Dingemanse & Dochtermann, 2013).

1.4 Objectives

There were four objectives to this study. First, I aimed to quantify grip strength and sprint speed in wild white-footed mice (*Peromyscus leucopus*). Second, I tested for the effects of multiple predictors, including body mass, test sequence, and trial sequence, in both performance traits. Third, I evaluated repeatability on a short-term and long-term basis to quantify the proportion of variation occurring *i-*) among individuals, *ii-*) among repeated tests within individuals, and *iii-*) among repeated trials within tests. Finally, I tested for the presence of trait correlation between grip strength and sprint speed to determine whether a performance trade-off occurs at the among- and/or within-individual levels, and if so whether the pattern of covariation follows the ‘big houses, big cars’ or a ‘sink or swim’ scenario.

2. Materials and Methods

2.1 Study Species

White-footed mice are small, nocturnal mammals with home ranges averaging 590 m² (Graves, Maldonado, & Wolff, 1988; Metzgar, 1973; Wolff, 1985). *P. leucopus* must perform multiple locomotor tasks in a variety of contexts to maximize their fitness. *P. leucopus* readily use trees as options for escape routes and diurnal refuge (Harney & Dueser, 1987; Wolff & Hurlbutt, 1982). Their polygynandrous mating system (Xia & Millar, 1991) suggests that compared to slow and weak individuals, fast and strong individuals could increase their opportunities of mating with more conspecifics. Females with greater locomotor performance relative to others could gain access to and/or outcompete conspecifics for higher-quality nesting sites to raise their offspring. In addition to mating success, locomotor performance traits allow *P. leucopus* to avoid and escape from common aerial and terrestrial predators (Barnum, Manville, Tester, & Carmen, 1992; Metzgar, 1967). Evidence in laboratory mice suggests that greater climbing and running performance traits are respectively associated with a large lower or upper scapular fossa (Green et al., 2012). Hence, a morphological constraint between climbing and running abilities in *P. leucopus* may generate a performance trade-off, where good climbers with a strong grip will be bad runners with a slow sprint speed, and *vice versa*.

2.2 Study Design

The study site was located on Cow Island at the Queen's University Biological Station (44°34'08"N; 76°19'08"W), a 7 ha island mostly covered with deciduous forest dominated by sugar maple (*Acer saccharum*) and oaks (*Quercus rubra* and *Q. alba*). Historically, Cow Island has been used to monitor natural *P. leucopus* populations (Bendell, 1959). The island was sampled using 214 Longworth traps and 200 nest boxes, overlaid onto the island in a 15 m × 15

m sampling grid alternating between 2 Longworth traps or 2 nest boxes per sampling location. Trapping occurred daily from May 18 to August 17 of 2016. Traps were set at sunset, insulated with polyester batting and baited with sunflower seeds, oats, and a piece of apple. Traps were checked at sunrise the following day. Newly captured individuals were assigned a unique identification code and permanently marked with two uniquely-numbered ear tags. For each capture, body mass, sex, age, and reproductive status were recorded. Age was categorized as either adult or juvenile based on colour patterns in pelage. Adult mice had complete brown pelage above the middorsal molt line, while juvenile mice had brown and grey pelage above the molt line (Collins, 1923). Reproductive status was recorded as a binary variable of reproductively active or not based on external sexual characteristics.

After standard manipulations were completed at the capture site, mice were placed back into their trap and transported to a laboratory located 600 m away from the study site, where performance was measured (see below). Up to 16 individuals per day were transported to and measured in the laboratory. Additional mice were released immediately, as well as mice that had been to the laboratory less than 7 days prior. Body mass measurements were taken using a precision scale (Mettler Toledo, Model ML1602T/00). All mice were released at the end of the day directly at their capture location.

2.2.1 Grip strength

A grip test was used to measure neuromuscular function as maximal muscle strength of combined forelimbs and hindlimbs. Each test was comprised of 4 repeated trials, where grip strength was assessed by the grasping force applied from the mouse on a 17×12 cm metal grid with 0.16cm wide metal wires spaced 1.27×1.27 cm apart. The metal grid was connected to a force gauge (Mecmesin basic force gauge model BFG 10, range $10\text{N} \pm 0.25\%$) and fixed to a

stand with adhesive rubber pads. The metal grid was attached to the gauge at a slight declined angle to limit mass-bias of grip strength measurements due to gravitation pull (i.e., large mice will have a greater gravitational force pushing them towards a perfectly horizontal grid in comparison to a lighter mouse). The mouse was lowered towards the metal grid while being held by the tail, until the mouse instinctively grabbed onto the metal grid. The mouse was then gently and consistently pulled backwards by the tail along the horizontal plane. The peak force (N) applied to the grid before the mouse lost grip was recorded as grip strength for that trial.

Although we did not have a way to motivate an individual to grip hard, being pulled by the tail would have detrimental effects in a natural setting (e.g., captured by a predator), so I assumed the animals were always maximally motivated. Following grip strength measurements, the mouse was returned into its trap with access to food for a resting period of at least 30 minutes.

2.2.2 Sprint speed

Sprint speed was measured on a 3 m long by 10 cm wide race track, with 30 cm high black acrylic walls and shelter units (one on either end). Along the bottom of the track, 24 stainless steel mesh sensors (5 cm long by 9 cm wide) were placed sequentially at every 12.5 cm. The sensors were connected by two MPR121 12-channel capacitive touch sensor detectors, which were sensitive enough to detect the touch of a mouse. Touch detection signals were sent to a custom recording software via an Arduino microprocessor. The processor ran a continuous millisecond timer, which reported a time value (ms) since the first sensor touched during a given trial (i.e., the first sensor touched was assigned a value of 1 ms, and the software recorded the time at which all other sensors were subsequently touched). The software did not record a time value for a sensor if an individual jumped over it. Those missing values were later interpolated (see below).

Each sprint speed test consisted of 10 repeated trials, alternating between high and low motivation levels. The mouse first was released into the starting shelter unit and maintained there for < 30 sec until a door was opened and the mouse was motivated to run down the track at their maximum speed (high motivation level). This was done by chasing the mouse with a meter stick covered with cardboard and duct tape (6.5×10 cm, curved) and using various methods of making noise, including hand clapping, yelling, and “evil laughing”, to prompt the mouse to run into a second shelter unit on the opposite side of the race track. The following trial was conducted under low motivation, where the mouse was released from the second shelter unit and allowed to return to the starting shelter unit, during which the observer stood behind the mouse and then quickly closed the door behind the mouse as it returned into the starting shelter unit. This was repeated four more times for a total of ten trials per test.

I first examined the raw data recorded by the race track to count number of missing time values (i.e., cases where a mouse would jump over a sensor or a contact failed to be detected by a sensor) and eliminate any aberrations (i.e., cases where a downstream sensor was falsely triggered before one or more upstream sensors). Then, I conducted an analysis to determine the optimal distance over which sprint speed should be calculated. Calculating maximum sprint speed across the shortest distance possible (i.e., 12.5 cm) yields the fastest sprint speed values (Fig. S1 in the Appendix). However, calculating speed over such a short distance introduces measurement error related to the random positioning of the mouse limbs as it runs down the track (i.e., in cases where a mouse lands on a pad with its hind limbs and then touches the following pad with its front limbs, the calculated speed would be higher than the speed at which the mouse was running). By contrast, calculating speed across the entire race track yields the slowest sprint speed values (Fig. S1). Although measurement error is greatly reduced, measuring speed over the

entire race track introduces variation in motivation (i.e., mice rarely run at a consistent speed across the entire race track). Using an inflection point analysis, I determined that calculating speed over 12 sensors (137.5 cm) represented the optimal compromise between minimising measurement error vs. introducing variation in motivation within a trial (Fig. S1). For analysis purposes, I interpolated values for undetected times across consecutive detected times within a trial. Accordingly, I did not interpolate times across distances greater than the calculated optimal distance. Trials with more than 20 missing time values were discarded, as they did not acceptably quantify sprint speed.

2.3 Statistical Analysis

2.3.1 Univariate models considering all trials

Grip strength and sprint speed were z-transformed (mean = 0, variance = 1) and first analysed using linear mixed models (LMM) in `ASReml-R` (Butler, Cullis, Gilmour, & Gogel, 2009). Two separate univariate LMMs were used to test how grip strength and sprint speed were influenced by body mass, sex, age, reproductive status, capture type, test sequence, trial sequence, handler (i.e., identity of the person conducting the test), time of day, and a two-way interaction between sex and reproductive status. The univariate analysis for sprint speed included additional fixed effects of motivation level, an interaction between trial sequence and motivation level, and the number of interpolated time values per trial. Categorical variables (sex, age, reproductive active status) were centered to allow interpretation of main effects in the presence of interactions in the model (Schielzeth, 2010). I calculated the upper and lower 95% confidence intervals (CI) by respectively adding and subtracting $1.96 \times$ standard error from the estimate. To independently visualize the effects of body mass, trial sequence, and test sequence while conditioning on all other variables in the model, I re-fitted the two univariate LMMs using `lmer()` function in R

package `lme4` (Bates, Mächler, Bolker, & Walker, 2014) and used the `visreg()` function in R package `visreg` (Breheny & Burchett, 2017) to calculate partial residuals.

2.3.2 Univariate models considering short- vs. long-term repeatability

Using data from all trials, univariate LMMs allow straightforward partitioning of the phenotypic variation (after conditioning on the fixed effects listed above) into three distinct components: among individuals, among tests within individuals, and among trials within tests (residuals). All models included individual identity as a random effect to quantify among-individual variance (V_{ind}). I included a ‘series’ random effect that estimates the variance among tests within a given individual (V_{test}). This extra random effect is a unique combination of individual identity and day (thus grouping all trials for a given day on a given individual). The residual variance (V_e) represents the variation among trials within a given test. Such partitioning of the variance allows estimating long-term and short-term repeatability (R) as follows:

$$R_{longterm} = \frac{V_{ind}}{V_{ind} + V_{test} + V_e}$$
$$R_{shortterm} = \frac{V_{ind} + V_{test}}{V_{ind} + V_{test} + V_e}$$

It is important to calculate short- vs. long-term repeatability to get a sense of the relative amount of variance in performance measurements taken at different temporal scales (Araya-Ajoy, Mathot, & Dingemanse, 2015). Long-term repeatability ($R_{longterm}$) represents the proportion of the total phenotypic variation attributed to genetic differences and permanent environmental effects (i.e., environmental variation causing among-individual differences over the time span within which the repeated measures were taken). By contrast, short-term repeatability ($R_{shortterm}$) adds V_{test} in its numerator, and thus necessarily equals or exceeds $R_{longterm}$. The V_{test} component captures short-term consistency caused by unique environmental conditions experienced by each

individual on a given test day. The remaining variance ($1 - R_{shortterm}$) represents the proportion of the variance among successive trials within a test. Hence, comparing $R_{shortterm}$ vs. $R_{longterm}$ yields an appreciation of the importance of variance across all 3 hierarchical levels in our data. For example, a very high $R_{shortterm}$ but a low $R_{longterm}$ would suggest taking repeated trials per test is not necessary, but taking repeated tests per individual is required to properly estimate among-individual differences (Falconer & Mackay, 1996).

2.3.3 Bivariate mixed model considering only maximum performance

A second analysis was based on the maximum value of grip strength and sprint speed measured on a given test (day) for an individual. Due to the high variance of performance measurements among trials within tests (V_e ; see below), I extracted the best trial per test as an indicator of the performance of that individual on that day. In other words, the grip strength of an individual for a given test (day) was recorded as the strongest grip force exhibited across the four grip trials, and sprint speed as the fastest sprint speed across the ten trials. Note that this approach is different from extracting ‘personal best’ scores, because I still considered all repeated tests made on a given individual on different days. Moreover, all tests were standardised to the same number of trials (four and ten for grip and speed tests, respectively), thus there is no sampling bias at this level of variation.

The bivariate model included individual identity as random effect to quantify V_{ind} , whereas V_e represented within-individual variation in maximum performance measures. Repeatability estimates for both grip strength and sprint speed performance were calculated as a ratio of V_{ind} to the total measurement variance ($V_{ind} + V_e$). I analysed the correlation between grip strength and sprint speed using a bivariate mixed model in ASRem1-R. With repeated pairs of measurements, a bivariate mixed model allowed partitioning the phenotypic correlation between

two traits at the among-individual (r_{ind}) and within-individual (r_e) levels with the use of two unstructured (co)variance matrices: one for the random effect of V_{ind} and another for V_e (Dingemanse & Dochtermann, 2013). I calculated the upper and lower 95% CI for r_{ind} and r_e using profile likelihood tests with the `nadiv` package (Wolak, 2012).

2.3.4 Phenotypic correlations using individual maxima

To compare my results with those obtained in many past studies, I extracted the maximum performance score obtained for each individual across all their repeated tests for both grip strength and sprint speed. Then I used these ‘personal best’ scores to compute the phenotypic correlation (r_P) between grip strength and sprint speed. Since I only had one ‘personal best’ value per individual, I simply fit a bivariate model without any fixed or random effect and allowed a single unstructured (co)variance matrix at the residual level. In a second bivariate model, I included as a fixed effect the number of repeated measures taken throughout the study on each individual. The change in r_P should reflect the extent of bias related to the use on individual maxima (Careau & Wilson, 2017a).

3. Results

3.1 Descriptive results

A total of 186 individuals (93 males and 93 females) were captured during the 2016 field season and brought into the laboratory at least once for performance measurements. On average, individuals experienced two laboratory visits, ranging from one to eight visits throughout the season (Fig. 1A). Due to natural mortality and migration, some individuals were only sampled within a restricted time window, while some individuals were repeatedly sampled throughout the season (Fig. 1B). In 52 out of the 444 total paired test phenotyping, an individual was only tested for one of the two performance measurements due to behavioural signs of distress during captivity or technical problems with the race track.

3.2 Univariate mixed models: All trials

Grip strength was quantified in 186 individuals across 431 tests for a total of 1,724 trials. Grip strength increased with body mass but did not differ between males and females (Table 1A, Fig. 2A). Juvenile mice exhibited weaker grip strength than adults (Table 1A). In addition, grip strength generally increased with each consecutive visit to the lab (test sequence; Table 1A, Fig. 2C). Within a test, however, there was no effect of trial sequence on grip strength (Table 1A, Fig. 2E).

Sprint speed was quantified in 168 individuals across 400 tests for a total of 3,515 trials. Sprint speed was not influenced by body mass or sex (Table 1B). Reproductively active individuals exhibited faster sprint speeds than non-reproductive individuals (Table 1B). Unlike grip strength, sprint speed did not increase with body mass or across consecutive lab visits (Table 1B, Fig. 2B, Fig. 2D). However, sprint speed increased with each successive trial within a test (Table 1B). Individuals displayed faster sprint speed during high motivation trials when

compared to low motivation trial speeds (Table 1B). In addition, a significant “motivation × trial” interaction indicated that sprint speed increased across consecutive high motivation trials within a test, yet sprint speeds remain constant across low motivation trials (Table 1B, Fig. 2F).

3.3 Univariate mixed models: Short- vs. long-term repeatability

While accounting for the fixed effects above, both grip strength and sprint speed were significantly repeatable on a short- and long-term basis (Table 1). Short-term repeatability was 0.46 for grip strength and 0.36 for sprint speed. As expected, long-term repeatability was lower, with estimates of 0.19 for grip strength and 0.17 for sprint speed, indicating that 17-19% of the total phenotypic variance is due to consistent individual differences (Table 1). The remaining variance (V_e) represents variation among repeated trials within tests, which is considerable and likely due to a combination of measurement error and variation in motivation across successive trials within a test (Table 1).

3.4 Bivariate mixed model: Performance trade-off

Although the analyses above considered all of the repeated tests and trials, here I only retained the daily best measure per test to better approximate the performance for an individual on a given day. Repeatability estimates for performance were similar to long-term estimates for both grip strength and sprint speed measurements (Table 2). At the among-individual level, a significant negative relationship was detected between grip strength and sprint speed ($r_{ind} = -0.39$; 95% CI: -0.68, -0.08). On average, individuals who have a strong grip are slow sprinters, whereas fast sprinters have a weaker grip (Fig. 3A). At the within-individual level, there was no significant relationship between grip strength and sprint speed ($r_e = 0.06$; 95% CI: -0.03, 0.16) (Fig. 3B). Still, the within-individual trend was positive, suggesting that on a given day, individuals that

performed stronger relative to their average grip strength tended to be faster relative to their average sprint speed (Fig. 3B).

3.5 Phenotypic correlation using individual maxima

Using each individual maximum performance across all repeated tests, I obtained a positive and significant correlation between grip strength and sprint speed ($r_P = 0.34$ [0.20, 0.47]). This r_P estimate, however, might be biased upward because individuals greatly varied in their number of repeated measurements (Fig. 1A). In fact, as expected, individuals with more repeated measures had, on average, higher ‘personal best’ scores for both grip strength and sprint speed (Fig. 4). When conditioning grip strength and sprint speed on the number of repeated measures taken on each individual, the r_P remains positive and significant, but the correlation is noticeably weaker ($r_P = 0.18$ [0.04, 0.34]).

4. Discussion

4.1 Overview

This study aimed to 1) quantify grip strength and sprint speed in a large number of wild *P. leucopus* individuals, 2) evaluate the effect of body mass, test sequence, and trial sequence for each performance trait, 3) partition variation across three separate levels of variation (among individuals, among tests within individuals, and among trials within tests), and 4) test for the presence of a performance trade-off at the among- vs. within-individual levels. Overall, I found that body mass and test sequence had a positive effect on grip strength but not on sprint speed, whereas trial sequence positively affected sprint speed in high motivation trials. After accounting for these and other sources of variance, both traits were significantly repeatable (i.e., individual differences were consistent through time). Finally, grip strength and sprint speed were significantly and negatively correlated at the among-individual level, but tended to be positively correlated at the within-individual level. Altogether, my study perfectly illustrates the importance of partitioning correlations at the among- vs. within-individual levels when testing for the presence of performance trade-offs. Had I decided to test for a trade-off using ‘personal best’ scores, as done in many past studies, I would have obtained a counter-intuitive positive correlation between grip strength and sprint speed.

4.2 Covariates of grip strength and sprint speed

Body mass was positively correlated with grip strength, but not sprint speed. These results are similar to those in the literature, where power performance traits such as grip strength and bite force, but not locomotor performance traits, are influenced by body mass in rodents (Djawdan & Garland Jr., 1988; Dohm, Hayes, & Garland Jr., 1996; Freeman & Lemen, 2008; Garland Jr., 1983). Compared to small individuals, individuals with greater body mass may concurrently

develop a stronger grip for climbing and clinging on various natural substrates (e.g., trees, branches). However, being larger may not be as beneficial for running performance, as a large body mass will impose a greater inertia and energetic demand on the individual to rapidly accelerate (despite potentially stronger limb muscles). Consequently, big mice ran at similar speeds to small mice, at least on my 3 m long race track.

An increase in grip strength was detected across test sequence, suggesting that individuals became stronger across the field season. Although confounded with the effect of age, test sequence effects on grip strength could be explained by either physiological (e.g., training) and/or behavioural (e.g., learning) processes. In contrast to the race track test for sprint speed, the grip strength test exposed individuals to conditions that would likely have detrimental effects in a natural setting (i.e., being pulled from their tail by a predator). Although mice may be habituated to run on a flat surface in their natural settings, being pulled by the tail would likely lead to a lethal outcome (i.e., being eaten). The novelty effect of the tail-pulling condition may provide optimal conditions for learning through repeated testing (Kormi-Nouri, Nilsson, & Ohta, 2005; Tulving & Kroll, 1995). Therefore, an individual may have used previous knowledge and experience to enhance grip strength performance across consequence tests.

Regarding sprint speed quantification, significantly positive effects of trial, motivation, and their interaction suggest my efforts to maximally motivate individuals yielded greater sprint speeds. The use of multiple methods of motivation (see 2.2.2 *Sprint Speed*), in addition to alternating between high and low motivation conditions, was enough to maintain stimulant uncertainty during sprint speed tests. Stimulant uncertainty is necessary to minimize the effects of habituation (Bolivar, 2009; Leussis & Bolivar, 2006), which may result in a decreased response during performance phenotyping. It must be noted that some individuals refusing to run

under high motivation readily ran as they returned to the starting shelter unit (under low motivation level). In fact, the fastest trial occurred under low motivation in 17% of sprint speed tests. Thus, varying motivation levels seems necessary to get a measure of sprint speed in some cases where the individual refuses to run when prompted to do so.

When comparing the univariate and bivariate models, we see minor changes to the effect sizes and significance levels of some predictor variables. One reason for the minor significance difference is the influence of motivation across repeated trials per test. In the univariate models, the variation in motivation occurring across repeated trials is captured from the inclusion of all recorded performance trials. However, the bivariate model contains only the daily best measurement of each performance trait per test, and therefore eliminates the variation at the trial level. For example, the effect of sex was not significant in the univariate model for grip strength, but sex was significant in the bivariate model. This may imply that females experienced less variation in motivation and frequently performed at their maximum grip strength, while males had greater variation of motivation across trials. Comparison of models with all performance values to models with only extracted daily best performance values can be a useful tool for detecting differences in motivation. Still, to determine the significance of a given predictor variable on performance, one should retain and analyse the best of the repeated trials within a given performance test, since it presumably reduces the influence of motivation and yields the best estimate of performance.

4.3 Short-term vs. long-term repeatability

The most straightforward method to compare the amount of among- vs. within-individual variance is to calculate repeatability. Repeatability allows us to determine the proportion of total variance attributed to variance among individuals, and in turn be utilized to determine the within-

individual variance proportion using $1 - R$ (Nakagawa & Schielzeth, 2010). My study, however, includes measurements made at three different levels and therefore variation occurs *i*) among individuals, *ii*) among tests within individuals, and *iii*) among trials within a test. The relative amount of variance in performance at these three temporal scales can be easily grasped by calculating short- and long-term repeatability (Araya-Ajoy et al., 2015). Long-term repeatability indicates that 19% and 17% of total variance in grip strength and sprint speed, respectively, is attributed to among-individual variance. As expected, short-term repeatability estimates were substantially higher (0.46 and 0.36, respectively), suggesting that 27% and 19% of variance for grip strength and sprint speed performance traits can be attributed to ‘within-day’ effects (i.e., temporally correlated factors unquantified in this study). The proportion of variance that remains unexplained, quantified as $1 - R_{shortterm}$, suggests that 54 - 64% of the variance occurs among successive trials with a test, which most likely represents variation in motivation (see above).

Short- and long-term repeatability estimates can be used to determine the importance of conducting repeated tests per individual and repeated trials per test, respectively. One importance of taking repeated measures is the reduction of uncertainty around individual estimates. Measuring a trait with $R = 0.8$ will accurately quantify the individual trait value within a few repeated measurements per individual. However, measuring a trait with $R = 0.2$ will require many repeated measurements to minimize the influence of within individual variation on individual estimates. In my case, short-term repeatability estimates justify the use of many repeated trials for each performance test, and suggest that many repeated tests per individual are also required to properly quantify individual differences in locomotor performance. Therefore, I encourage researchers to disclose both short-term and long-term repeatability estimates of

performance traits for purposes of comparing measurement variance across different temporal scales.

4.3 Trade-off detection by covariance partitioning

Detecting a trade-off between moderately repeatable traits can be complicated by the fact that their phenotypic covariance is shaped by covariances occurring both at the among- and within-individual levels. Therefore, solely focusing on phenotypic correlations is misleading in a scenario where the among- and within-individual correlations differ. Two contrasting cases of such scenarios have been proposed. In the ‘big houses, big cars’ scenario, trade-offs occur at the within-individual level, but can be masked by a positive relationship between traits among individuals (Reznick et al., 2000). Taking the egg quantity vs. size trade-off, individuals that have acquired more resources can lay more eggs and larger eggs than individuals that have acquired less resources, thus masking the allocation trade-off that inevitably occurs within each individual (Stearns, 1992).

Wilson et al (2014) transposed this idea to performance and argued that trade-offs went undetected because they were masked by variation in individual “quality”. However, two main issues arise with the use of quality. First, quality is difficult to define. The definition of an individual’s quality varies on the context of the question being asked (Wilson & Nussey, 2010). As a result, using variation in quality among individuals as a reason for undetected negative covariances between traits may be a false portrayal of some unquantified trait(s) (Wilson & Nussey, 2010). Second, it is difficult to correct for quality in a wild population. Quality is comprised of both intrinsic and extrinsic components, where the intrinsic component represents the underlying relationship between morpho-physiological and performance traits, and the extrinsic component represents external factors influencing an individual’s performance (i.e.,

training, diet) (Walker & Caddigan, 2015). Ideally, correcting for the extrinsic component of quality will unmask any underlying trade-off(s) due to intrinsic mechanical components. Walker and Caddigan (2015) highlighted the difficulty in quantifying intrinsic and extrinsic components of quality when studying performance traits in wild animals and suggested that the physiological interpretation of quality has yet to be determined in a natural population. Therefore, quality must be clearly and correctly quantified among individuals to accurately explain the masking of within-individual trade-offs under the ‘big houses, big cars’ scenario.

By contrast, in the ‘sink-or-swim’ scenario, trade-offs occur at the among-individual level, but can be masked by a positive relationship between traits within individuals (Careau & Wilson, 2017a). Especially regarding performance traits, the ‘sink-or-swim’ scenario is driven by individual variation in phenotypes (e.g., morphological and physiological traits) that enhance certain performance while impairing others (e.g., speed vs. endurance based on fast- vs. slow-twitch muscle fibre types; see Dohm et al., 1996; Garland Jr., 1988). Positive within-individual correlations are driven by unquantified changes in condition (e.g., fat content, hormonal state) or extrinsic quality (e.g., diet, training). For example, athletes performing in multi-event sports (e.g., decathlon) face performance trade-offs among different functional group events (i.e., throwing, jumping, and running), but may perform better-than-usual across all events during a given competition as a result of an improved training regime adopted prior to the competition (Careau & Wilson, 2017b).

My results offer support for the ‘sink-or-swim’ scenario, as I found a trade-off between grip strength and sprint speed at the among-individual level, but the two traits tended to be positively correlated within individuals. Further research is needed to elucidate the underlying cause of the results. At the among-individual level, consistent individual differences in scapular

morphology could explain why mice with strong grip are slow runners, and vice versa (Green et al., 2012). At the within-individual level, a positive trend could be caused by unquantified changes in factors that positively affect both performance traits (e.g., condition, diet, and micro-habitat).

Testing for the ‘big houses, big cars’ or ‘sink-or-swim’ scenarios requires the partitioning of covariance and the among- and within-individual levels through a multivariate mixed model (MMM) approach. Historically, however, researchers testing for performance trade-offs did not use MMMs. Instead, phenotypic correlations were computed using individual maximum (i.e., ‘personal best’) values for each performance traits. This practice alone could be the reason why performance trade-offs are notoriously hard to detect and could even explain some of the counter-intuitive positive correlations obtained. In our case, the phenotypic correlation (r_p) computed using ‘personal best’ values from each individual across all repeated tests was positive and significant, but this r_p estimate is biased due to the unequal number of repeated measurements per individual (Fig. 1A and 4). This is because the likelihood of quantifying an individual’s maximum performance increases with the number of repeated measures taken on that individual (Careau & Wilson, 2017a, 2017b). When conditioning on the number of repeated measures in the phenotypic correlation analysis, the r_p was weakened, yet remained positive and significant. These results are perhaps not surprising to those familiar with the phenomenon of attenuation (Adolph & Hardin, 2007; Head et al., 2012; Spearman, 1904), given the relatively low level of repeatability and the contrasting among- and within- individual correlations obtained in this study. What is surprising is the time it took for researchers to realize the problems with individual maxima and modify their practice. My study emphasizes the importance of minimizing bias related to maximum values and unequal sampling across individuals. Most

importantly, the practical solution of using MMMs and all repeated measures clearly represents a promising avenue for research on performance trade-offs.

4.4 Conclusion

Results from my study present evidence for a performance trade-off at the among-individual level in a wild animal population. Many different research directions can be taken upon quantifying trade-offs. For example, research may continue in an evolutionary perspective, where the selective importance of each performance trait is quantified by its relationship with fitness. One common fitness proxy is survivability via capture-recapture data (Lebreton, Burnham, Clobert, & Anderson, 1992), which accompanies most wild animal datasets (including mine). Furthermore, the ecological importance of closely related performance traits could be assessed by looking at multiple suites of performance traits (Charters et al., 2018). Alternatively, the trade-off between two performance traits can be further understood by studying their biomechanical (i.e., muscle movement and structure) and physiological (i.e., cellular function) properties. Detecting a trade-off at these levels provides information on resource allocation throughout the organism.

Biases that have been historically present in performance data can be eliminated using my proposed statistical approach, and by doing so, trait relationships that often only occurred conceptually can be uncovered. Still, quantifying complex performance traits in wild animals remains a difficult task, as many environmental and biological aspects can influence how an individual performs during a test. For example, an individual's maximum performance capacity may be buffered by their behaviour during trait quantification (Garland Jr. & Carter, 1994). While it may be desirable to eliminate variation introduced by motivation in performance measures, the trial-to-trial variability in running speed measured in the context of an escape

response may provide insight in the evolution of predictability of behaviour (Cleasby, Nakagawa, & Schielzeth, 2015; Martin, Pirotta, Petelle, & Blumstein, 2017; Roche, Careau, & Binning, 2016). Therefore, I recommend that future research concurrently monitor survival, physiological, and behavioural aspects of biology when quantifying performance traits.

References

- Adolph, S. C., & Hardin, J. S. (2007). Estimating phenotypic correlations: Correcting for bias due to intraindividual variability. *Functional Ecology*, *21*(1), 178–184.
- Araya-Ajoy, Y. G., Mathot, K. J., & Dingemanse, N. J. (2015). An approach to estimate short-term, long-term and reaction norm repeatability. *Methods in Ecology and Evolution*, *6*(12), 1462–1473.
- Barnum, S. A., Manville, C. J., Tester, J. R., & Carmen, W. J. (1992). Path selection by *Peromyscus leucopus* in the presence and absence of vegetative cover. *Journal of Mammalogy*, *73*(4), 797–801.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
- Bendell, J. F. (1959). Food as a control of a population of white-footed mice, *Peromyscus leucopus noveboracensis* (Fischer). *Canadian Journal of Zoology*, *37*(2), 173–209.
- Bennett, A. F., Huey, R. B., & John-Alder, H. (1984). Physiological correlates of natural activity and locomotor capacity in two species of lacertid lizards. *Journal of Comparative Physiology B*, *154*(2), 113–118.
- Bolivar, V. J. (2009). Intrasession and intersession habituation in mice: From inbred strain variability to linkage analysis. *Neurobiology of Learning and Memory*, *92*(2), 206–214.
- Breheny, P., & Burchett, W. (2017). Visualization of regression models using visreg. *R Package*, *9*, 1–15.
- Butler, D. G., Cullis, B. R., Gilmour, A. R., & Gogel, B. J. (2009). *ASReml–R Reference Manual*. VSN International Ltd. VSN International Ltd.
- Careau, V., & Garland Jr., T. (2012). Performance, personality, and energetics: Correlation,

- causation, and mechanism. *Physiological and Biochemical Zoology*, 85(6), 543–571.
- Careau, V., & Wilson, R. S. (2017a). Of uberfleas and krakens: Detecting trade-offs using mixed models. *Integrative and Comparative Biology*, 57, 362–371.
- Careau, V., & Wilson, R. S. (2017b). Performance trade-offs and ageing in the “world’s greatest athletes”. *Proceedings of the Royal Society B: Biological Sciences*, 284, 20171048.
- Charters, J. E., Heiniger, J., Clemente, C. J., Cameron, S. F., Amir, A. F., Nasir, A., ... Wilson, R. S. (2018). Multidimensional analyses of physical performance reveal a off between suites of traits. *Functional Ecology*, 32(6), 1541–1553.
- Cleasby, I. R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour: Statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, 6(1), 27–37.
- Collins, H. H. (1923). Studies of the pelage phases and of the nature of color variations in mice of the genus *Peromyscus*. *Journal of Experimental Zoology*, 38(1), 45–107.
- Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, 82(1), 39–54.
- Djawdan, M., & Garland Jr., T. (1988). Maximal running speeds of bipedal and quadrupedal rodents. *American Society of Mammalogists*, 69(4), 765–772.
- Dohm, M. R., Hayes, J. P., & Garland Jr., T. (1996). Quantitative genetics of sprint running speed and swimming endurance in laboratory house mice (*Mus domesticus*). *Evolution*, 50(4), 1688–1701.
- Eaton, R. L. (1974). The cheetah: The biology, ecology, and behaviour of an endangered species. In *Ecology* (pp. 41–87). New York: Van Nostrand Reinhold Company.
- Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics* (4th edn.).

Harlow: Longman.

- Feder, M. E., Bennett, A. F., & Huey, R. B. (2000). Evolutionary physiology. *Annual Review of Ecology and Systematics*, 31, 315–341.
- Freeman, P. W., & Lemen, C. A. (2008). A simple morphological predictor of bite force in rodents. *Journal of Zoology*, 275(4), 418–422.
- Friedman, W. A., Garland Jr., T., & Dohm, M. R. (1992). Individual variation in locomotor behavior and maximal oxygen consumption in mice. *Physiology & Behavior*, 52(1), 97–104.
- Garland Jr., T. (1983). The relation between maximal running speed and body mass in terrestrial mammals. *The Zoological Society of London*, 199, 157–170.
- Garland Jr., T. (1988). Genetic basis of activity metabolism. I. Inheritance of speed, stamina, and antipredator displays in the garter snake *Thamnophis sirtalis*. *Evolution*, 42(2), 335–350.
- Garland Jr., T., & Carter, P. A. (1994). Evolutionary physiology. *Annual Review of Physiology*, 56(1), 579–621.
- Garland Jr., T., Else, P. L., Hulbert, A. J., & Tap, P. (1987). Effects of endurance training and captivity on activity metabolism of lizards. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 252(3), R450–R456.
- Graves, S., Maldonado, J., & Wolff, J. O. (1988). Use of ground and arboreal microhabitats by *Peromyscus leucopus* and *Peromyscus maniculatus*. *Canadian Journal of Zoology*, 66(1), 277–278.
- Green, D. J., Richmond, B. G., & Miran, S. L. (2012). Mouse shoulder morphology responds to locomotor activity and the kinematic differences of climbing and running. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 318(8), 621–638.

- Harney, B. A., & Dueser, R. D. (1987). Vertical stratification of activity of two *Peromyscus* species: an experimental analysis. *Ecology*, 68(4), 1084–1091.
- Head, A. W., Hardin, J. S., & Adolph, S. C. (2012). Methods for estimating peak physiological performance and correlating performance measures. *Environmental and Ecological Statistics*, 19(2), 127–137.
- Herrel, A., & Bonneaud, C. (2012). Trade-offs between burst performance and maximal exertion capacity in a wild amphibian, *Xenopus tropicalis*. *Journal of Experimental Biology*, 215(17), 3106–3111.
- Huey, R. B., Dunham, A. E., Overall, K. L., & Newman, R. A. (1990). Variation in locomotor performance in demographically known populations of the lizard *Sceloporus merriami*. *Physiological Zoology*, 63(5), 845–872.
- Husak, J. F., Fox, S. F., Lovern, M. B., & Van Den Bussche, R. A. (2006). Faster lizards sire more offspring: Sexual selection on whole-animal performance. *Evolution*, 60(10), 2122.
- Jayne, B. C., & Bennett, A. F. (1990). Scaling of speed and endurance in garter snakes: a comparison of cross-sectional and longitudinal allometries. *Journal of Zoology*, 220(2), 257–277.
- Karlsson, J., Sjödin, B., Jacobs, I., & Kaiser, P. (1981). Relevance of muscle fibre type to fatigue in short intense and prolonged exercise in man. *Ciba Foundation Symposium*, 82, 59–74.
- Kormi-Nouri, R., Nilsson, L. G., & Ohta, N. (2005). The novelty effect: Support for the novelty-encoding hypothesis. *Scandinavian Journal of Psychology*, 46(2), 133–143.
- Lailvaux, S. P., & Husak, J. F. (2014). The life history of whole-organism performance. *The Quarterly Review of Biology*, 89(4), 285–318.
- Lappin, A. K., & Husak, J. F. (2005). Weapon performance, not size, determines mating success

- and potential reproductive output in the collared lizard (*Crotaphytus collaris*). *The American Naturalist*, *166*(3), 426–436.
- Le Galliard, J. F., & Ferrière, R. (2008). Evolution of maximal endurance capacity: Natural and sexual selection across age classes in a lizard. *Evolutionary Ecology Research*, *10*(2), 157–176.
- Lebreton, J., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs*, *62*(1), 67–118.
- Leussis, M. P., & Bolivar, V. J. (2006). Habituation in rodents: A review of behavior, neurobiology, and genetics. *Neuroscience and Biobehavioral Reviews*, *30*(7), 1045–1064.
- Magrath, M. J. L., & Elgar, M. A. (1997). Paternal care declines with increased opportunity for extra-pair matings in fairy martins. *Proceedings of the Royal Society B: Biological Sciences*, *264*(1389), 1731–1736.
- Márquez, R., & Cejudo, D. (2000). Defensive behavior as an escape strategy in four species of Gallotia (Sauria, Lacertidae) from the Canary Islands (Spain). *Copeia*, *2000*(2), 601–605.
- Martin, J. G. A., Pirotta, E., Petelle, M. B., & Blumstein, D. T. (2017). Genetic basis of between-individual and within-individual variance of docility. *Journal of Evolutionary Biology*, *30*(4), 796–805.
- Metzgar, L. H. (1967). An experimental comparison of screech owl predation on resident and transient white-footed mice (*Peromyscus leucopus*). *Journal of Mammalogy*, *48*(3), 387–391.
- Metzgar, L. H. (1973). Exploratory and feeding home ranges in *Peromyscus*. *Journal of Mammalogy*, *54*(3), 760–763.

- Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews*, 85(4), 935–956.
- O’Steen, S., Eby, S. L., & Bunce, J. A. (2010). Dynamically honest displays: Courtship locomotor performance indicates survival in guppies. *Functional Ecology*, 24(5), 1045–1053.
- Oufiero, C. E., Walsh, M. R., Reznick, D. N., & Garland Jr., T. (2011). Swimming performance trade-offs across a gradient in community composition in Trinidadian killifish (*Rivulus hartii*). *Ecology*, 92(1), 170–179.
- Petrie, M., & Kempenaers, B. (1998). Extra-pair paternity in birds: Explaining variation between species and populations. *Trends in Ecology and Evolution*, 13(2), 52–57.
- Reznick, D., Nunney, L., & Tessier, A. (2000). Big houses, big cars, superfleas and the costs of reproduction. *Trends in Ecology and Evolution*, 15(10), 421–425.
- Roche, D. G., Careau, V., & Binning, S. A. (2016). Demystifying animal ‘personality’ (or not): why individual variation matters to experimental biologists. *Journal of Experimental Biology*, (November).
- Sardell, R. J., Arcese, P., Keller, L. F., & Reid, J. M. (2011). Sex-specific differential survival of extra-pair and within-pair offspring in song sparrows, *Melospiza melodia*. *Proceedings of the Royal Society B: Biological Sciences*, 278(1722), 3251–3259.
- Schaller, G. B. (1972). *The Serengeti Lion*. Chicago: University of Chicago Press.
- Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2), 103–113.
- Sneddon, L. U., Huntingford, F. A., Taylor, A. C., Orr, J. F., Sneddon, L. U., Huntingford, F. A., & Taylor, A. C. (2000). Weapon strength and competitive success in the fights of shore

- crabs (*Carcinus maenas*). *Journal of Zoology*, 250, 397–403.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Physiology*, 15(1), 72–101.
- Stearns, S. C. (1989). Trade-offs in life-history evolution. *Functional Ecology*, 3(3), 259–268.
- Stearns, S. C. (1992). *The Evolution of Life Histories*. Oxford: Oxford University Press.
- Stephens, P. R., & Wiens, J. J. (2008). Testing for evolutionary trade-offs in a phylogenetic context: Ecological diversification and evolution of locomotor performance in emydid turtles. *Journal of Evolutionary Biology*, 21(1), 77–87.
- Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review*, 2(3), 387–390.
- Van Damme, R., Wilson, R. S., Vanhooydonck, B., & Aerts, P. (2002). Performance constraints in decathletes. *Nature*, 415(6873), 755.
- van Noordwijk, A. J., & de Jong, G. (1986). Acquisition and allocation of resources: Their influence on variation in life history tactics. *The American Naturalist*, 128(1), 611–624.
- Vanhooydonck, B., Van Damme, R., & Aerts, P. (2001). Speed and stamina trade-off in lacertid lizards. *Evolution*, 55(5), 1040–1048.
- Walker, J. A., & Caddigan, S. P. (2015). Performance trade-offs and individual quality in decathletes. *Journal of Experimental Biology*, 218(22), 3647–3657.
- Wilson, A. J., & Nussey, D. H. (2010). What is individual quality? An evolutionary perspective. *Trends in Ecology and Evolution*, 25(4), 207–214.
- Wilson, R. S., Niehaus, A. C., David, G., Hunter, A., & Smith, M. (2014). Does individual quality mask the detection of performance trade-offs? A test using analyses of human physical performance. *Journal of Experimental Biology*, 217(4), 545–551.

- Wolak, M. E. (2012). nadiv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. *Methods in Ecology and Evolution*, 3(5), 792–796.
- Wolff, J. O. (1985). The effects of density, food, and interspecific interference on home range size in *Peromyscus leucopus* and *Peromyscus maniculatus*. *Canadian Journal of Zoology*, 63, 2657.
- Wolff, J. O., & Hurlbutt, B. (1982). Day refuges of *Peromyscus leucopus* and *Peromyscus maniculatus*. *Journal of Mammalogy*, 63(4), 666–668.
- Xia, X., & Millar, J. S. (1991). Genetic evidence of promiscuity in *Peromyscus leucopus*. *Behavioral Ecology and Sociobiology*, 28(3), 171–178.

Tables and Figures

Table 1. Estimates and 95% confidence intervals (CI) extracted from two separate univariate mixed models of (A) grip strength and (B) sprint speed (both standardised to a mean of 0 and a variance of 1) in wild white-footed mice (*Peromyscus leucopus*). Fixed effects include body mass, capture type (trap/nest box), sex (male/female), age (juvenile/adult), reproductive status (active/not active), an interaction between sex and reproductive status, time of day, test sequence, and trial sequence. For sprint speed, models contained additional fixed effects of motivation, interaction of motivation by trial sequence, and number of missing values per trial. Significant fixed effect estimates are bolded. Random effects variance components included individual identity (V_{ind}), test series (V_{test}), and residual (V_e). Using these variance components allowed for calculating long- and short-term repeatability.

Source	(A) Grip Strength			(B) Sprint Speed		
	Estimate	95% CI		Estimate	95% CI	
		lower	upper		lower	upper
Fixed Effects						
Body mass	0.48	0.41	0.56	0.02	-0.05	0.10
Sex [male]	0.13	-0.01	0.28	0.03	-0.12	0.17
Age [juvenile]	-0.41	-0.54	-0.28	0.01	-0.12	0.13
Repro status [active]	-0.04	-0.21	0.12	0.25	0.09	0.41
Sex × Repro status	-0.21	-0.53	0.10	-0.27	-0.58	0.05
Capture type [trap]	0.25	-0.01	0.52	-0.13	-0.33	0.07
Time of day	-0.01	-0.08	0.07	0.04	-0.02	0.09
Test sequence	0.21	0.15	0.27	0.05	-0.01	0.11
Trial sequence	0.00	-0.03	0.02	0.06	0.04	0.08
Motivation [high]				0.66	0.61	0.70
Motivation × Trial				0.12	0.08	0.17
Missing values				0.12	0.11	0.13
Variance						
V_{ind}	0.10	0.05	0.15	0.10	0.06	0.14
V_{test}	0.14	0.10	0.18	0.11	0.08	0.15
V_e	0.28	0.25	0.30	0.37	0.35	0.39
Repeatability (R)						
Long-term	0.19	0.11	0.27	0.17	0.11	0.23
Short-term	0.46	0.41	0.52	0.36	0.31	0.42

Table 2. Estimates and 95% confidence intervals (CI) extracted from the bivariate mixed model of daily best (A) grip strength and (B) sprint speed (both standardised to a mean of 0 and a variance of 1) in wild white-footed mice (*Peromyscus leucopus*). Fixed effects are the same as in Table 1 (except for trial sequence and motivation \times trial, which were not fitted here because only the maximum value was retained per test). Given that $N = 167$ individuals had a total of 389 repeated paired measurements of grip strength and sprint speed, the bivariate mixed model included random effects variance components for individual identity (V_{ind}) and residual (V_e), as well as an among-individual correlation (r_{ind}) and a residual (within-individual) correlation (r_e).

Source	(A) Grip Strength			(B) Sprint Speed		
	Estimate	95% CI		Estimate	95% CI	
		lower	upper		lower	upper
Fixed Effects						
Body mass	0.51	0.42	0.60	0.08	-0.05	0.22
Sex [male]	0.17	0.01	0.33	0.05	-0.19	0.29
Age [juvenile]	-0.49	-0.65	-0.33	0.01	-0.22	0.23
Repro status [active]	-0.03	-0.22	0.16	0.41	0.11	0.70
Sex \times Repro status	-0.21	-0.58	0.16	-0.32	-0.88	0.24
Capture type [trap]	0.30	-0.01	0.61	-0.13	-0.51	0.24
Time of day	-0.02	-0.11	0.07	0.10	-0.01	0.20
Test sequence	0.24	0.16	0.31	0.03	-0.08	0.14
Motivation [high]				0.88	0.65	1.12
Missing values				0.04	0.00	0.08
Variance						
V_{ind}	0.10	0.04	0.16	0.20	0.09	0.32
V_e	0.32	0.26	0.37	0.62	0.51	0.73
Repeatability (R)	0.24	0.12	0.36	0.25	0.12	0.37
Correlation						
r_{ind}		-0.39 (-0.68, -0.08)				
r_e		0.07 (-0.03, 0.16)				

Figure captions

Figure 1. A) Frequency and B) temporal distribution of repeated measures made on wild white-footed mice (*Peromyscus leucopus*) on Cow Island, Ontario, Canada, in 2016. In (B), repeated observations made on a given individual are connected by a grey line, showing how different individuals were repeatedly sampled throughout the summer (a shallow vs. steep slope represent a longer vs. shorter delay between measurements in comparison, respectively).

Figure 2. Grip strength and sprint speed as function of (A-B) body mass (g), (C-D) test sequence (repeated tests on different days; see Fig. 1), and (E-F) trial sequence (repeated trials within a test) in wild white-footed mice (*Peromyscus leucopus*). Shown are partial residuals adjusted for the other fixed and random effects in the model (see *Statistical analysis*). A solid trend line indicates a significant relationship.

Figure 3. Representation of the (A) among-individual correlation (r_{ind}) and (B) within-individual correlation (r_e) between grip strength and sprint speed in 186 wild white-footed mice (*Peromyscus leucopus*). In (A), individual deviations from the population mean are illustrated using best linear unbiased predictors ($BLUP \pm SE$; red dots) associated with the random effect of individual identity. In (B), deviations from individual means are represented as the residuals (open circles) extracted from a bivariate mixed model.

Figure 4. Individual maximum grip strength and sprint speed values as a function of total number of repeated measures. Presented are the maximum performance measurements for grip strength and sprint speed of every individual ($N_{grip} = 186$, $N_{sprint} = 160$; grey circle) and the linear regressions for each performance trait as a function of total repeated measures (grip strength: $R^2 = 0.17$, $P < 0.001$; sprint speed: $R^2 = 0.19$, $P < 0.001$). In addition, population mean values (black dots \pm standard errors) are given for each number of repeated measures.

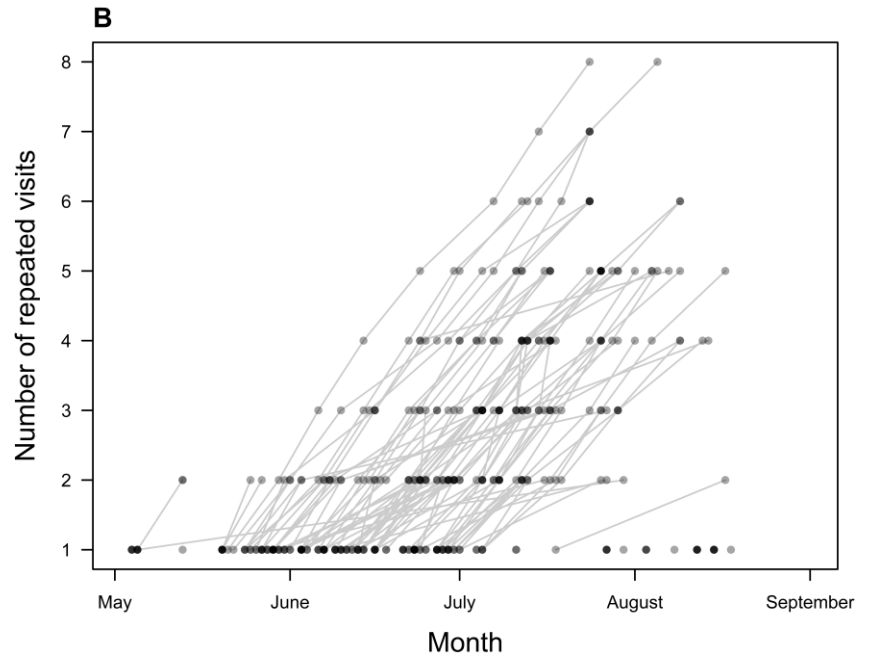
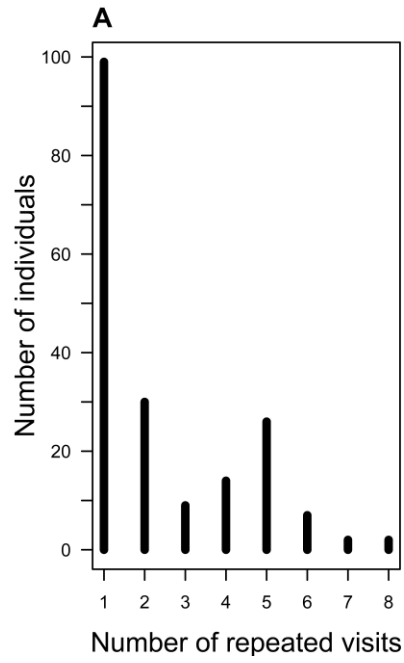


Figure 1

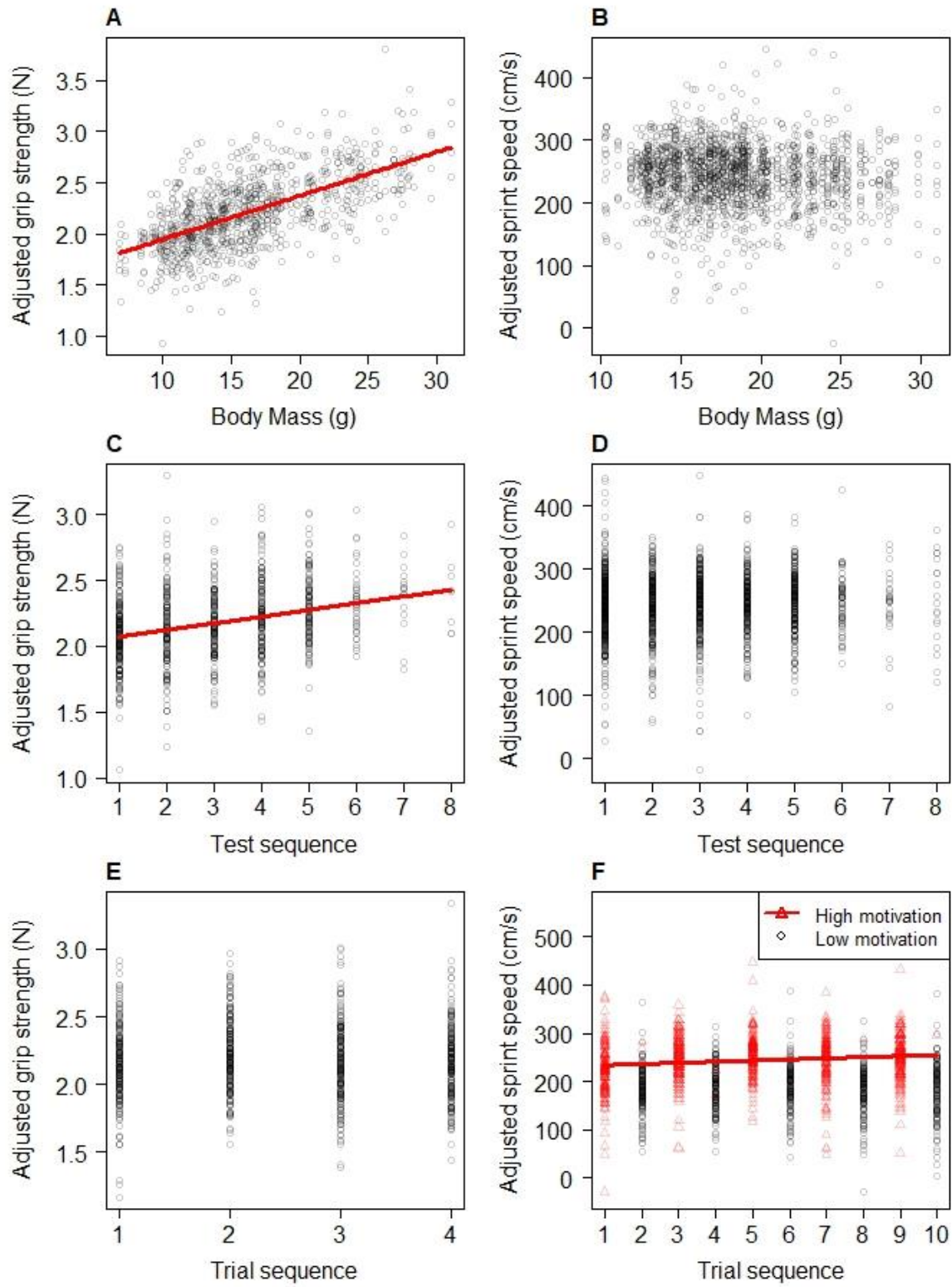


Figure 2

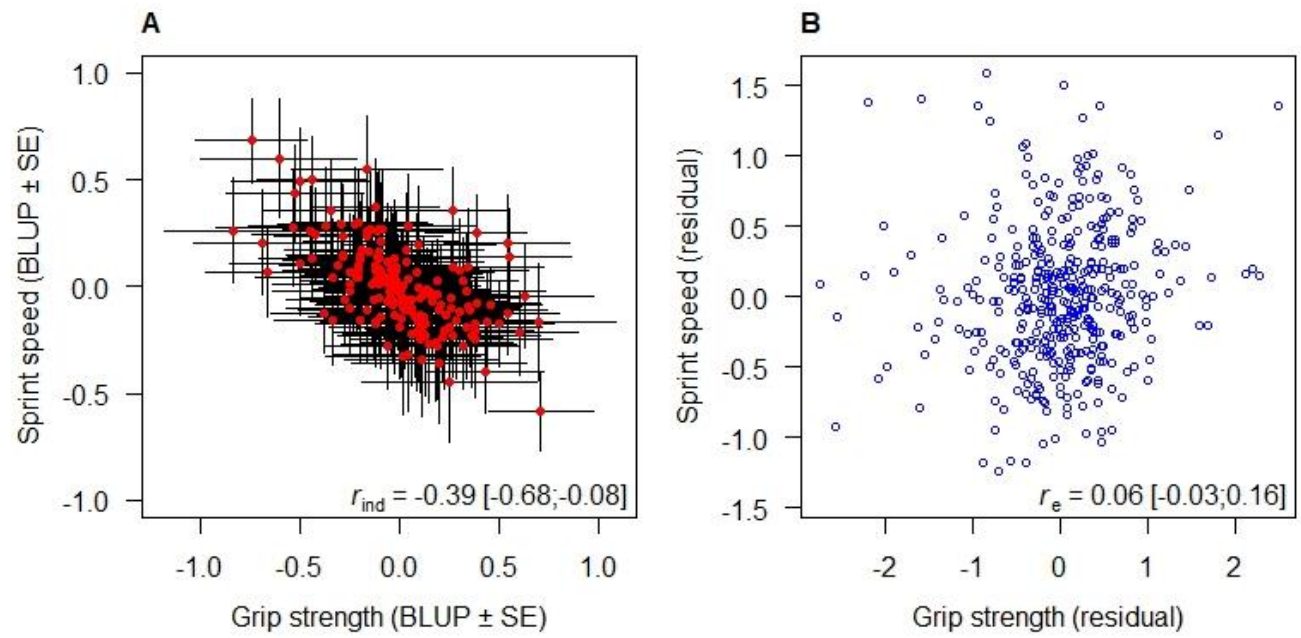


Figure 3

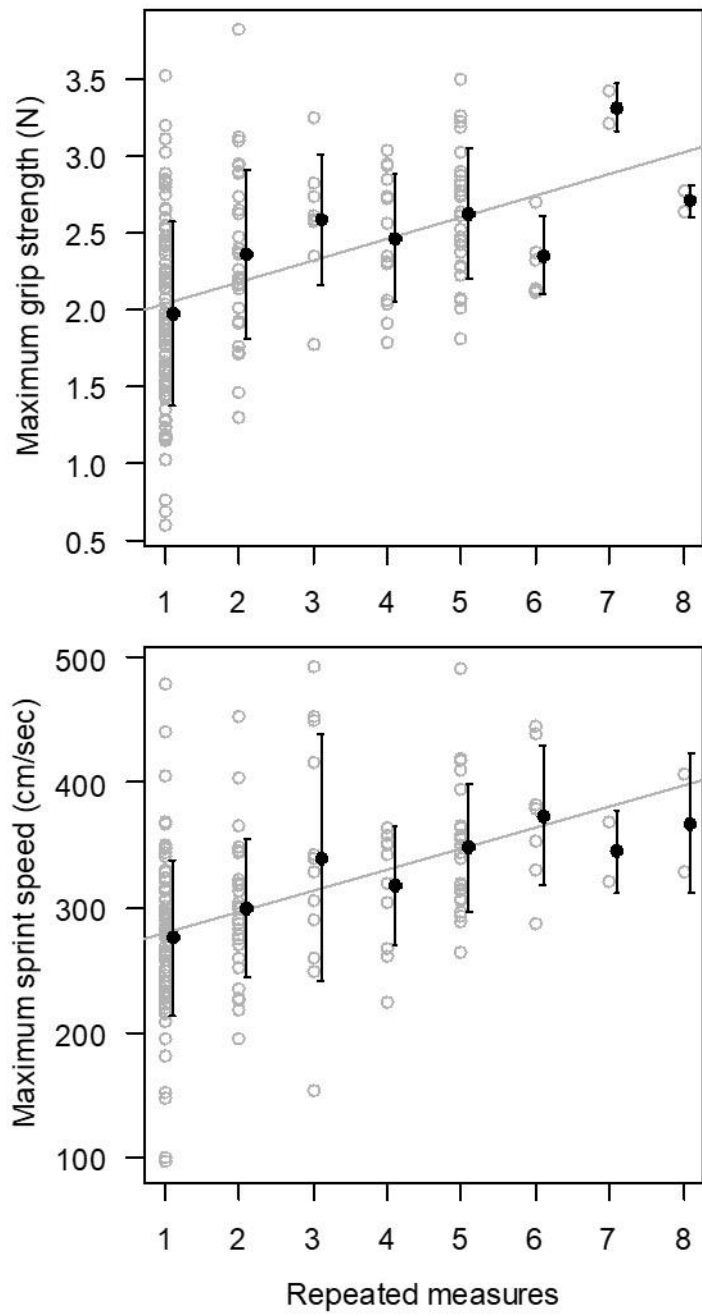


Figure 4

Appendix

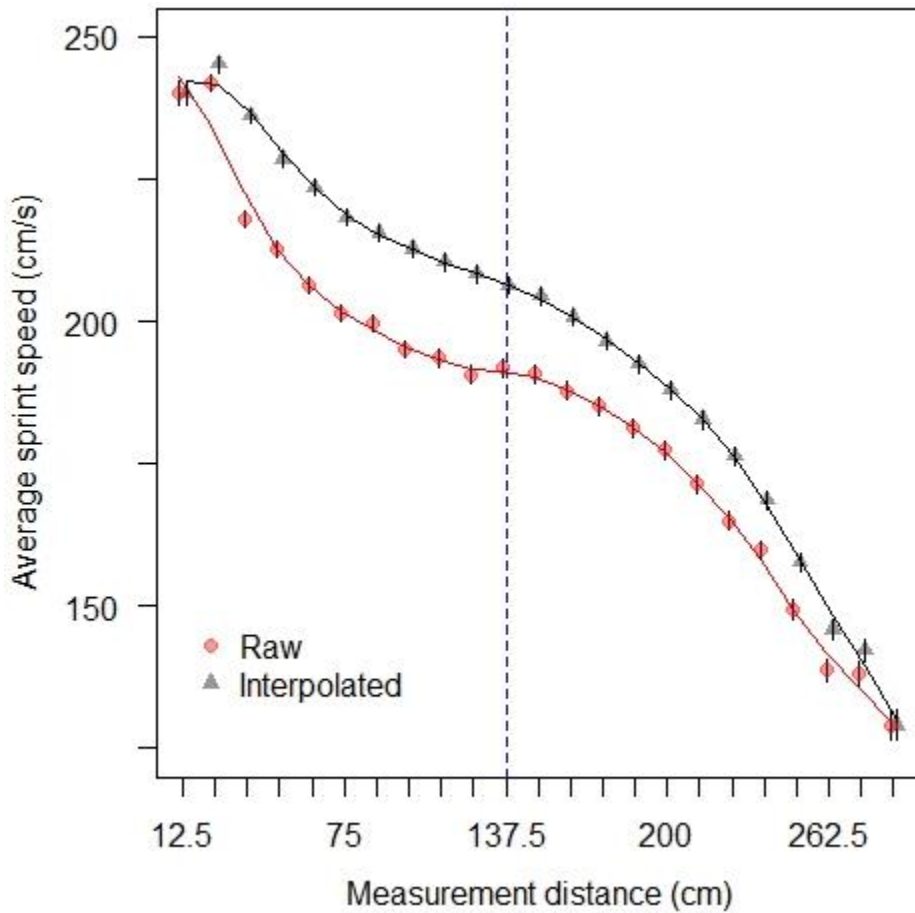


Figure S1. Average \pm standard error of all maximum sprint speed values (across all trials; $n = 3,515$) as a function of the distance used to calculate the running speed on a 3m long race track with 24 sensors equally spaced along the track (12.5 cm). Raw values encompass all trials in which enough sensors were triggered to calculate raw speed. Interpolated values include all detection times as well as interpolated time values between two detections, given that they do not exceed the given distance. The dashed vertical line indicates the measurement distance closest to the inflection point for both curves.