Does Behavior Evolve First? Correlated Responses to Selection for Voluntary Wheel-Running Behavior in House Mice

Running Head: Does Behavior Evolve First?

Rahim H. Khan[1]
Justin S. Rhodes[2]
Isabelle A. Girard[3]
Nicole E. Schwartz[1,4]
Theodore Garland, Jr.[1,*]

[1] Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, California 92521

[2] Department of Psychology, University of Illinois, Champaign, Illinois 61820

[3] Research Core Facilities Program, University of California, Davis, California 92521

[4] Present Address: Biological Sciences, Chapman University, Orange, CA 92866

[*] Corresponding author; email: tgarland@ucr.edu; phone 951-827-3524

Keywords:
adaptation, artificial selection, behavior, life history, performance, trade-off.

Opening Quotation

"It is unavoidable that we biologists, because of our limitations, divide ourselves into categories of specialization and then pretend that these categories exist in the biological world. As everyone knows, organisms are functionally indivisible and cannot be split into the conventional compartments of morphology, physiology, behaviour and genetics. Each of these is only one aspect of the organism as a whole and since it is the organism which deals with the physical environment, where do we start?" (Bartholomew 1964, p. 8)

ABSTRACT

How traits at multiple levels of biological organization evolve in a correlated fashion in response to directional selection is poorly understood, but two popular models are the very general "behavior evolves first" (BEF) hypothesis and the more specific "morphology - performance - behavior - fitness" (MPBF) paradigm. Both acknowledge that selection often acts relatively directly on behavior and, when behavior evolves, other traits will as well, but most with some lag. However, this proposition is exceedingly difficult to test in nature. Therefore, we studied correlated responses in the High Runner (HR) mouse selection experiment, in which 4 replicate lines have been bred for voluntary wheel-running behavior and compared with 4 non-selected Control (C) lines. We analyzed a wide range of traits measured at generations 20-24 (with a focus on new data from generation 22), coinciding with the point at which all HR lines were reaching selection limits (plateaus). Significance levels (226 P values) were compared across trait types by ANOVA and we used the positive False Discovery Rate (pFDR) to control for multiple comparisons. This meta-analysis showed that, surprisingly, the measures of performance (including maximal oxygen consumption during forced exercise) showed no evidence of having diverged between the HR and C lines, nor did any of the life history traits (e.g., litter size), whereas body mass had responded (decreased) at least as strongly as wheel running. Overall, results suggest that the HR lines of mice had evolved primarily by changes in motivation, rather than performance ability, at the time they were reaching selection limits. In addition, neither the BEF nor the MPBF models of hierarchical evolution provide a particularly good fit to the HR mouse selection experiment.

1. Introduction

How traits at multiple levels of biological organization evolve in a correlated fashion in response to directional selection is poorly understood. The “behavior evolves first” hypothesis (BEF) (Rhodes and Kawecki 2009) recognizes that, in nature, selection often acts relatively directly on behavior and that behavior appears to be evolutionarily labile (Blomberg et al. 2003; but see Revell et al. 2008). Darwin (1859, Chapter 6) might be credited with the earliest mention of BEF when he singled out the Dipper as a bird for which "the acutest observer by examining the dead body … would never have suspected its sub-aquatic habits. … In such cases, and many others could be given, habits have changed without a corresponding change of structure" (but see Smith et al. 2022). Other quotations to this effect are easy to find, such as "Behavior is an animal's way of interacting with its environment and it is therefore a prime target for natural selection" (Crusio 1995, p. 323). Selection acting on behavior can have far-reaching implications. To quote Mayr (1982, p. 612), "Many if not most acquisitions of new structures in the course of evolution can be ascribed to selection forces exerted by newly acquired behaviors ... Behavior, thus, plays an important role as the pacemaker of evolutionary change." He goes on to claim that "Most adaptive radiations were apparently caused by behavioral shifts." On the other hand, Mayr (1958, p. 356) also noted that "there is no general answer to the question, 'Structure first or behavior first?'"

In an influential paper, Arnold (1983) discussed how selection can be measured in natural populations (Lande and Arnold 1983) through the use of path analysis and noted the key role of whole-organism performance abilities (e.g., how fast an animal can run when maximally motivated) as transducers between lower-level traits and Darwinian fitness (survival and reproductive success). He did not explicitly recognize the pivotal role of behavior in the hierarchy of biological organization, and instead "use[d] 'morphology' as a shorthand for any measurable or countable aspect of structure, physiology or behavior." Since then, various workers have expanded on Arnold's original ideas, in particular by arguing that behavior is a key potential "filter" between performance abilities and components of Darwinian fitness (Garland et al. 1990). The argument posits that, in nature, selection often acts on what an animal does in a given situation, such as when it encounters a predator, i.e., its behavioral choices (e.g., see Bateson 1988). For example, if a rabbit chose to remain motionless as a predator approached, such that crypsis might allow it to

avoid detection, then its coloration would be under correlational selection along with its behavior (Brodie III 1992). However, in this scenario, the rabbit's maximal sprinting ability would be irrelevant. On the other hand, if the rabbit chose to run away at top speed when it encountered a predator, then the behavioral choice along with its maximal sprinting ability would be under correlational selection. Similar arguments have been made in the context of thermoregulatory behavior (Huey et al. 2003). This larger view of how selection acts at different levels essentially encompssses the BEF hypothesis and has come to be known as the “morphology - performance - behavior - fitness" (MPBF) paradigm (Garland and Carter 1994; Garland and Losos 1994; Careau and Garland 2012; Lailvaux and Husak 2014; Storz et al. 2015), which has also been referred to as the “mechanism - performance - behavior - fitness" paradigm (Dantzer et al. 2016). The MPBF paradigm (an expanded version specific to the present study is shown in Fig. 1) is consistent with the idea that behavior may generally be "in the vanguard of evolution" (Plomin 1990, p. 183). However, it is also consistent with the possibility that behavior can sometimes inhibit evolution (Huey et al. 2003; Duckworth 2009).

Behavior, of course, cannot evolve without changes in aspects of the brain and neurobiology (e.g., motivation and reward circuitry, the neural control of muscular contractions), and so some neurobiological traits that underpin behavior (which can be termed "lower-level" or "subordinate" traits) will also evolve, such as the sizes of particular brain regions or the density of receptors for specific neurotransmitters (Bronikowski et al. 2004; Katz 2011; Fischer and O’Connell 2017; Sheehan et al. 2018; Schmill et al. 2023). If the expression of a behavior evolves enough, then it will become eventually reach a point at which further expression becomes limited or constrained by whole-organism performance abilities (Bennett 1989; Horning 2012), which will limit further evolution of the behavior in question. If this leads to evolution of the performance abilities (e.g., endurance capacity), then subordinate traits that determine or constrain such abilities will also necessarily evolve (e.g., muscle contractile properties, circulating hormone levels, physical arrangements of bones, tendons, and muscles) (Husak et al. 2009; Taylor and Thomas 2014; Garland et al. 2016; Green et al. 2021).

Given that behavioral evolution requires at least neurobiological change and, with large behavioral shifts, also changes in performance abilities that are determined by various aspects of morphology and physiology (e.g., endocrine function), life history traits might also

be affected. For example, an evolutionary change in activity levels might involve changes in the motivation for, or reward perceived from, physical activity (Kuhn et al. 2016; Lightfoot et al. 2018; Klimentidis et al. 2022). Such changes would likely involve brain regions (particularly those in the reward system) that are also involved in the control of feeding, social behavior or maternal care (Kalivas and Nakamura 1999; Kelley and Berridge 2002; Fischer and O'Connell 2017; Schmill et al. 2023). Changes in feeding behavior (e.g., appetite) could have implications for energy balance, growth rate, and body size, which could affect maturation rate and litter size. Of course, pup mass at weaning and related fitness components (Fig. 1) would also be affected by maternal care behavior. These sorts of changes in other traits can also be viewed through the lens of pleiotropic gene action and the likelihood of "universal" pleiotropy (Wagner and Zhang 2011), i.e., the idea that essentially every gene affects more than one trait, and perhaps all traits.

The purpose of the present study was to test for correlated evolution at multiple levels of biological organization (Fig. 1) in a long-term artificial selection experiment. Specifically, the High Runner mouse experiment includes four replicate HR lines bred for the average number of revolutions run on days five and six of a six-day exposure to wheels as young adults (Swallow et al. 1998a). The HR lines are compared with four non-selected Control (C) lines. All four HR lines responded rapidly to selection and reached apparent selection limits (plateaus) around generations 17-27, depending on line and sex (Careau et al. 2013, 2015), at which point they were running ~2.5-3-fold more revolutions per day than the C lines. Analyses separated by sex indicated that the timing of the selection limit was similar between females (generation 21.0 ± 4.2) (mean ±95% confidence interval) and males (generation 19.8 ± 3.1), but the height of the plateau (relative to C lines) was 28% higher in females (8,175 ± 1,130 revolutions per day) than males (6,385 ± 914.3 revolutions per day) (Careau et al. 2013). Therefore, we sampled mice from within this generational range to provide a "snapshot" of phenotypic divergence at or near a selection plateau. Although numerous previous publications on the HR mice have reported differences from the C lines at various generations (reviews in Garland 2003; Rhodes et al. 2005; Swallow et al. 2009; Wallace and Garland 2016), none has attempted to synthesize the differences at or near when selection limits were reached.

The core data set analyzed here was from generation 22 and includes three measures of whole-organism performance (maximal oxygen consumption during forced

treadmill exercise [VO_2max], maximal sprint speed, rotarod performance), in addition to basal metabolic rate (BMR), measures of open-field behavior from Bronikowski et al. (2001), blood hemoglobin and hematocrit, body mass, and the masses of various internal organs (Fig. 1). Data from adjacent generations (20-24) included life history traits (e.g., litter size, pup mass, growth rate), micro-analyses of wheel-running behavior (e.g., bout duration), physical activity in cages when wheels are not available, maternal care behavior, behaviors in tests related to food reward and learning (Supplemental Methods), wheel-running responses to stimulatory drugs, plasma corticosterone, skeletal measurements, and morphometrically estimated lung diffusing capacity (Weibel 1970/71, 1990) (see Methods for a full listing of traits and their categorizations).

Meta-analysis of this broad range of traits allows us to determine whether, as expected under both BEF and MPBF, we would see the amount of divergence in phenotypes decreasing in order from behavior through performance and down to subordinate traits. (Fig. 1 shows some of the traits included in the present study, more details are provided in the Methods section, and a complete listing is found in the online supplemental materials). We can also determine if other behaviors were altered, as would be expected if they share neural pathways with those involved in the motivational and physical control of locomotion and voluntary exercise (Garland et al. 2011b; Lightfoot et al. 2018). In addition, we can test if aspects of life history have evolved (Lailvaux and Husak 2014; Orr and Garland 2017). Finally, complex traits often encompass trade-offs (e.g., Glazier 2009; Cohen et al. 2020; Douhard et al. 2021; Garland et al. 2022). As argued previously (Girard et al. 2002), for several reasons one might expect trade-offs between physical activity and aspects of Darwinian fitness, such as growth rate, fertility, litter size or pup mass, and perhaps maternal care behavior. Another expected trade-off would be higher endurance ability leading to reduced maximal sprint speeds (Dlugosz et al. 2009; Castro et al. 2022a). On the other hand, we expected that High Runner mice might perform better on the rota-rod test of neuromuscular coordination, reflecting enhanced abilities that may be required for wheel running.

2. Methods

2.1. The High Runner Mouse Selection Experiment

Beginning in 1993, four replicate lines of house mice have been bred for high voluntary wheel-running behavior (HR lines), based on the average number of revolutions run on days five and six of a six-day exposure to wheels (Wahman type, 1.12 m circumference) attached to standard housing cages (Swallow et al. 1998a). The starting population was 224 individuals from the outbred, genetically variable, Hsd:ICR strain. Following two generations of random mating in our laboratory, mice were randomly divided into eight lines. Four of these were subsequently bred for high running and four were bred without regard to running as Control (C) lines. Each line has been maintained by 10 mating pairs per generation. Within-family selection is used and sibling-mating is disallowed in all lines.

This study examined various traits at generation 22, which corresponds approximately to the time that selection limits were being reached in all four HR lines (Careau et al. 2013). We would expect that traits important for wheel running would have diverged by this point in the experiment. The sample included 160 mice (half male and half female), 20 from each of the 8 lines, 1 male and 1 female from each of 10 families, that were randomly chosen.

2.2. Measurement Timeline for Generation 22

Mice were first measured in three separate batches (for logistical reasons), over a period of 38 days, for several traits, in the following order: maximal oxygen consumption (VO_2max), wheel running over 6 days, sprint speed, rota-rod performance, open-field behavior, and basal metabolic rate (BMR). Mice were bred after all breeder mice had been measured for BMR.

Dissections were conducted on non-breeder mice immediately after BMR was measured. Male breeder mice were dissected next, seven days after removal from being paired with females, and female breeders were measured last, seven days after weaning of their litters. Given the large effects that pregnancy and lactation have on organ masses and other aspects of physiology and anatomy, we did not perform combined-sex analyses for organ masses.

All procedures were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Wisconsin-Madison.

2.3. Voluntary Wheel-running Behavior

As in the routine selection protocol (Swallow et al. 1998a), each day we recorded revolutions in each 1-minute interval over a period of ~23.5 h. We then computed the total number of revolutions, the number of 1-minute intervals that had at least one revolution (minutes of wheel activity), and the mean speed of running (revolutions/interval). Here, we present analyses of the average values for days 5 and 6 of the 6-day test, i.e., the values that are used as the selection criterion. As in previous studies, a measure of wheel freeness was also included as a covariate in analyses of wheel running (e.g., Swallow et al. 1998a; Garland et al. 2011a), which was measured by accelerating each wheel to a constant speed and then recording the number of revolutions until it stopped (Copes et al. 2015).

2.4. Maximal Oxygen Consumption (VO_2max)

VO_2max sets an upper limit to the intensity of work that can be sustained aerobically, without the relatively rapid onset of fatigue (Seeherman et al. 1981; Dlugosz et al. 2013) and, in principle, should be limited by physical rather than motivational factors (but see Noakes 2004, 2012; Noakes and Gibson 2004). Hence, it would be expected to increase at some point as a correlated response to selection for wheel running, which is a sustained behavior that occurs for many hours per day.

The testing protocol had been used in several other studies in our lab (e.g., see Friedman et al. 1992; Hayes et al. 1992; Dohm et al. 1994, 2001; Swallow et al. 1998b), and was as follows. Two minutes of baseline data were collected on ambient air. A mouse was then placed in a small Plexiglas chamber held just above the surface of the treadmill belt, thus allowing inflow of room air. Chamber inner dimensions were 13 x 6.3 x 5 cm at the highest and 13 x 6.3 x 2 cm at the lowest portion of the wedge-shaped extension over an electrified grid. Mice were first placed in the chamber while the treadmill was stopped, and resting O_2 consumption (VO_2) was recorded for 1.5–2 min. The treadmill was then started at an initial speed of 1.5 km/h. Mice were induced to run by being prodded with a straightened paper clip inserted through a hole at the rear of the chamber and/or by a mild

electric current (50–110 V, 3–12 mA) provided through a horizontal grid of twelve 2-mm bars spaced 5 mm apart at the end of the moving belt. Treadmill speed was then increased every 2 min by 0.5 km/h. All mice reached at least 2.5 km/h; the maximum speed attained by any mouse in this study was 4.0 km/h. Trials were ended when VO_2 failed to increase with increasing speed and/or the mouse failed to keep pace with the treadmill. VO_2 generally decreased before a trial was ended (i.e., while the mouse was still running). After the treadmill was stopped, mice were left in the chamber for 1.5–2 min. Mice were then removed from the chamber, and baseline data were again recorded for 2 min. Body mass of each animal was recorded on the first day of measurement. Time of day, speed at which the trial ended, and a subjective assessment of trial quality (essentially cooperativity, scored as 5 categories from poor to excellent [1-5]) were recorded at the end of each trial (Swallow et al. 1998b).

2.5. Sprint Speed

Forced maximal sprint-running speed should be supported largely by anaerobic metabolism, and hence captures a different aspect of performance ability than does VO_2max. Sprint speed was measured on a 7-m long by 7.5 cm wide photocell-lined racetrack, with short-pile plastic artificial grass substrate (Friedman et al. 1992; Dohm et al. 1994). The race- track had sheet metal walls which extended above the substrate 27 cm on one side and 50 cm on the other. Twelve sets of photocells spaced at 0.5 m intervals (first set at 1.0 m from start) were interfaced to a computer. Each individual was tested five times in quick succession on each of two consecutive days. Mice were chased along the racetrack with a meter stick covered with cardboard (6.5 X 30 cm) as fast as they would run; they were then gently encouraged to walk back to the start of the track. The fastest 1.0 m intervals (three consecutive photocell stations) from each trial day were compared to assess repeatability; the single fastest 1.0 m ever recorded was analyzed as "maximal" sprint running speed (Garland et al. 1995).

2.6. Rota-rod

The rota-rod is a standard performance measure used to indicate motor coordination or fatigue resistance (Dunham and Miya 1957; Kinnard and Carr 1957; Weaver and Miya 1961; Jones and Roberts 1968; Sofia 1969; Norton 1982). Mice were tested on the Jones &

Roberts 7650 accelerating rota-rod treadmill (manufacturer Ugo Basile). Mice were placed individually into one of the five sections that was rotating at the minimum speed of 5 RPM. Total time for placing all five mice onto the rota-rod was kept to less than 30 seconds. Mice were swung by their tails from a position beneath the cylinder so that they would voluntarily climb atop the already rotating cylinder (mice will struggle if the experimenter attempts to place the mouse on the cylinder from above). When the last of the mice was placed on the cylinder, the rota-rod was switched to accelerating mode and timing was started; the cylinder accelerated constantly from 5 rpm to 50 rpm during the first 7 minutes, after which it steadily rotated at 50 rpm. Counter-trip plates built into the base of the rota-rod detected the time after starting that the mouse fell off the rotating cylinder, with an arbitrary upper limit of 600 seconds. The rota-rod was cleaned with a sponge between trials. After every third or fourth trial the counter-trip plates were also cleaned. Upon the completion of each day's trial, the rota-rod and plates were cleaned with ethanol.

2.7. Basal Metabolic Rate (BMR)

Basal metabolic rate sets a lower limit to the rate of energy expenditure by a fasted endotherm during its normally inactive phase (Hulbert and Else 2004; McNab 2012). Mice were fasted overnight and placed in glass metabolism chambers the next morning. The chambers were part of an open-circuit respirometry system. Up to seven mice were monitored simultaneously. Each mouse and a control chamber received dry air at 200 cm^3 min^{-1} from upstream thermal mass flow controllers (Sierra Instruments, Inc., Monterey, California, SideTrack Model 844). Water and CO_2 were removed from the excurrent air with Drierite and Ascarite, respectively. Excurrent air from each chamber was monitored every 5s for at least 7.5min of each hour (more if fewer than seven mice were being measured) by an Applied Electrochemistry S-3A/II oxygen analyzer (Ametek, Pittsburgh, Pennsylvania) interfaced to a microcomputer. Air was diverted by an automated system with solenoid valves under programmed control. We calculated VO_2 for the last 5 min before switching to the next chamber using the appropriate equation given by Hill (1972). With the flow rate and chamber volumes we used, this protocol ensured clearing of the respirometry system downstream of the metabolism chamber before the start of the 5-min interval. The data analysis program corrected for drift in the control channel (baseline) using linear regression to calculate predicted baseline values throughout the course of the intervening sampling.

The analysis program calculated the lowest and second lowest 5-min intervals of oxygen consumption of the day for each mouse (Hayes et al. 1992; Dohm et al. 2001).

2.8. Open-Field Behavior

We reanalyzed data from Bronikowski et al.(2001) with SAS Procedure Mixed, as noted below in section 2.12.

2.9. Life History Traits for Females

Analyses of litter size, total litter mass, and mean pup mass followed the procedures outlined previously for generation 21 (Girard et al. 2002), except that SAS Procedure Mixed was used rather than SAS Procedure GLM.

2.10. Organ Masses, Hematocrit, Blood Hemoglobin Content

Methods followed Carter et al. (1999). Heparinized microcapillary tubes were used to take blood samples from the sub-orbital sinus. Tubes were centrifuged for 6.5 min in a Clay-Adams Autocrit Ultra 3 microfuge. Hematocrit (Hct) was determined immediately following centrifugation. For measurement of hemoglobin ([Hb]), 25 ml blood samples (drawn from an additional heparinized microcapillary tube) were added to 5 ml of Drabkin's reagent. Concentration of cyanmethemoglobin was determined at 540 nm with a Beckman spectrophotometer (Sigma Technical Bulletin No. 525) and human hemoglobin standard (Sigma Catalog No. 525-18). Hct and [Hb] were determined in duplicate, and means analyzed.

After blood sampling, mice were sacrificed by cervical dislocation. Following a midventral incision, the heart was lifted with forceps and the ventricles cut free from the atria and major blood vessels. The ventricles were blotted and any coagulated blood removed. The gall bladder was excised before removing the liver for weighing. Finally, the right triceps surae (which includes the lateral and medial heads of the gastrocnemius muscle, the soleus, and the plantaris [also known as the flexor digitorum superficialis]) was removed by cutting the muscle from the lateral condyle of the tibia and medial condyle of the fibula, followed by cutting the Achilles' tendon approximately midway between its origin and the muscle's insertion. Wet mass of tissues was recorded to the nearest 0.1 mg on an

electronic balance; tissues were then frozen on dry ice, and stored at −80°C (Carter et al. 1999).

2.11. Statistical Analyses of Generation 22

Because some of the mice were dissected after being bred, the statistical analyses for the various organ masses and measures of hemoglobin and hematocrit were separated by sex and we used breeder versus non-breeder as a cofactor (nuisance variable). For traits measured before breeding, we performed both separate-sex and combined-sex analyses. Following numerous previous studies of these lines of mice, data were analyzed as mixed models in SAS Procedure Mixed, with REML estimation and Type III Tests of Fixed Effects (Copes et al. 2015; Acosta et al. 2017; Cadney et al. 2021; Castro et al. 2022b). Linetype (HR vs. C), sex, and mini-muscle status (see below) were fixed effects. Replicate line was a random effect nested within linetype. Effects of linetype were tested relative to line, with 1 and 6 degrees of freedom, as dictated by the design of the selection experiment. For analyses of the wheel-running traits, we allowed for separate estimates of the among-line variance because we have previously shown this to differ between the HR and C lines (Garland et al. 2011a). Effects of sex and of the linetype-by-sex interaction were also tested with 1 and 6 degrees of freedom.

Mini-muscle status was tested relative to the residual d.f. As previously described, the mini-muscle phenotype is characterized primarily by a 50% reduction in hindlimb muscle mass (Garland et al. 2002; Houle-Leroy et al. 2003). The underlying genetic variant is a C-to-T transition located in a 709-bp intron between exons 11 and 12 of the *Myosin heavy polypeptide 4* gene (Kelly et al. 2013) that behaves as a simple Mendelian recessive (Garland et al. 2002). Mini-muscle status was determined based on a comparison of triceps surae muscle masses in relation to body mass (Garland et al. 2002).

Body mass was used as a covariate for VO_2max, sprint speed, rota-rod, organ masses and tail length, litter size, total litter mass, and mean pup mass. Additional covariates were age and, depending on the trait, time of day, and z-transformed (time of day) squared (orthogonal polynomial). Covariates were tested relative to the residual d.f.

Some variables were $\log_{10}$-transformed to improve normality of residuals, the homogeneity of their spread when used as independent variables or linearity of relations with covariates. Statistical significance was judged at $P \leq 0.05$. Least squares means

(LSMs) from SAS Procedure Mixed are presented to compare groups. Long after starting the selection experiment, simulations confirmed that statistical power to detect linetype differences was relatively low (Castro et al. 2021).

2.12. Compilation of Results for Generations 20-24 and Meta-analysis

In addition to the *P* values reported in this study for mice from generation 22, we obtained *P* values from both published and unpublished studies that used mice from generations 20-24. In all cases, we used *P* values from separate-sex analyses. For some traits previously published with analyses that used SAS Procedure GLM (Bronikowski et al. 2001; Girard et al. 2002), data were reanalyzed using SAS Procedure Mixed to make the *P* values directly comparable with the new analyses reported here. (Note that *P* values are monotonically related to measures of effect size and are suitable for use here because d.f. for testing effects of selective breeding were always 1 and 6.) We used the positive False Discovery Rate (pFDR) to control for multiple comparisons (SAS Procedure Multtest).

To compare divergence of traits at different levels of biological organization, we formed eight categories, recognizing that the traditional four categories in the MPBF paradigm do not capture the diversity of traits for which we had data. *Subordinate Traits* included, for one or both sexes, 51 measures of bone dimensions or masses (e.g., Kelly et al. 2006), 16 other morphological measures (e.g., organ masses, tail length), and 10 physiological measures (e.g., blood hemoglobin content, plasma corticosterone concentration). *Whole-organism Physiology* included basal metabolic rate (BMR), Respiratory Exchange Ratio (RER) at measurement of VO_2max, body mass change across the 6-day period of wheel access for choosing breeders, estrus cycle length (methods following Girard and Garland 2002), growth rate, and the response of wheel-running behavior to drugs targeting certain neurotransmitters (Rhodes et al. 2001). *Performance* encompassed VO_2max, maximal sprint speed, and rota-rod. *Behavior* included multiple aspects of the open-field test (Bronikowski et al. 2001), maternal care (Girard et al. 2002), cooperativity during VO_2max trials (following Swallow et al. 1998b), home-cage activity (Rhodes et al. 2001), play behavior (Whitehead et al. 2023), and food-reward behavior (see Supplemental Methods). *Wheel Running Other* are measures of wheel running other than during days 5 and 6 of the 6-day test used to select breeders. *Life History* includes litter

characteristics at birth and at weaning (from the present study and Girard et al. 2002). Finally, *Body Size* includes both mass and length as adults.

We used a two-way ANOVA (SPSS UNIANOVA procedure) of ranked P values (to reduce heteroskedasticity) to test for effects of trait type (N = 8 categories), sex, and their interaction.

3. Results

3.1. Wheel Running

Based on combined-sex analyses of transformed wheel revolutons (Supplemental Table S1), mice from the HR lines ran 2.19-fold (based on back-transformed LSMs) more revolutions/day than those from the Control lines (Fig. 2A, $P = 0.0017$) and females ran 1.54-fold more revolutions than males ($P = 0.0008$), with no interaction between linetype and sex ($P = 0.3350$) and no effect of mini-muscle status ($P = 0.2825$). For the duration of daily running (Fig. 2B), HR lines ran 1.24-fold more minutes per day than C lines ($P = 0.1059$) and females ran 1.32-fold more minutes than males ($P = 0.0003$), with no significant interaction between linetype and sex ($P = 0.3093$) and no effect of mini-muscle status ($P = 0.1598$). Mice from HR lines ran at ~1.71-fold higher average speeds (Fig. 2C, $P = 0.0001$) and females ran ~1.18-fold faster than males ($P = 0.1138$), with a significant interaction ($P = 0.0163$), such that the linetype difference was +91% for females, but only +51% for males. Maximum wheel-running speed (Fig. 2D) showed a pattern similar to that for average speed, with a significant effect of both linetype ($P = 0.0002$, +48%) and sex ($P = 0.0103$, females +16%), and an interaction ($P = 0.0279$), where the linetype difference was +61% for females and +34% for males.

3.2. Organismal Performance and BMR

Neither VO_2max (with body mass as a covariate) nor the score of trial quality was significantly related to linetype, sex or mini-muscle status (Supplemental Table S1, see also Supplemental Table S2). Maximal sprint speed also did not differ between the linetypes or sexes, but mini-muscle mice had reduced sprint speeds compared to normal-muscled individuals ($P = 0.0026$) in the combined-sex analyses. Rota-rod performance was unaffected by linetype, sex or mini-muscle. Body mass was negatively related to rota-rod

performance (P = 0.1162 for females, P = 0.0025 for males). Wth body mass as a covariate, females had significantly higher BMR (P = 0.0049), and mini-muscle had higher BMR than normal mice (P < 0.0001) (Supplemental Table S1).

3.3. Open-Field Behavior

Of 11 different measures of open-field behavior, the only significant or near-significant effects of linetype were for turning behavior, where, for both sexes, mice from the Control lines turned in both directions more frequently than did those from the HR lines (the four P values ranged from 0.0373 to 0.0641: Supplemental Table S1). However, the difference between the number of right and left was not affected by linetype or mini-muscle status.

3.4. Body Size, Growth Rate, and Mass Change during Wheel Running

As presented in Supplemental Table S1 and Supplemental Table S2, adult mice were weighed at several times (Fig. 3) and the consistent result was that HR mice were significantly smaller than controls, females were smaller than males, and mini-muscle mice smaller than normal-muscled mice. However, weaning body mass was not significantly affected by linetype, sex or mini-muscle. Growth rate (g/day) from weaning to adult wheel testing was reduced in HR mice and in females, as well as in mini-muscle individuals (Fig. 3A). Body mass at dissection was significantly reduced in HR mice of both sexes.

All mice tended to lose body mass across the six days of wheel access (Fig 3D), and, for females, mice from the HR lines lost less mass than did C mice.

Based on separate-sex analyses, body length measured at the time of dissection was significantly reduced in HR mice of both sexes (Supplemental Table S2). With body length as a covariate, mini-muscle individuals of both sexes had lower body masses (Supplemental Table S2).

3.5. Life History Traits for Females

Dams from the HR lines were smaller at weaning than those from C lines, but we found no statistically significant differences for litter size, mean pup mass, total litter mass (all with dam body mass as a covariate) or percent females (Supplemental Table S2).

3.6. Organ Masses, Hemoglobin, Hematocrit

Adult mice were dissected after being bred. All organ masses, and tail length, were significantly positively related to body mass. Based on separate-sex analyses, the only statistically significant linetype effect was for males, where HR mice had larger livers (Supplemental Table S2).

With body mass as a covariate, mini-muscle individuals of both sexes had smaller triceps surae muscles (as expected), but larger heart ventricles, kidneys, and livers, as well as longer tails. Spleen mass was significantly larger only for females. Hematocrit was significantly lower than for normal-muscle individuals for both sexes; blood hemoglobin content trended in the same direction.

3.7. Meta-analysis of Significance Levels by Trait Type

A total of 226 *P* values from traits measured at generations 20-24 were available (Supplemental Table S3). For analysis, they were placed into eight categories, roughly corresponding to the levels of biological organization shown in Figure 1. Subordinate traits (N = 77) includes various aspects of morphology (e.g., bone dimensions, organ masses) and physiology [e.g., circulating concentration of corticosterone (Girard and Garland 2002), pulmonary diffusing capacity (T. Garland, Jr., I. Girard, J. S. Rhodes, and S. F. Perry, unpublished results)]. Whole-organism physiology (N = 18) includes such traits as growth rate of body mass, basal metabolic rate, Respiratory Exchange Ratio at VO_2max, and the response to drugs based on their wheel running. Performance (N = 7) includes VO_2max, sprint speed, rota-rod. Behavior (N = 69) includes open-field, maternal, home-cage activity, cooperativity during VO_2max tests, play, and food reward tests. Wheel running was separated into observations on days 5 and 6 of the standard 6-day tests used for choosing breeders (N = 26) versus measures taken on days 1-4 or over longer time periods (N = 8). Life history traits (N = 11) included offspring characteristics at birth and at weaning. From generation 20-21, we had live pups at birth, total litter mass at birth, mean pup mass at birth, litter size at weaning, total litter mass at weaning, mean offspring mass at weaning, and percent female at weaning (Girard et al. 2002). From generation 22 (present study), we had values at weaning for litter size, mean offspring mass, total litter mass, and percent female. Finally, we included 10 measures of adult body size (mass and/or length). Note that the

sexes were far from equally distributed among the trait types (see Supplemental Table S3). In some cases, this is because a given trait or set of traits is sex-limited (e.g., maternal behavior), but in others it reflects the fact that a given study only included a single sex. Overall, however, the P values were approximately equally split between the sexes (116 for females, 110 for males).

Figure 4 shows *P* values in relation to trait type (see Supplemental Table S3 for a complete listing). A two-way ANOVA (SPSS UNIANOVA procedure) of ranked P values (to reduce heteroskedasticity) by trait type (N = 8 categories) and sex indicated a strong effect of type ($F = 16.23$, d.f. = 7,211, $P < 0.001$), but no significant effect of sex ($F = 1.38$, d.f. = 1,211, $P = 0.242$) nor a trait type X sex interaction ($F = 0.61$, d.f. = 6,211, $P = 0.610$). This ANOVA failed a Levene's test for heteroskedasticity of residuals across cells ($P = 0.020$) because, as can be seen in Figure 4, some trait categories, such as Behavior, have much more variable P values than do others, such as Body Size. In any case, results were similar with bootstrapping. Note that, for this analysis, "Bones" and "Morphology, Physiology" in Figure 4 were included in the category of Subordinate Traits.

Figure 4 shows that, *P* values for measures of wheel running on days 5 and 6 were generally low (81% with $P < 0.05$), as would be expected, with the notable exception of female minutes/day (duration of daily running). The other eight measures of wheel running also had low *P* values (75% with $P < 0.05$).

For the subordinate traits, *P* values ranged from 0.0032 to 0.9843, with 22% of the 51 bone *P* values falling below 0.05 and 12% of the other morphology (e.g., organ masses) and physiology *P* values falling below 0.05. A third of the *P* values for whole-organism physiology were < 0.05, but none of the seven measures of whole-organism performance (e.g., maximal sprint speed) had $P < 0.05$. Only 10% of the 69 behavioral traits (other than wheel running) had a $P < 0.05$. Interestingly, no aspect of maternal behavior had $P < 0.05$, a pattern that extended to life history traits. Finally, the trait category having the strongest overall divergence between the HR and Control lines -- even more so than for wheel-running behavior measured on days 5 and 6 -- was body size, with 100% of the 10 measures having $P < 0.05$.

Of the 226 traits, 68 had a $P \leq 0.05$, with 22 passing a pFDR of 0.05 (with 0.0063 being the highest *P* value that passes) and 68 (coincidentally) passing a pFDR of 0.10 (with 0.0500 being the highest P value that passes).

4. Discussion

Overview

The goal of the present study was to examine correlated responses to selection around the time a selection plateau was being reached in the High Runner mouse selection experiment. To do so, we analyzed new data obtained at generation 22, and also compiled statistical results from both published and unpublished data from generations 20-24. We interpret our results in the context of two prevalent models of correlated evolution. Based on the MPBF and BEF models (see Introduction), we expected the greatest divergence between the HR and Control lines for wheel running and its components (average running speed, duration of daily activity), with less divergence in traits at the level of life history (e.g., litter size) and performance (maximal oxygen consumption during forced exercise [VO_2max], maximal sprint speed, rota-rod), and the least amount of divergence for such lower-level traits as organ masses and bone dimensions. However, our results indicate that neither the BEF nor the MPBF models are accurate predictors of what has occurred in the HR mice experiment. In particular, we found no evidence for divergence in organismal performance abilities (Fig. 4 and Fig. 5), which suggests that increases in motivation (a primary cause of the expression of behavior) drove the evolution of wheel-running behavior up to the point selection limits were being reached and also that performance ability in the base population (e.g., VO_2max) was higher than was being used. This interpretation is illustrated in Figure 6B.

Hierarchical Divergence between the HR and Control Lines

Of the 226 traits considered, 68 had nominal $P \leq 0.05$, and 22 of these pass a pFDR of 0.05 (Supplemental Table S2). Considering those 22, the vast majority (16) are, unsurprisingly, aspects of wheel-running behavior. Interestingly, the trait under selection (revolutions/day on days 5&6, $P = 0.0007$) did not have the lowest P value; rather, five other aspects of running had lower P values (range = 0.0002 to 0.0004). Most of these lower P values reflect the fact that both HR males and especially HR females have diverged from the C lines primarily in terms of running speed, not running duration (e.g., see Swallow et al. 1998a; Garland et al. 2011a; Claghorn et al. 2017b; Hiramatsu et al. 2017; Cadney et al.

2021). Various other aspects of running behavior pass the pFDR at 5%, including running distance over longer exposure durations (several weeks) and also on the first day of exposure. Thus, the increased wheel-running behavior of the HR lines is general and not restricted to the specifics of the selection criterion.

The only other behavioral trait that passes the pFDR of 5% is a measure of physical activity in the home cage when wheels are not provided (Ambulations Day 2, $P = 0.0004$). Studies at later generations have shown this to be a robust correlated response: when deprived of wheels, HR mice are more active in home cages, and this activity contributes to increased energy requirements and food consumption (e.g., Malisch et al. 2009; Copes et al. 2015; Acosta et al. 2017). Thus, the HR mice have been considered "hyperactive" in a general sense (e.g., Rhodes and Garland 2003; Rhodes et al. 2005). Given that no other behavioral traits pass the pFDR of 5%, we conclude that physical activity is a largely independent axis of behavior in these lines of mice.

Unexpectedly, the non-behavioral trait showing the greatest divergence between the HR and Control lines (as judged by *P* values) is adult body mass (both sexes), which reflects the lower growth rate of HR mice from weaning to adulthood (see also Girard et al. 2007). Adding body mass at the start of wheel testing to the separate-sex analyses shown in Supplemental Table S2 indicates it is a significant negative covariate of wheel running distance on days 5&6 for females (P = 0.0101) but not for males (P = 0.6545). However, mass and running distance are confounded because HR mice of both sexes run more and are smaller. In other generations, body mass is usually not significantly correlated with wheel-running behavior in these mice (e.g., Garland et al. 2011a). Thus, we suspect that the reduction in body mass is not part of the adaptive suite of traits that allow or promote wheel running, but rather caused by other changes, in particular increased circulating corticosterone concentrations (Girard and Garland 2002; Malisch et al. 2007; Singleton and Garland 2019). Experimentally elevated levels of glucocorticoids are known to suppress growth in mammals (e.g., see Dantzer et al. 2013), including in the HR and C lines of mice (Singleton and Garland 2019). In the present study, the decreased adult body mass of HR mice is attributable to reduced growth rate for both sexes between weaning, when mass does not differ (P = 0.9584), and adulthood. Circulating corticosterone can affect many traits, including aspects of both motivation and ability for wheel running, and whether the increased concentrations in HR mice are adaptive is unclear (Girard and Garland 2002;

Malisch et al. 2007; Garland et al. 2016; Singleton and Garland 2019). Another possibility is that the reward received from running "competes" with that received from eating. No experiments have been done to address this hypothesis explicitly, but some evidence regarding changes in dietary preferences might be interpreted as supportive of the possibility (Acosta et al. 2017; Thompson et al. 2018). Importantly, HR mice would not appear to face a limit that restricts the amount of food they can eat in a way that would reduce growth rates. First, as noted above, HR lines have evolved higher running distances mainly by running faster, so they are not time-limited relative to C mice. Second, both HR and C mice can eat much more food than they normally do even when housed with wheels if they are challenged with cold (Koteja et al. 2001; Vaanholt et al. 2007).

Previously at generation 20, we reported that HR dams were significantly smaller than C dams when giving birth (Girard et al. 2002). Dam mass at birth was significantly positively correlated with litter size and with total live litter mass, but not with mean pup mass. As HR dams were smaller when giving birth but had litter sizes similar to those of C dams, the former give birth to relatively larger litters for their body mass, although the difference is not statistically significant (Girard et al. 2002). In contrast, at generation 21, dam body size when their pups were weaned was not significantly smaller in HR than C mice, with no statistical differences in litter characteristics, and those results are confirmed here for generation 22 (Girard et al. 2002). Thus, we find no evidence that compromises to these life history traits could be related to the selection limits observed in the HR lines (Careau et al. 2013).

Several measures of hindlimb bone dimensions (corrected for variation in body mass) (Kelly et al. 2006; Wallace et al. 2012) have nominal $P \leq 0.05$, reflecting divergence in lower-level or subordinate traits that would be expected to affect running performance. A number of studies both before and since generation 22 have documented divergence in skeletal traits between the HR and C lines. For example, in a sample of both sexes at generation 11, we found that mice from HR lines had more symmetrical hindlimb bone lengths, larger femoral heads, along with relatively wider distal femora and deeper proximal tibiae, which suggests larger knee surface areas (Garland and Freeman 2005; Castro and Garland 2018). Moreover, HR mice had larger total femoral nutrient canal areas (Schwartz et al. 2018). Other differences have neen identified in other generations

(e.g., Middleton et al. 2008; Castro et al. 2021, 2022b), although the magnitude of difference can vary among generations (Castro et al. 2021).

Notably, none of the three measures of performance show statistically significant divergence, even though limb bone dimensions might reasonably be expected to affect maximal sprint speed or rota-rod performance, though not VO_2max (Fig. 4 and Fig. 5). In a previous study at generation 10, VO_2max was increased by 6% in male HR mice measured at a somewhat older age (P = 0.0190), which may represent a Type I error (Swallow et al. 1998b). In later generations, VO_2max has proven to be consistently higher in the HR lines (e.g., see Kolb et al. 2010; Hiramatsu et al. 2017; Singleton and Garland 2019; Cadney et al. 2021; Schwartz et al. 2023). Two previous studies are relvant here. First, in a study of 35 adult males from the Hsd:ICR strain (as used to start the selection experiment) VO_2max was positively correlated with subsequent wheel running, but never significantly so across seven days of wheel running (Friedman et al. 1992). Regardless of the lack of statistical significance, the tendency for a positive phenotypic correlation suggests that selection for high wheel-running behavior would also involve some selection for high VO_2max. In a study of adult male rats, none of three measures of performance (VO_2max, running endurance, maximal sprint speed) were predictive of subsequent wheel-running behavior (Lambert et al. 1996).

If we consider nominal $P \leq 0.05$, additional types of traits appear diverged between the HR and Control lines. Of particular interest are behaviors related to food reward, presumably reflecting alterations in the general motivational system of the HR lines, as has been well-documented in later generations (Rhodes et al. 2005; Belke and Garland 2007; Saul et al. 2017; Thompson et al. 2018; Schmill et al. 2023). Turning behavior of HR mice in the open-field test also has nominal $P \leq 0.05$, but what this represents is unclear, as it is not commonly reported in the context of "normal" rodent behavior (Bronikowski et al. 2001; Careau et al. 2012). Rather, turning behavior is typically associated with abnormal situations, such as neurological mutants, knockouts or other adverse manipulations (e.g. Fornaguera and Schwarting 2002; Kalueff et al. 2007) or induced stress responses (e.g. Mundorf et al. 2020, 2022). Conceivably, this behavior is related to the "acrobatics" that sometimes occur in our mice in wheels (see video that accompanies Girard et al. 2001). Studies of open-field behavior at later generations have generally not found statistical

differences between the HR and C lines, but they did not measure turning behavior (Jonas et al. 2010; Careau et al. 2012; Cadney et al. 2021).

The wheel-running response of HR lines to cocaine, which we view as an aspect of whole-organism physiology, also differs from that of C mice. This differential response suggests alterations in the dopaminergic neuromodulatory system (Rhodes et al. 2001) (and possibly serotonin signaling; see also Claghorn et al. 2016), an inference that has been supported by subsequent studies (e.g., see Bronikowski et al. 2004 for a study of dopamine receptor expression levels in the hippocampus at generation 27) and implicates changes in the reward system of the HR mice (Rhodes et al. 2005). Other morphological traits that likely have functional significance for wheel running differ between HR and C lines at $P \leq 0.05$, including several additional bone dimensions, the mass of the foot bones (Kelly et al. 2006), and liver and calf muscle masses.

How Well Do the Standard Models for Hierarchical Evolution Describe the HR Mouse Selection Experiment?

Both the BEF and MPBF models begin with the premise that selection in nature often acts most directly on behavior. Thus, on the face of it, they are germane to the High Runner mouse selection experiment, in which selection is imposed directly on voluntary wheel-running behavior (specifically, selection is imposed by linking wheel running directly to mating opportunity: Fig. 1). Neither model addresses the relative rate at which components of fitness (e.g., litter size) should evolve and so neither can really be inconsistent with our results, which find no changes in fitness components (see previous section). Similarly, neither model directly addresses the likelihood that other behaviors will evolve as a correlated response to selection on a particular behavior. In any case, we did find evidence for correlated responses in other behaviors (physical activity when wheels are not available, turning in the open-field test, food reward).

Under MPBF, organismal performance should evolve once behavior has evolved enough to begin taxing physiological abilities. Although we analyzed mice at a selection limit for a behavior, we found no evidence that the physical ability to express that behavior had evolved (measurements of VO_2max, sprinting abilities, and rota-rod performance). We interpret these results as consistent with the idea of motivation underpinning the evolution of

high wheel running up to this point, which also implies that performance abilities were "excessive" (Gans 1979) in the base population (see Fig. 6B).

Here it is important to note that we have not yet obtained direct measures of motivation for wheel running. Indeed, as noted by Bolles (1967, p. 1), "Although some writers have suggested behavioral criteria to define motivation, these attempts to specify what is meant by motivation are not very compelling." We agree with Bolles (1975) definition of motivation as a cognitive state for which multiple measures can be used to assess its level. The most appropriate measure may depend on the specific question, and it is likely that no single measure can adequately capture differences in motivation. With this theoretical background, we had planned a study in which mice would be trained to press a lever to free a brake on the wheels. Once this was accomplished, we planned to increase the number of presses required to free the brake until we reached a "giving up" point. We hypothesized that mice from the HR lines would have a higher giving up point, and that such a result could reasonably be interpreted as an indication of higher motivation for wheel running. After some preliminary trials, we eventually studied 32 female mice from generation 41 (Belke and Garland 2007). However, we found that the standard (from the literature) 90 s of wheel access as a reinforcer during training worked for almost all C mice, but for few of the HR mice. When reinforcer duration was increased to 30 min, almost all HR and C mice could be trained. We interpreted these results as evidence the selective breeding may have altered the motivational system in a way that reduces the reinforcing value of shorter running durations. We also offered the novel hypothesis that there may be a trade-off in the motivational system for activities of long versus short duration.

We did find divergence in some measures of whole-organism physiology, including growth rate from weaning to adulthood, body mass change during wheel access, and the wheel-running response to drugs that are known to affect reward-related behaviors, including wheel running (Rhodes et al. 2001, 2005), the last of these supporting the argument that motivation had evolved. We also found divergence in subordinate traits, including several measures of hindlimb bones. Thus, selection for wheel-running behavior has seemingly “skipped” over organismal performance, which is inconsistent with both the MPBF and BEF models (see also below in the Trade-offs section).

That conclusion should be tempered by the fact that we did not measure endurance capacity directly in the present study. However, in an unpublished study of swimming

endurance at generation 18, using methods described previously (Friedman et al. 1992; Dohm et al. 1994, 1996), we found no difference between HR and C mice of either sex. A study at generation 49 found HR mice to have higher endurance than C mice during forced treadmill exercise, for both sexes (Meek et al. 2009). VO_2max is one important component of running endurance ability, but we found no statistical difference between HR and C mice in the present study at generation 22 or in an unpublished study of males from generation 21 (Supplemental Table S3). (As noted above, we have found VO_2max to be consistently higher in HR mice at later generations, supporting the relatively early evolution of motivation versus performance abilities.) In principle, some of the skeletal differences we found could affect running economy, which also may impinge on endurance capacity, but studies at other generations have not found evidence for linetype effects on the cost of transport when body mass in used as a covariate in analyses (Koteja et al. 1999; Rezende et al. 2006, 2009).

Neither MPBF nor BEF specifically addresses the expectation for correlated evolution of body size. Body size is correlated with many traits, especially among species (Calder 1984; Schmidt-Nielsen 1984; Stearns 1992; Brown and West 2000; Taylor and Thomas 2014; Garland and Albuquerque 2017; Cloyed et al. 2021). In the present study, body mass was correlated with such traits as bone dimensions, organ masses, metabolic rate, and litter size, but we always used body mass as a covariate for our analyses, so any divergence in those traits would not simply reflect body size. As noted above, the reduced growth rate of HR mice may be a byproduct (likely nonadaptive) of their increased circulating corticosterone concentrations, which typically suppress growth in mammals. A negative correlated response in body size might in some sense represent a trade-off (Garland et al. 2022), given that endurance capacity is generally expected to be positively correlated with body size (e.g., see Garland and Else 1987; Garland 1994; Garland and Albuquerque 2017; Cloyed et al. 2021). On the other hand, swimming endurance is not significantly related to body size in Hsd:ICR mice (Dohm et al. 1996) and an unreplicated selection experiment with laboratory rats that targeted endurance running during forced treadmill exercise found that the high-selected line evolved lower body mass as compared with a down-selected line (Koch and Britton 2001).

Finally, neither the MPBF nor the BEF model addresses if, or how rapidly, life history traits should evolve when selection acts on behavior. However, a trade-off might occur

between maternal energetic investment in such behaviors as locomotor activity versus the production of offspring and caring for them after birth. Furthermore, in the present experiment, selection for high voluntary wheel-running behavior may have led to alterations in reward-related neural pathways or endocrine functions that also potentially affect maternal behavior (Girard et al. 2002; Rhodes et al. 2005; Garland et al. 2016). Thus, one might expect such traits as fertility, litter size, or pup mass to change (Girard et al. 2002). In any case, we found no differences in these traits.

Trade-offs or the Lack Thereof

Trade-offs are expected to be an important aspect of the evolution of life history and performance traits (Garland et al. 2022). Elsewhere, we reviewed several reasons why one might expect a trade-off between high levels of physical activity and female reproductive success, including diversion of energy from reproduction to locomotion, diminished maternal care when locomotor activity is elevated (including reduced time spent in thermoregulation and/or lactation), changes in the motivation for maternal behaviors whose neuroendocrine basis overlaps with the control of physical activity, and conflicts in the effects of hormones whose circulating levels change during reproduction (Girard et al. 2002). Nevertheless, we found no evidence for trade-offs with respect to weaning success or other litter characteristics either here or in the previous study.

The reduction in growth rate between weaning and sexual maturity, and the resulting reduction in adult body mass (Swallow et al. 1999), could be viewed as a trade-off, but this did not adversely affect any of the litter traits measured. Reduced body mass could adversely affect the ability of a small endotherm to maintain body temperature when housed below the thermal neutral zone, and room temperature is indeed below the thermal neutral zone for laboratory house mice see (Hylander and Repasky 2016; Gordon 2017), but HR mice do not have lower body temperatures when at rest during the day. Moreover, body temperatures of HR mice are elevated relative to those of C mice during nightly wheel running (Rhodes et al. 2000).

Locomotor ability is also a prime candidate for the occurrence of trade-offs, based on variation in fiber types and other aspects of muscle function and biomechanics that cause trade-offs among force generation, speed of contraction, and ability to sustain force (Wilson and James 2004; Castro et al. 2022a; Mendoza et al. 2023; and references therein). For

example, muscles that are "designed" for stamina generally cannot contract rapidly. Hence, we expected a possible trade-off involving daily wheel-running distance -- an endurance activity -- and maximal sprint-running speed. Interestingly, we found no statistical differences between the HR and C lines of mice for measures of the duration of daily running, but strong divergence in average (and maximum) wheel-running speeds, as we have reported in many other studies at both earlier and later generations (e.g., Swallow et al. 1998a; Garland et al. 2011a). However, the maximum speeds observed running on wheels (e.g., 1.036 m/s for the highest minute of running for the highest of 160 individual mice) are substantially lower than those measured in the racetrack (e.g., 2.604 m/s for the fastest individual) (see also Dohm et al. 1996; Claghorn et al. 2017a). Moreover, running on wheels is highly intermittent, and has evolved to be even more intermittent in the HR lines (Girard et al. 2001). Thus, it is unclear if one should expect the evolution of increased wheel-running speed to impact sprint speeds.

As noted above, we did not measure endurance capacity in the present study, although at generation 49 HR mice had higher treadmill endurance than C mice (Meek et al. 2009). Nevertheless, given that VO_2max is one important determinant of the maximal aerobic speed and running endurance, but we found no statistical difference in VO_2max between HR and C mice in the present study or in an unpublished study of males from generation 21 (Supplemental Table S3), our results would not necessarily lead to an expectation of reduced sprint-running ability.

Both rota-rod performance and wheel running involve coordination, so one might expect a positive correlation. On the other hand, the rota-rod test involves a rod of ~3 cm diameter, much smaller than that for the wheels used here, and the trials lasted a maximum of 10 minutes, which is much shorter than the duration of daily running but much longer than the duration of a continuous wheel-running bout (Girard et al. 2001). Thus, whether one would expect this measure of performance to trade-off with wheel running or evolve in the same direction is unclear. In any case, rota-rod performance tended to be higher in HR than C mice in the combined-sex analysis ($P = 0.0580$), although not when body mass was included as a covariate ($P = 0.6004$). A lack of improved motor coordination would be consistent with the idea that motivation, but not ability, had evolved at the time the HR lines were reaching selection limits.

Although we found no general differences in performance measures between the HR and C mice for either sex, the subset of individuals with the mini-muscle phenotype had lower maximal sprint speeds (P = 0.0026 in the combined-sex analyses) (see also Dlugosz et al. 2009), but did not differ in rota-rod performance (P = 0.6658) or VO_2max (P = 0.7279), nor for any of the four measures of wheel running (all P $\geq$ 0.16). Some studies at later generations have reported that mini-muscle individuals run more distance, at higher speeds, and/or for shorter durations as compared with normal-muscle mice (e.g., see Hannon et al. 2008; Dlugosz et al. 2009; Gomes et al. 2009), and have an increased VO_2max (Hiramatsu et al. 2017).

Bone Traits

Eleven bone traits (22% of those measured) appear as diverged between the HR and Control lines with a nominal P $\leq$ 0.05 (Supplemental Table S3). As discussed elsewhere, at least some of these bone differences may represent functional adaptations in the HR mice, such as the increased size of the femoral head the HR and C lines (e.g., Garland and Freeman 2005; Kelly et al. 2006; Middleton et al. 2008; Schutz et al. 2014; Schwartz et al. 2018; Castro et al. 2021, 2022b). On the other hand, HR mice have higher circulating corticosterone concentrations and glucocorticoids are known to affect bone properties (Gordon et al. 1993; Henneicke et al. 2011; Kinlein et al. 2017). Therefore, some of the bone differences between HR and C mice could be non-adaptive or even maladaptive byproducts of increased corticosterone levels in the selected lines (e.g., thicker femurs).

Limitations, Caveats, and Future Directions

In general, the present study lacks measures of various phenotypes that would be of particular interest with respect to the MPBF and BEF models. For example, we do not have any direct measures of endurance capacity near generation 22, although we do know that endurance capacity during forced treadmill exercise had increased significantly by generation 49 (Meek et al. 2009). Hence, the conclusion that performance abilities *sensu lato* had not evolved by generation 22 should be viewed with due caution. However, we did measure VO_2max at generations 21 and 22, finding no statistical divergence between HR and Control lines of mice, and we also found no difference in maximal sprint speed or rota-rod performance at generation 22. As with most protocols that aim to measure performance

ability, such as sprint speed, it is difficult to know if subjects are maximally motivated, and so some element of "behavior" may be involved (Garland and Losos 1994; e.g., see Losos et al. 2002), thus blurring trait categories. Still, sprint speed is commonly used as a measure of locomotor performance ability in both laboratory mice (Friedman et al. 1992; Dohm et al. 1994, 1996) and wild rodents (Djawdan and Garland Jr 1988; Garland et al. 1988, 1995).

We have no direct measures of brain traits in the generations near 22, although we know that non-cerebellar brain mass (with body mass as a covariate) had significantly increased by generation 39, and that the sizes of particular brain regions the size of brain regions has also increased in later generations (e.g., midbrain, hippocampus: Kolb et al. 2013; Schmill et al. 2023). Abundant evidence for alterations in the brain reward system or in traits that affect the reward system has accumulated in later generations (Johnson et al. 2003; Bronikowski et al. 2004; Rhodes et al. 2005; Belke and Garland 2007; Keeney et al. 2008, 2012; Garland et al. 2016; Acosta et al. 2017; Thompson et al. 2017). Included in the present range of generations are pharmacological studies showing altered responses of wheel running to drugs that would affect the dopaminergic neuromodulatory system (Rhodes et al. 2001). These pharmacological studies, combined with the lack of divergence in performance abilities, are consistent with the BEF model.

The placement of traits into categories for analysis (cf. Bartholomew 1964) is somewhat subjective and, in some cases, reasonable people might disagree. For example, we categorized the responses of wheel running to drugs as whole-organism physiology, with the rationale that they reflect neurophysiological changes, even though the measured "output" is clearly a behavior. As another example, the length of the estrus cycle is also considered as whole-organism physiology, but one might argue for it being a life history trait. As we are making available all of the P values and the categorizations used for analyses (Supplemental Table S3), readers may recategorize and reanalyze as they see fit, but we believe that the overall conclusions discussed above are robust to such potential changes.

A final caveat relates to the time of year when these mice were sampled. Wheel running shows a strong seasonal pattern in these mice, especially in the HR lines, with greater running during the winter and lower levels during the summer (see Appendix S5 in Careau et al. 2013). Wheel running in the present sample from generation 22 was measured beginning 4 June 1999, and values were substantially higher both two generations before and two after, which were measured during the winter (Careau et al.

2013). We do not know about possible seasonal changes in other traits, but if they are of greater magnitude in the HR lines, then the apparent differences between HR and C mice may also vary seasonally.

In closing, we note that neither the MPBF nor the BEF model addresses how hierarchical evolution might proceed beyond a selection limit. Indeed, little theoretical or empirical work addresses this subject at either the phenotypic or genetic level, although we have offered a verbal model that relates to the evolution of multiple solutions, such as differences among the replicate HR lines (Hillis and Garland 2023). Further insight regarding this topic will be crucial to future studies of adaptation, because, as the High Runner mouse selection experiment amply demonstrates, adaptation is a moving target (e.g., see Rose et al. 2005; Edgell et al. 2009; Crisci et al. 2016; Castro et al. 2021; Scott and Dalziel 2021; Hillis 2022).

Acknowledgments

We thank numerous other members of the Garland lab and R. B. Huey for helpful discussions over many years. Two anonymous reviewers provided helpful comments on the manuscript. This work was supported by US NSF grants to T.G., most recently IOS-2038528.

Data and Code Accessibility Statement

Data and code are available from the corresponding author on reasonable request.

Figure Legends

Figure 1. Path diagram of the “morphology - performance - behavior - fitness" (MPBF) paradigm, modified from Storz et al. (2015) to illustrate some of the traits included in the present study (mainly those from the new focal data set at generation 22). Single-headed arrows imply causality, whereas double-headed arrows indicate correlations. Various lower-level subordinate traits act together to determine whole-organism performance abilities, such as maximal sprint speed. Organismal performance refers to what an animal can do when maximally motivated (Garland and Losos 1994; Careau and Garland 2012). These abilities constrain the expression of behavior, thus playing a permissive role. Not shown on this diagram is motivation (or "will" or "drive"), which sets the propensity to express a behavior, within the confines of the "performance space" (Bennett 1989). At least one depiction of the MPBF paradigm has included motivation as a factor directly affecting behavior (Higham et al. 2011), but here it is implicit as we did not have direct measures of motivation for wheel running. In the High Runner mouse experiment, voluntary wheel-running behavior is directly tied to mating opportunity for both males and females due to the way artificial selection is imposed (see Methods). If a female becomes pregnant and gives birth, then the components of fitness shown here may be greater than zero, and all of them may be correlated with female body size. As indicated in the Introduction, the "behavior evolves first" (BEF) hypothesis is a more general (and older) idea about hierarchical evolution that existed before the modern focus on measurement of whole-organism performance traits (e.g., maximal sprint speed).

Figure 2. Adult voluntary wheel-running behavior on days 5+6. A) Mean wheel revolutions per day (circumference 1.12 m), B) duration of daily wheel running, C) mean revolutions per minute, D) maximum revolutions per minute. Values are least squares means and standard errors from combined-sex analyses in SAS Procedure Mixed (*P*-values are from Differences of Least Squares Means). See text and Supplemental Table S1 for statistical results (analyses were done on transformed values for revolutions, but untransformed data are shown here). Mice from the High Runner lines evolved mainly by increasing running speed, not the duration of daily running.

Figure 3. Body mass and growth rate. A) Average daily growth rate measured by subtracting weaning mass from body mass measured at the start of the 6-day wheel test, then dividing by the number of days involved (age as a covariate). B) Adult body mass at the start of wheel testing. C) Mass at the end of the 6-day wheel test. D) Change in body mass across the 6-day wheel test. Values are least squares means and standard errors from combined sex-analyses in SAS Procedure Mixed. See Supplemental Table S1 for statistical results. Mice from the HR lines have evolved reduced adult body mass, caused by reduced post-weaning growth rates (body mass at weaning does not significantly differ).

Figure 4. *P* values in relation to eight categories of trait (for this analysis, "Bones" and "Morphology, Physiology" were included in the category of Subordinate Traits). Two-way analysis of variance (see text) indicates a highly significant effect of trait category but no effect of sex, nor a sex * trait type interaction. A key result is that none of the available measures of Performance (e.g., VO_2max) had diverged significantly between the selectively bred HR lines and the non-selected Control lines of mice. Interestingly, measures of adult body size (mainly mass) have *P* values as low as for wheel running itself. Numbers after wheel traits indicate the days on which the trait was measured. Artificial selection was based on wheel revolutions run on days 5 and 6 of a 6-day exposure period when mice are young adults. The notable outlier for running on days 5-6 is for female minutes/day (duration of daily running). As noted in the text and elsewhere (e.g., Garland et al. 2011a), females in the HR lines have evolved longer daily running distances mainly by increased running speed, whereas HR males have increased primarily in speed but also in the duration of daily running.

Figure 5. Trait categories as shown in Figure 1. Left vertical axis for bars indicates percentage of P-values that were below 0.05 (none for Organismal Performance), whereas right axis for black line indicates mean P-values by trait category (highest for Organismal Performance). Values in parentheses are number of traits in each category. For purposes of this summarization, as compared with Figure 4, Lower-level Subordinate traits includes Bones, Morphology, Physiology, and Whole-organism Physiology. Behavior includes all measures of wheel running. Components of Fitness includes Life History traits and all

measures of Body Size. The apparent lack of response to selection for Performance traits (e.g., maximal oxygen consumption) is inconsistent with the "morphology - performance - behavior - fitness" (MPBF) paradigm, although not with the less specific "behavior evolves first" (BEF) hypothesis.

Figure 6. Four of many possible scenarios for how motivation and ability may have evolved in relation to artificial selection for high voluntary wheel-running behavior. In the present study, measurements were taken approximately when the plateau for wheel running was being reached. A) In the base population before selection began, most mice ran a daily distance that matched both their motivation *for* running and their ability *to* run. In this case, the response to positive selection would have entailed simultaneous increases in both motivation and ability, and a selection limit (plateau) would be reached when neither could increase further for genetic and/or functional reasons (two different ways to view the cause of selection limits). B) In the base population, the ability of mice to run greatly exceeded their motivation (consistent with common human experience!), so increased wheel running occurred by parallel increases in motivation for running (or the reward received from running), with a selection limit occurring when motivation matched ability and neither could evolve further. The data presented in the present study matches this scenario better than the others. However, this simple model would not account for the fact that increases in performance ability have been documented at later generations. C) The converse of B: motivation for running initially greatly exceeded ability, which had to evolve for daily running distance to respond to selection. D) Ability exceeded motivation for running in the base population, and after some number of generations motivation matched ability, at which point both increased up to the selection limit.

Literature Cited

Acosta W., T.H. Meek, H. Schutz, E.M. Dlugosz, and T. Garland, Jr. 2017. Preference for Western diet coadapts in High Runner mice and affects voluntary exercise and spontaneous physical activity in a genotype-dependent manner. Behav Processes 135:56–65.

Arnold S.J. 1983. Morphology, performance and fitness. Am Zool 23:347–361.

Bartholomew G.A. 1964. The roles of physiology and behaviour in the maintenance of homeostasis in the desert environment. Pp. 7–29 in Homeost Feedback Mech, Symposia of the Society for Experimental Biology. Academic Press, New York, NY.

Bateson P. 1988. The active role of behaviour in evolution. Pp. 191–207 in M.-W. Ho and S. Fox eds. Evol Process Metaphors. Wiley, Chichester, U.K.

Belke T.W. and T. Garland, Jr. 2007. A brief opportunity to run does not function as a reinforcer for mice selected for high daily wheel-running rates. J Exp Anal Behav 88:199–213.

Bennett A.F. 1989. Integrated studies of locomotor performance. Pp. 191–202 in D.B. Wake and G. Roth eds. Complex Org Funct Integr Evol Vertebr. John Wiley & Sons, Ltd.

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

Bolles R.C. 1967. Theory of motivation (2nd ed.). Harper and Row, New York.

_____. 1975. Theory of motivation. 2nd ed. (2nd ed.). Harper and Row, New York.

Brodie III E.D. 1992. Correlational selection for color pattern and antipredator behavior in the garter snake *Thamnophis ordinoides*. Evolution 46:1284–1298.

Bronikowski A.M., P.A. Carter, J.G. Swallow, I.A. Girard, J.S. Rhodes, and T. Garland, Jr. 2001. Open-field behavior of house mice selectively bred for high voluntary wheel-running. Behav Genet 31:309–316.

Bronikowski A.M., J.S. Rhodes, T. Garland, Jr., T.A. Prolla, T.A. Awad, and S.C. Gammie. 2004. The evolution of gene expression in mouse hippocampus in response to selective breeding for increased locomotor activity. Evolution 58:2079–2086.

Brown J.H. and G.B. West, eds. 2000. Scaling in biology. Oxford University Press, New York.

Cadney M.D., L. Hiramatsu, Z. Thompson, M. Zhao, J.C. Kay, J.M. Singleton, R.L. Albuquerque, et al. 2021. Effects of early-life exposure to Western diet and voluntary exercise on adult activity levels, exercise physiology, and associated traits in selectively bred High Runner mice. Physiol Behav 234:113389.

Calder W.A. 1984. Size, function, and life history. Harvard University Press, Cambridge, Mass.

Careau V., O.R.P. Bininda-Emonds, G. Ordonez, and T. Garland, Jr. 2012. Are voluntary wheel running and open-field behavior correlated in mice? Different answers from comparative and artificial selection approaches. Behav Genet 42:830–844.

Careau V. and T. Garland, Jr. 2012. Performance, personality, and energetics: correlation, causation, and mechanism. Physiol Biochem Zool 85:543–571.

Careau V., M.E. Wolak, P.A. Carter, and T. Garland, Jr. 2013. Limits to behavioral evolution: the quantitative genetics of a complex trait under directional selection. Evolution 67:3102–3119.

_____. 2015. Evolution of the additive genetic variance–covariance matrix under continuous directional selection on a complex behavioural phenotype. Proc R Soc B Biol Sci 282:20151119.

Carter P.A., J. Garland Theodore, M.R. Dohm, and J.P. Hayes. 1999. Genetic variation and correlations between genotype and locomotor physiology in outbred laboratory house mice (*Mus domesticus*). Comp Biochem Physiol A Mol Integr Physiol 123:155–162.

Castro A.A. and T. Garland, Jr. 2018. Evolution of hindlimb bone dimensions and muscle masses in house mice selectively bred for high voluntary wheel-running behavior. J Morphol 279:766–779.

Castro A.A., T. Garland, Jr., S. Ahmed, and N.C. Holt. 2022a. Trade-offs in muscle physiology in selectively bred High Runner mice. J Exp Biol 225:jeb244083.

Castro A.A., F.A. Karakostis, L.E. Copes, H.E. McClednon, A.P. Trivedi, N.E. Schwartz, and T. Garland, Jr. 2022b. Effects of selective breeding for voluntary exercise, chronic exercise, and their interaction on muscle attachment site morphology in house mice. J Anat 240:279–295.

Castro A.A., H. Rabitoy, G.C. Claghorn, and T. Garland, Jr. 2021. Rapid and longer-term effects of selective breeding for voluntary exercise behavior on skeletal morphology in house mice. J Anat 238:720–742.

Claghorn G.C., I.A.T. Fonseca, Z. Thompson, C. Barber, and T. Garland, Jr. 2016. Serotonin-mediated central fatigue underlies increased endurance capacity in mice from lines selectively bred for high voluntary wheel running. Physiol Behav 161:145–154.

Claghorn G.C., Z. Thompson, J.C. Kay, G. Ordonez, T.G. Hampton, and T. Garland, Jr. 2017a. Selective breeding and short-term access to a running wheel alter stride characteristics in house mice. Physiol Biochem Zool 90:533–545.

Claghorn G.C., Z. Thompson, K. Wi, L. Van, and T. Garland, Jr. 2017b. Caffeine stimulates voluntary wheel running in mice without increasing aerobic capacity. Physiol Behav 170:133–140.

Cloyed C.S., J. Grady, V. Savage, J.C. Uyeda, and A.I. Dell. 2021. The allometry of locomotion. Ecology 102:e03369.

Cohen A.A., C. Coste, X.-Y. Li, S. Bourg, and S. Pavard. 2020. Are trade-offs really the key drivers of aging and lifespan? Funct Ecol 34:153–166.

Copes L.E., H. Schutz, E.M. Dlugosz, W. Acosta, M.A. Chappell, and T. Garland, Jr. 2015. Effects of voluntary exercise on spontaneous physical activity and food consumption in mice: results from an artificial selection experiment. Physiol Behav 149:86–94.

Crisci J.L., M.D. Dean, and P. Ralph. 2016. Adaptation in isolated populations: when does it happen and when can we tell? Mol Ecol 25:3901–3911.

Crusio W.E. 1995. Natural selection on hippocampal circuitry underlying exploratory behavior in mice: quantitative-genetic analysis. Pp. 323–342 in Behav Brain Res Nat Semin Setting NATO Adv Study Inst Ser Behav Soc Sci. Kluwer Academic Press, Dordrecht, The Netherlands.

Dantzer B., A.E. Newman, R. Boonstra, R. Palme, S. Boutin, M.M. Humphries, and A.G. McAdam. 2013. Density triggers maternal hormones that increase adaptive offspring growth in a wild mammal. Science 340:1215–1217.

Dantzer B., S.E. Westrick, and F. van Kesteren. 2016. Relationships between endocrine traits and life histories in wild animals: insights, problems, and potential pitfalls. Integr Comp Biol 56:185–197.

Darwin C.R. 1859. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life [1st edition] (1st ed.). John Murray, London.

Djawdan M. and T. Garland Jr. 1988. Maximal running speeds of bipedal and quadrupedal rodents. J Mammal 69:765–772.

Dlugosz E.M., M.A. Chappell, D.G. McGillivray, D.A. Syme, and T. Garland, Jr. 2009. Locomotor trade-offs in mice selectively bred for high voluntary wheel running. J Exp Biol 212:2612–2618.

Dlugosz E.M., M.A. Chappell, T.H. Meek, P. Szafrańska, K. Zub, M. Konarzewski, J.H. Jones, et al. 2013. Phylogenetic analysis of mammalian maximal oxygen consumption during exercise. J Exp Biol jeb.088914.

Dohm M.R., J.P. Hayes, and T. Garland, Jr. 1996. Quantitative genetics of sprint running speed and swimming endurance in laboratory house mice (*Mus domesticus*). Evolution 50:1688–1701.

Dohm M.R., J.P. Hayes, and T. Garland, Jr. 2001. The quantitative genetics of maximal and basal rates of oxygen consumption in mice. Genetics 159:267–277.

Dohm M.R., C.S. Richardson, and T. Garland. 1994. Exercise physiology of wild and random-bred laboratory house mice and their reciprocal hybrids. Am J Physiol-Regul Integr Comp Physiol 267:R1098–R1108.

Douhard F., M. Douhard, H. Gilbert, P. Monget, J.-M. Gaillard, and J.-F. Lemaître. 2021. How much energetic trade-offs limit selection? Insights from livestock and related laboratory model species. Evol Appl 14:2726–2749.

Duckworth R.A. 2009. The role of behavior in evolution: a search for mechanism. Evol Ecol 23:513–531.

Dunham N.W. and T.S. Miya. 1957. A note on a simple apparatus for detecting neurological deficit in rats and mice. J Am Pharm Assoc Am Pharm Assoc 46:208–209.

Edgell T.C., B.R. Lynch, G.C. Trussell, and A.R. Palmer. 2009. Experimental evidence for the rapid evolution of behavioral canalization in natural populations. Am Nat 174:434–440.

Fischer E.K. and L.A. O'Connell. 2017. Modification of feeding circuits in the evolution of social behavior. J Exp Biol 220:92–102.

Fornaguera J. and R.K.W. Schwarting. 2002. Time course of deficits in open field behavior after unilateral neostriatal 6-hydroxydopamine lesions. Neurotox Res 4:41–49.

Friedman W.A., T.J. Garland, and M.R. Dohm. 1992. Individual variation in locomotor behavior and maximal oxygen consumption in mice. Physiol Behav 52:97–104.

Gans C. 1979. Momentarily excessive construction as the basis for protoadaptation. Evolution 33:227–233.

Garland, Jr. T. 2003. Selection experiments: an under-utilized tool in biomechanics and organismal biology. Pp. 23–56 in V. L. Bels, J.-P. Gasc, and A. Casinos eds. Vertebr Biomech Evol. BIOS Scientific Publishers Ltd., Oxford, U.K.

Garland, Jr. T. and R.L. Albuquerque. 2017. Locomotion, energetics, performance, and behavior: a mammalian perspective on lizards, and vice versa. Integr Comp Biol 57:252–266.

Garland, Jr. T., A.F. Bennett, and C.B. Daniels. 1990. Heritability of locomotor performance and its correlates in a natural population. Experientia 46:530–533.

Garland, Jr. T. and P.A. Carter. 1994. Evolutionary physiology. Annu Rev Physiol 56:579–621.

Garland, Jr. T., C.J. Downs, and A.R. Ives. 2022. Trade-offs (and constraints) in organismal biology. Physiol Biochem Zool 95:82–112.

Garland, Jr. T. and P.L. Else. 1987. Seasonal, sexual, and individual variation in endurance and activity metabolism in lizards. Am J Physiol-Regul Integr Comp Physiol 252:R439–R449.

Garland, Jr. T. and P.W. Freeman. 2005. Selective breeding for high endurance running increases hindlimb symmetry. Evolution 59:1851–1854.

Garland, Jr. T., F. Geiser, and R.V. Baudinette. 1988. Comparative locomotor performance of marsupial and placental mammals. J Zool Lond 215:505–522.

Garland, Jr. T., T.T. Gleeson, B.A. Aronovitz, C.S. Richardson, and M.R. Dorm. 1995. Maximal sprint speeds and muscle fiber composition of wild and laboratory house mice. Physiol Behav 58:869–876.

Garland, Jr. T. Jr. 1994. Phylogenetic analyses of lizard endurance capacity in relation to body size and body temperature. Pp. 237–259 in L.J. Vitt and E.R. Pianka eds. Lizard Ecol Hist Exp Perspect. Princeton University Press, Princeton, N.J.

Garland, Jr. T., S.A. Kelly, J.L. Malisch, E.M. Kolb, R.M. Hannon, B.K. Keeney, S.L. Van Cleave, et al. 2011a. How to run far: multiple solutions and sex-specific responses to selective breeding for high voluntary activity levels. Proc R Soc B Biol Sci 278:574–581.

Garland, Jr. T. and J.B. Losos. 1994. Ecological morphology of locomotor performance in squamate reptiles. Pp. 240–302 in P.C. Wainwright and S.M. Reilly eds. Ecol Morphol Integr Org Biol. University of Chicago Press, Chicago.

Garland, Jr. T., M.T. Morgan, J.G. Swallow, J.S. Rhodes, I. Girard, J.G. Belter, and P.A. Carter. 2002. Evolution of a small-muscle polymorphism in lines of house mice selected for high activity levels. Evolution 56:1267–1275.

Garland, Jr. T., H. Schutz, M.A. Chappell, B.K. Keeney, T.H. Meek, L.E. Copes, W. Acosta, et al. 2011b. The biological control of voluntary exercise, spontaneous physical activity and daily energy expenditure in relation to obesity: human and rodent perspectives. J Exp Biol 214:206–229.

Garland, Jr. T., M. Zhao, and W. Saltzman. 2016. Hormones and the evolution of complex traits: insights from artificial selection on behavior. Integr Comp Biol 56:207–224.

Girard I. and T. Garland, Jr. 2002. Plasma corticosterone response to acute and chronic voluntary exercise in female house mice. J Appl Physiol 92:1553–1561.

Girard I., M.W. McAleer, J.S. Rhodes, and T. Garland, Jr. 2001. Selection for high voluntary wheel-running increases speed and intermittency in house mice (*Mus domesticus*). J Exp Biol 204:4311–4320.

Girard I., E.L. Rezende, and T. Garland Jr. 2007. Leptin levels and body composition of mice selectively bred for high voluntary locomotor activity. Physiol Biochem Zool 80:568–579.

Girard I., J.G. Swallow, P.A. Carter, P. Koteja, J.S. Rhodes, and T. Garland, Jr. 2002. Maternal-care behavior and life-history traits in house mice (*Mus domesticus*) artificially selected for high voluntary wheel-running activity. Behav Processes 57:37–50.

Glazier D.S. 2009. Trade-offs. Pp. 44–60 in W.M. Rauw ed. Resour Alloc Theory Appl Farm Anim Prod. CAB International, Cambridge, Mass.

Gomes F.R., E.L. Rezende, J.L. Malisch, S.K. Lee, D.A. Rivas, S.A. Kelly, C. Lytle, et al. 2009. Glycogen storage and muscle glucose transporters (GLUT-4) of mice selectively bred for high voluntary wheel running. J Exp Biol 212:238–248.

Gordon C.J. 2017. The mouse thermoregulatory system: Its impact on translating biomedical data to humans. Physiol Behav 179:55–66.

Gordon, Kenneth R., Levy, Cesar, Perl, Mordechai, and Weeks, Ophelia I. 1993. Adaptive Modeling In A Mammalian Skeletal Model System. Growth Dev Aging 57:101–110.

Green P.A., M.J. McHenry, and A. Rico-Guevara. 2021. Mechanoethology: the physical mechanisms of behavior. Integr Comp Biol 61:613–623.

Hannon R.M., S.A. Kelly, K.M. Middleton, E.M. Kolb, D. Pomp, and T. Garland, Jr. 2008. Phenotypic effects of the “Mini-Muscle” allele in a large HR x C57BL/6J mouse backcross. J Hered 99:349–354.

Hayes J.P., T. Garland, and M.R. Dohm. 1992. Individual variation in metabolism and reproduction of mus: are energetics and life history linked? Funct Ecol 6:5.

Henneicke H., M. Herrmann, R. Kalak, T.C. Brennan-Speranza, U. Heinevetter, N. Bertollo, R.E. Day, et al. 2011. Corticosterone selectively targets endo-cortical surfaces by an osteoblast-dependent mechanism. Bone 49:733–742.

Higham T.E., P.G. Korchari, and L.D. McBrayer. 2011. How muscles define maximum running performance in lizards: an analysis using swing- and stance-phase muscles. J Exp Biol 214:1685–1691.

Hillis D.A. 2022. *Genetic basis of aerobically supported voluntary exercise: results from a selection experiment with house mice*. University of California, Riverside, Riverside, California.

Hillis D.A. and T. Garland, Jr. 2023. Multiple solutions at the genomic level in response to selective breeding for high locomotor activity. Genetics 223:iyac165.

Hiramatsu L., J.C. Kay, Z. Thompson, J. Singleton, G.C. Claghorn, R.L. Albuquerque, B. Ho, et al. 2017. Maternal exposure to Western diet affects adult body composition and voluntary wheel running in a genotype-specific manner in mice. Physiol Behav 179:235–245.

Horning M. 2012. Constraint lines and performance envelopes in behavioral physiology: the case of the aerobic dive limit. Front Physiol 3.

Houle-Leroy P., H. Guderley, J.G. Swallow, and T. Garland, Jr. 2003. Artificial selection for high activity favors mighty mini-muscles in house mice. Am J Physiol-Regul Integr Comp Physiol 284:R433–R443.

Huey R.B., P.E. Hertz, and B. Sinervo. 2003. Behavioral drive versus behavioral inertia in evolution: a null model approach. Am Nat 161:357–366.

Hulbert A.J. and P.L. Else. 2004. Basal metabolic rate: history, composition, regulation, and usefulness. Physiol Biochem Zool 77:869–876.

Husak J.F., D.J. Irschick, S.D. McCormick, and I.T. Moore. 2009. Hormonal regulation of whole-animal performance: Implications for selection. Integr Comp Biol 49:349–353.

Hylander B.L. and E.A. Repasky. 2016. Thermoneutrality, Mice, and Cancer: A Heated Opinion. Trends Cancer 2:166–175.

Johnson R.A., J.S. Rhodes, S.L. Jeffrey, T. Garland, Jr., and G.S. Mitchell. 2003. Hippocampal brain-derived neurotrophic factor but not neurotrophin-3 increases more in mice selected for increased voluntary wheel running. Neuroscience 121:1–7.

Jonas I., K.A. Schubert, A.C. Reijne, J. Scholte, T. Garland, Jr., M.P. Gerkema, A.J.W. Scheurink, et al. 2010. Behavioral traits are affected by selective breeding for increased wheel-running behavior in mice. Behav Genet 40:542–550.

Jones B.J. and D.J. Roberts. 1968. The quantiative measurement of motor inco-ordination in naive mice using an acelerating rotarod. J Pharm Pharmacol 20:302–304.

Kalivas P.W. and M. Nakamura. 1999. Neural systems for behavioral activation and reward. Curr Opin Neurobiol 9:223–227.

Kalueff A.V., C.L. Jensen, and D.L. Murphy. 2007. Locomotory patterns, spatiotemporal organization of exploration and spatial memory in serotonin transporter knockout mice. Brain Res 1169:87–97.

Katz P.S. 2011. Neural mechanisms underlying the evolvability of behaviour. Philos Trans R Soc B Biol Sci 366:2086–2099.

Keeney B.K., T.H. Meek, K.M. Middleton, L.F. Holness, and T. Garland, Jr. 2012. Sex differences in cannabinoid receptor-1 (CB1) pharmacology in mice selectively bred for high voluntary wheel-running behavior. Pharmacol Biochem Behav 101:528–537.

Keeney B.K., D.A. Raichlen, T.H. Meek, R.S. Wijeratne, K.M. Middleton, G.L. Gerdeman, and T. Garland, Jr. 2008. Differential response to a selective cannabinoid receptor antagonist (SR141716: rimonabant) in female mice from lines selectively bred for high voluntary wheel-running behaviour. Behav Pharmacol 19:812–820.

Kelley A.E. and K.C. Berridge. 2002. The neuroscience of natural rewards: relevance to addictive drugs. J Neurosci 22:3306–3311.

Kelly S.A., T.A. Bell, S.R. Selitsky, R.J. Buus, K. Hua, G.M. Weinstock, T. Garland, Jr., et al. 2013. A novel intronic single nucleotide polymorphism in the *Myosin heavy polypeptide 4* gene is responsible for the Mini-Muscle phenotype characterized by major reduction in hind-limb muscle mass in mice. Genetics 195:1385–1395.

Kelly S.A., P.P. Czech, J.T. Wight, K.M. Blank, and T. Garland, Jr. 2006. Experimental evolution and phenotypic plasticity of hindlimb bones in high-activity house mice. J Morphol 267:360–374.

Kinlein S.A., Z. Shahanoor, R.D. Romeo, and I.N. Karatsoreos. 2017. Chronic corticosterone treatment during adolescence has significant effects on metabolism and skeletal development in male C57BL6/N mice. Endocrinology 158:2239–2254.

Kinnard W.J.J. and C.J. Carr. 1957. A preliminary procedure for the evaluation of central nervous system depressants. J Pharmacol Exp Ther 121:354–361.

Klimentidis Y.C., M. Newell, M.D. van der Zee, V.L. Bland, S. May-Wilson, G. Arani, C. Menni, et al. 2022. Genome-wide association study of liking for several types of physical activity in the UK Biobank and two replication cohorts. Med Sci Sports Exerc 54:1252–1260.

Koch L.G. and S.L. Britton. 2001. Artificial selection for intrinsic aerobic endurance running capacity in rats. Physiol Genomics 5:45–52.

Kolb E.M., S.A. Kelly, and T. Garland, Jr. 2013. Mice from lines selectively bred for high voluntary wheel running exhibit lower blood pressure during withdrawal from wheel access. Physiol Behav 112–113:49–55.

Kolb E.M., S.A. Kelly, K.M. Middleton, L.S. Sermsakdi, M.A. Chappell, and T. Garland, Jr. 2010. Erythropoietin elevates $VO_{2,max}$ but not voluntary wheel running in mice. J Exp Biol 213:510–519.

Koteja P., J.G. Swallow, P.A. Carter, and T. Garland, Jr. 1999. Energy cost of wheel running in house mice: implications for coadaptation of locomotion and energy budgets. Physiol Biochem Zool 72:238–249.

Koteja P., J.G. Swallow, P.A. Carter, and T. Garland. 2001. Maximum cold-induced food consumption in mice selected for high locomotor activity: implications for the evolution of endotherm energy budgets. J Exp Biol 204:1177–1190.

Kuhn S.L., D.A. Raichlen, and A.E. Clark. 2016. What moves us? How mobility and movement are at the center of human evolution. Evol Anthropol Issues News Rev 25:86–97.

Lailvaux S.P. and J.F. Husak. 2014. The life history of whole-organism performance. Q Rev Biol 89:285–318.

Lambert M.I., C. Van Zyl, R. Jaunky, E.V. Lambert, and T.D. Noakes. 1996. Tests of running performance do not predict subsequent spontaneous running in rats. Physiol Behav 60:171–176.

Lande R. and S.J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 1210–1226.

Lightfoot J.T., E.J.C. De Geus, F.W. Booth, M.S. Bray, M. den Hoed, J.A. Kaprio, S.A. Kelly, et al. 2018. Biological/genetic regulation of physical activity level: consensus from GenBioPAC. Med Sci Sports Exerc 50:863–873.

Losos J.B., D.A. Creer, and J.A. Schulte II. 2002. Cautionary comments on the measurement of maximum locomotor capabilities. J Zool 258:57–61.

Malisch J.L., C.W. Breuner, E.M. Kolb, H. Wada, R.M. Hannon, M.A. Chappell, K.M. Middleton, et al. 2009. Behavioral despair and home-cage activity in mice with chronically elevated baseline corticosterone concentrations. Behav Genet 39:192–201.

Malisch J.L., W. Saltzman, F.R. Gomes, E.L. Rezende, D.R. Jeske, and T. Garland, Jr. 2007. Baseline and stress-induced plasma corticosterone concentrations of mice selectively bred for high voluntary wheel running. Physiol Biochem Zool 80:146–156.

Mayr E. 1958. Behavior and systematics. Pp. 341–363 in A. Roe and G.G. Simpson eds. Behav Evol. Yale University Press, New Haven, CT.

_____. 1982. The growth of biological thought. Belknap Press of Harvard University Press, Cambridge, Mass., and London.

McNab B.K. 2012. Extreme measures: the ecological energetics of birds and mammals. University of Chicago Press, Chicago.

Meek T.H., B.P. Lonquich, R.M. Hannon, and T. Garland, Jr. 2009. Endurance capacity of mice selectively bred for high voluntary wheel running. J Exp Biol 212:2908–2917.

Mendoza E., D.S. Moen, and N.C. Holt. 2023. The importance of comparative physiology: mechanisms, diversity and adaptation in skeletal muscle physiology and mechanics. J Exp Biol 226:jeb245158.

Middleton K.M., S.A. Kelly, and T. Garland, Jr. 2008. Selective breeding as a tool to probe skeletal response to high voluntary locomotor activity in mice. Integr Comp Biol 48:394–410.

Mundorf A., H. Matsui, S. Ocklenburg, and N. Freund. 2020. Asymmetry of turning behavior in rats is modulated by early life stress. Behav Brain Res 393:112807.

_____. 2022. Analyzing Turning Behavior after Repeated Lithium, Ketamine, or NaCl Injection and Chronic Stress Exposure in Mice. Symmetry 14.

Noakes T.D. 2004. Logical limitations to the “catastrophe” models of fatigue during exercise in humans. Br J Sports Med 38:648–649.

Noakes T.D. 2012. Fatigue is a brain-derived emotion that regulates the exercise behavior to ensure the protection of whole body homeostasis. Front Physiol 3:13 pages.

Noakes T.D. and A.S.C. Gibson. 2004. Logical limitations to the ‘“catastrophe”’ models of fatigue during exercise in humans. Br J Sports Med 38:648-649(shorter version; longer is 30 pages).

Norton S. 1982. Methods in behavioral toxicology. Pages 353-373 in A. W. Hayes, ed. Principles and methods of toxicology. P. 750 in . Raven Press, New York.

Orr T.J. and T. Garland, Jr. 2017. Complex reproductive traits and whole-organism performance. Integr Comp Biol 57:407–422.

Plomin R. 1990. The role of inheritance in behavior. Science 248:183–188.

Revell L., L. Harmon, and D. Collar. 2008. Phylogenetic signal, evolutionary process, and rate. Syst Biol 57:591–601.

Rezende E.L., F.R. Gomes, M.A. Chappell, and T. Garland, Jr. 2009. Running behavior and its energy cost in mice selectively bred for high voluntary locomotor activity. Physiol Biochem Zool 82:662–679.

Rezende E.L., S.A. Kelly, F.R. Gomes, M.A. Chappell, and T. Garland Jr. 2006. Effects of size, sex, and voluntary running speeds on costs of locomotion in lines of laboratory mice selectively bred for high wheel-running activity. Physiol Biochem Zool 79:83–99.

Rhodes J.H. and T. Garland, Jr. 2003. Differential sensitivity to acute administration of Ritalin, apomorphine, SCH 23390, and raclopride in mice selectively bred for hyperactive wheel-running behavior. Psychopharmacology (Berl) 167:242–250.

Rhodes J.S., S.C. Gammie, and T. Garland, Jr. 2005. Neurobiology of mice selected for high voluntary wheel-running activity. Integr Comp Biol 45:438–455.

Rhodes J.S., G.R. Hosack, I. Girard, A.E. Kelley, G.S. Mitchell, and T. Garland, Jr. 2001. Differential sensitivity to acute administration of cocaine, GBR 12909, and fluoxetine

in mice selectively bred for hyperactive wheel-running behavior. Psychopharmacology (Berl) 158:120–131.

Rhodes J.S. and T.J. Kawecki. 2009. Behavior and neurobiology. Pp. 263–300 in T. Garland, Jr. and M.R. Rose eds. Exp Evol Concepts Methods Appl Sel Exp. University of California Press, Berkeley.

Rhodes J.S., P. Koteja, J.G. Swallow, P.A. Carter, and T. Garland. 2000. Body temperatures of house mice artificially selected for high voluntary wheel-running behavior: repeatability and effect of genetic selection. J Therm Biol 25:391–400.

Rose M.R., H.B. Passananti, A.K. Chippindale, J.P. Phelan, M. Matos, H. Teotonio, and L.D. Mueller. 2005. The effects of evolution are local: evidence from experimental evolution in *Drosophila*. Integr Comp Biol 45:486–491.

Saul M., P. Majdak, S. Perez, M. Reilly, T. Garland, Jr., and J.S. Rhodes. 2017. High motivation for exercise is associated with altered chromatin regulators of monoamine receptor gene expression in the striatum of selectively bred mice. Genes Brain Behav 16:328–341.

Schmidt-Nielsen K. 1984. Scaling: why is animal size so important? Cambridge University Press, Cambridge.

Schmill M.P., Z. Thompson, D. Lee, L. Haddadin, S. Mitra, R. Ezzat, S. Shelton, et al. 2023. Hippocampal, whole midbrain, red nucleus, and ventral tegmental area volumes are increased by selective breeding for high voluntary wheel-running behavior. Genes Brain Behav 98:245–263.

Schutz H., H.A. Jamniczky, B. Hallgrímsson, and T. Garland, Jr. 2014. Shape-shift: semicircular canal morphology responds to selective breeding for increased locomotor activity. Evolution 68:3184–3198.

Schwartz N.E., M. McNamara, J.M. Orozco, J.O. Rashid, A.P. Thai, and T. Garland, Jr. 2023. MANUSCRIPT A test of the aerobic capacity model for vertebrate energetics: selective breeding for high voluntary exercise in mice increases maximal (VO_2max), but not basal metabolic rate. J Exp Biol Accepted.

Schwartz N.E., B.A. Patel, T. Garland, Jr., and A.M. Horner. 2018. Effects of selective breeding for high voluntary wheel-running behavior on femoral nutrient canal size and abundance in house mice. J Anat 233:193–203.

Scott G.R. and A.C. Dalziel. 2021. Physiological insight into the evolution of complex phenotypes: aerobic performance and the O_2 transport pathway of vertebrates. J Exp Biol 224:jeb210849.

Seeherman H.J., C. Richard Taylor, G.M.O. Maloiy, and R.B. Armstrong. 1981. Design of the mammalian respiratory system. II. Measuring maximum aerobic capacity. Respir Physiol 44:11–23.

Sheehan M.J., C.H. Miller, C.C. Vogt, and R.A. Ligon. 2018. Behavioral evolution: can you dig it? Curr Biol 28:R19–R21.

Singleton J. and T. Garland, Jr. 2019. Influence of corticosterone on growth, home-cage activity, wheel running, and maximal oxygen consumption in replicate lines of house mice selectively bred for high voluntary wheel-running behavior. Physiol Behav 198:27–41.

Smith N.A., K.L. Koeller, J.A. Clarke, D.T. Ksepka, J.S. Mitchell, A. Nabavizadeh, R.C. Ridgley, et al. 2022. Convergent evolution in dippers (Aves, Cinclidae): The only wing-propelled diving songbirds. Anat Rec 305:1563–1591.

Sofia R.D. 1969. Comparison of Two Methods for Measuring Drug-Induced Neurotoxicity. J Pharm Sci 58:900–901.

Stearns S.C. 1992. The evolution of life histories. Oxford University Press, Oxford; New York.

Storz J.F., J.T. Bridgham, S.A. Kelly, and T. Garland, Jr. 2015. Genetic approaches in comparative and evolutionary physiology. Am J Physiol - Regul Integr Comp Physiol 309:R197–R214.

Swallow J.G., P.A. Carter, and T. Garland, Jr. 1998a. Artificial selection for increased wheel-running behavior in house mice. Behav Genet 28:227–237.

Swallow J.G., T. Garland, Jr., P.A. Carter, W.-Z. Zhan, and G.C. Sieck. 1998b. Effects of voluntary activity and genetic selection on aerobic capacity in house mice (*Mus domesticus*). J Appl Physiol 84:69–76.

Swallow J.G., J.P. Hayes, P. Koteja, and T. Garland, Jr. 2009. Selection experiments and experimental evolution of performance and physiology. Pp. 301–351 in T. Garland, Jr. and M.R. Rose eds. Exp Evol Concepts Methods Appl Sel Exp.

Swallow J.G., P. Koteja, P.A. Carter, and T. Garland, Jr. 1999. Artificial selection for increased wheel-running activity in house mice results in decreased body mass at maturity. J Exp Biol 202:2513–2520.

Taylor G. and A. Thomas. 2014. Evolutionary biomechanics: selection, phylogeny, and constraint. Oxford University Press.

Thompson Z., D. Argueta, T. Garland, Jr., and N. DiPatrizio. 2017. Circulating levels of endocannabinoids respond acutely to voluntary exercise, are altered in mice selectively bred for high voluntary wheel running, and differ between the sexes. Physiol Behav 170:141–150.

Thompson Z., E.M. Kolb, and T. Garland, Jr. 2018. High-runner mice have reduced incentive salience for a sweet-taste reward when housed with wheel access. Behav Processes 146:46–53.

Vaanholt L.M., T. Garland, S. Daan, and G.H. Visser. 2007. Wheel-running activity and energy metabolism in relation to ambient temperature in mice selected for high wheel-running activity. J Comp Physiol B 177:109–118.

Wagner G.P. and J. Zhang. 2011. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. Nat Rev Genet 12:204–213.

Wallace I.J. and T. Garland, Jr. 2016. Mobility as an emergent property of biological organization: insights from experimental evolution. Evol Anthropol 25:98–104.

Wallace I.J., S.M. Tommasini, S. Judex, T. Garland, Jr., and B. Demes. 2012. Genetic variations and physical activity as determinants of limb bone morphology: An experimental approach using a mouse model. Am J Phys Anthropol 148:24–35.

Weaver J.E. and T.S. Miya. 1961. Effects of certain ataraxic agents on mice activity. J Pharm Sci 50:910–912.

Weibel E.R. 1970. Morphometric estimation of pulmonary diffusion capacity. I. Model and method. Respir Physiol 11:54–75.

_____. 1990. Morphometry: stereological theory and practical methods. Pp. 199–252 in J. Gil ed. Models Lung Dis - Microsc Struct Methods. Marcel Dekker, Inc., New York, Basel.

Whitehead N.N., S.A. Kelly, J.S. Demes, N.E. Schwartz, and T. Garland, Jr. 2023. Locomotor play behavior evolves by random genetic drift but not as a correlated response to selective breeding for high voluntary wheel-running behavior. Behav Processes 213:104973.

Wilson R.S. and R.S. James. 2004. Constraints on muscular performance: trade–offs between power output and fatigue resistance. Proc R Soc Lond B Biol Sci 271:S222–S225.

Supplemental Files

Supplemental Methods.docx

Supplemental Table S1. Results for combined-sex analyses:
LG23_Both_Sex_23_Supplemental_Table_S1.xlsx

Supplemental Table S2. Results for separate-sex analyses:
Table_2_Sep_Sex_Analyses_version_15_UPLOAD.xlsx

Supplemental Table S3. Trait Categorizations and Compilation of *P* values from generations 20-24
Effect_Sizes_19_out_from_SPSS_for_Manuscript.xlsx