The Journal of Experimental Biology 208, 2447-2458
Published by The Company of Biologists 2005
doi:10.1242/jeb.01631

Maximal metabolic rates during voluntary exercise, forced exercise, and cold exposure in house mice selectively bred for high wheel-running

Enrico L. Rezende*, Mark A. Chappell, Fernando R. Gomes, Jessica L. Malisch and Theodore Garland, Jr

Department of Biology, University of California, Riverside, California 92521, USA

*Author for correspondence (e-mail: enrico.rezende@email.ucr.edu)

Accepted 31 March 2005

Summary

Selective breeding for high wheel-running activity has generated four lines of laboratory house mice (S lines) that run about 170% more than their control counterparts (C lines) on a daily basis, mostly because they run faster. We tested whether maximum aerobic metabolic rates ($\dot{V}_{O_2max}$) have evolved in concert with wheel-running, using 48 females from generation 35. Voluntary activity and metabolic rates were measured on days 5+6 of wheel access (mimicking conditions during selection), using wheels enclosed in metabolic chambers. Following this, $\dot{V}_{O_2max}$ was measured twice on a motorized treadmill and twice during cold-exposure in a heliox atmosphere (He-O_2). Almost all measurements, except heliox $\dot{V}_{O_2max}$, were significantly repeatable. After accounting for differences in body mass (S<C) and variation in age at testing, S and C did not differ in $\dot{V}_{O_2max}$ during forced exercise or in heliox, nor in maximal running speeds on the treadmill. However, running speeds and $\dot{V}_{O_2max}$ during voluntary exercise were significantly higher in S lines. Nevertheless, S mice never voluntarily achieved the $\dot{V}_{O_2max}$ elicited during their forced treadmill trials, suggesting that aerobic capacity *per se* is not limiting the evolution of even higher wheel-running speeds in these lines. Our results support the hypothesis that S mice have genetically higher motivation for wheel-running and they demonstrate that behavior can sometimes evolve independently of performance capacities. We also discuss the possible importance of domestication as a confounding factor to extrapolate results from this animal model to natural populations.

Key words: artificial selection, exercise, experimental evolution, locomotor activity, maximum oxygen consumption, running performance, thermogenesis, mouse.

Introduction

Studies of locomotor performance have been crucial for understanding how behavior and physiology interact at ecological and evolutionary levels (e.g. Bennett and Huey, 1990; Autumn et al., 1999; Irschick and Garland, 2001; Chappell et al., 2004). The 'centrality of organismal performance' paradigm postulates that natural selection typically acts most directly on behavior, which is limited by whole-organismal performance abilities, and these are in turn determined by lower-level (subordinate) morphological, physiological and biochemical traits (Garland and Carter, 1994). In various groups of vertebrates, more active species tend to have elevated locomotor abilities and/or or maximal aerobic capacity (Bennett and Ruben, 1979; Garland, 1999; Irschick and Garland, 2001; Weibel et al., 2004), suggesting that activity levels and metabolism have evolved in a correlated fashion. That should not necessarily be the case unless (1) activity levels are limited by aerobic capacity and/or (2) the additive genetic correlation between these traits is positive and conserved among vertebrates. Although interspecific comparative studies are more traditional, techniques of experimental evolution, including artificial selection experiments, provide powerful tools with which to test hypotheses related to the correlated evolution of behavior, performance and subordinate traits (Bennett, 2003; Garland, 2003; Bradley and Folk, 2004; Swallow and Garland, 2005).

We tested whether selection for increased locomotor activity affected aerobic capacity in four replicate lines of house mice bred for high voluntary wheel-running (Selected or S lines) as compared with their four random-bred Control (C) lines (Swallow et al., 1998a; Garland, 2003). The S lines have run about 170% more than C (revolutions/day) since generation 16 (see Garland, 2003), and most of the increase in distance has been achieved by increasing average running speed rather than time spent running (the relative importance of each component is sex-dependent, however; Koteja et al., 1999a,b; Rhodes et al., 2000; Swallow et al., 1998b; Girard et al., 2001). Based on videotape analyses of instantaneous running speeds, Girard et al. (2001) concluded that S females from generation 23 ran twice as fast as C females (about 0.76 m s^{-1} and 0.38 m s^{-1}, respectively), as well as more intermittently, during the time of peak wheel-running. Intuitively, $\dot{V}_{O_2max}$ might be expected to evolve in concert with higher running activity because it

determines the ceiling of sustainable exercise. Hence, individuals with higher $\dot{V}_{O2max}$ – and higher maximum aerobic speeds (MAS) – might have been favored by the selection regime, at least once they had evolved sufficiently high activity levels to tax their aerobic limits (see Koteja et al., 1999a; Girard et al., 2001). Indeed, a study at generation 10 reported a small (6%) but statistically significant increase in treadmill $\dot{V}_{O2max}$ of the S lines as compared with C (Swallow et al., 1998b).

Since the analyses of Swallow et al. (1998b), these lines have been through more than 20 additional generations of selection. Males from S lines that ran on average 75% more than C at generation 10 (Swallow et al., 1998b) now run about 190% more than C (generation 29; Rhodes et al., 2003). It is possible, therefore, that $\dot{V}_{O2max}$ has evolved even further. The first goal of the present study was to test whether $\dot{V}_{O2max}$ has continued to coadapt with additional selection for high wheel-running. To avoid potential problems arising from differences between S and C in motivation to run, we employed two separate protocols to estimate $\dot{V}_{O2max}$: forced exercise on a treadmill and cold-exposure in a He-O_2 (heliox) atmosphere. The latter protocol does not involve 'motivation' in any conventional sense and, in small rodents, may elicit different values of oxygen consumption than are attained in forced exercise protocols (e.g. see Chappell et al., 1995; Rezende et al., 2004a).

Our second goal was to measure directly maximum metabolic rates during voluntary wheel-running. Since generation 16, running distance by the S lines has been at an apparent selection limit or plateau (e.g. fig. 2A in Rhodes et al., 2003; Garland, 2003). Because running speed, rather than running time, was the main factor explaining differences between S and C, and maximal running speeds in S lines are close to their predicted MAS (Girard et al., 2001), this 'ceiling' in running activity might be related to constrained aerobic capacity. Pharmacological studies support the hypothesis that physiological and/or biomechanical factors might limit further evolution of wheel-running, because none of the drugs tested has increased wheel-running in S lines whereas several increased voluntary running activity in C (Rhodes et al., 2001, 2003, 2005; Rhodes and Garland, 2003; Li et al., 2004). We also tested whether mice from S lines occasionally run voluntarily at speeds that exceed MAS, i.e. run 'wind sprints'. Studies of various vertebrates, including human beings, indicate that they generally do not choose to exercise at speeds above their MAS (or anaerobic threshold in humans; Taigen and Beuchat, 1984; Powers and Howley, 2001; Chappell et al., 2004), but what about animals that have been purpose-bred for high voluntary activity?

Materials and methods

The selection experiment

We studied females from generation 35 of artificial selection for high voluntary wheel-running activity (Swallow et al., 1998a; Garland, 2003). The original population was the outbred, genetically variable Hsd:ICR strain of laboratory house mice *Mus domesticus* (Schwarz and Schwarz, 1943). After two generations of random mating, mice were randomly paired and assigned to 8 closed lines (10–15 pairs in each, about 10 families per line were used per generation). In each subsequent generation, when offspring were 6–8 weeks old, they were housed individually with access to a running wheel for 6 days. Daily wheel-running was monitored by an automated system.

Wheel-running was quantified as the total number of revolutions run on days 5 and 6 of the 6-day test. In the four S lines, the highest-running male and female from each family were selected as breeders to propagate the lines to the next generation. In the four C lines, a male and a female were randomly chosen from each family. Within all lines, the chosen breeders were randomly paired, except that sibling matings were not allowed. Selection was suspended for generations 32–35 as the colony was transferred from the University of Wisconsin-Madison to California.

Protocol

Forty eight females (6 per line, each from a different family) were measured in the following protocol. After weaning at 21 days of age, two individuals each from C and S lines were randomly mixed, four per cage, and maintained with food and water *ad libitum*. Measurements began at about 8 weeks of age; the schedule for each female is summarized in Table 1. In short, animals had access to wheels for a total of 6 days, mimicking the conditions of the selection experiment. Mice had access to wheels from days 1–4 (Table 1), as used routinely during selective breeding. On day 5, they were placed inside the wheel metabolic chamber (see below), and $\dot{V}_{O_2}$ during wheel-running was recorded during days 5 and 6 (i.e. over a 48 h period). Measurements were then performed twice for each individual on the treadmill and in a heliox atmosphere during consecutive days (Table 1, details below). To avoid potential circadian effects, treadmill and heliox trials were performed between 20:00 and 22:00 h, which corresponds to the period of highest voluntary wheel activity on a 12 h:12 h light cycle with lights on at 7:00 h (Girard and Garland, 2002; Rhodes et al., 2003).

Metabolic rate during voluntary activity

Performance during voluntary exercise was measured using the protocol and equipment reported in Chappell et al. (2004), which also provides figures and URLs for photographs of the setup. Briefly, we enclosed one of the wheels (circumference 1.12 m) and its attached standard plastic mouse cage, as used in the routine selection protocol, within an airtight Lucite housing. Mice could enter and exit the wheel at will through an access port cut into the side of the mouse cage. The mouse cage contained bedding (wood shavings), a food hopper and a drinking tube. Food and water were available *ad libitum* during measurements.

Two such metabolic wheel enclosures were housed in a temperature control cabinet (range between 18 and 27°C in a daily cycle) and photoperiod (12 h:12 h L:D, dark period

Table 1. *Sequence of measurements of $\dot{V}_{O_2}$ during voluntary exercise on wheels, forced exercise on a motorized treadmill, and acute cold exposure in a He-O_2 atmosphere (heliox)*

Days 1–4	Day 5	Day 6	Day 8	Day 9	Day 10	Day 11
Access to wheel	$\dot{V}_{O_2}$ recording in wheel (48 h)		Treadmill 1	Treadmill 2	Heliox 1	Heliox 2

Mice were weighed on days 1 (when first placed into wheels), 5 (when moved into wheel metabolic chambers), 6 (when removed from wheel chambers), and before each treadmill and heliox trial.

19:00–07:00 h, as in the room) housing the breeding colony. Paired incurrent and excurrent ports provided for airflow, and an internal fan rapidly recirculated air within the enclosure to facilitate mixing. The respirometry system and measurements were identical to those explained in Chappell et al. (2004). Mice were weighed (±0.05 g) before entering the wheel (day 0), as well as before and after trials in the metabolic chamber. At the completion of days 4 or 5 of wheel-running, mice were transported with their respective cages from their acclimation wheels to a wheel metabolic enclosure. Given that only two animals could be measured at once, measurements were randomly scheduled across lines, except that we roughly attempted to control for age effects (i.e. mice that were born first were also measured first), and we always measured one S and one C female concomitantly.

Although each mouse was measured in its own cage (with its own bedding, etc.), it was logistically difficult to clean the metabolic chambers between measurements. We tested rotational resistance before and after each measurement by spinning wheels to high speed (~80 revs min^{-1}) with an electric drill fitted with a rubber friction disk, and then monitoring the time needed for speed to decay to zero. Order of measurement, wheel number, and resistance did not significantly affect any variable, however, and were not included in the final statistical analyses.

We dried subsampled air with magnesium perchlorate (CO_2 was not removed in order to avoid the large volumes of soda lime or frequent scrubber changes that otherwise would be necessary for long-duration tests), and calculated $\dot{V}_{O_2}$ (ml min^{-1}) as:

$$\dot{V}_{O_2} = \dot{V}\,[(F_{iO_2}-F_{eO_2})-F_{eO_2}\times(F_{eCO_2}-F_{iCO_2})]\,/\,(1-F_{eO_2})\,, \quad (1)$$

where $\dot{V}$ is flow rate (ml min^{-1} STP; Standard Temperature and Pressure), F_i and F_e are the fractional gas concentrations (O_2 or CO_2) in incurrent and excurrent air, respectively. Experimentation with step changes in gas concentrations (described above) indicated that the best resolution of $\dot{V}_{O_2}$ was obtained with 7-point nearest-neighbor smoothing repeated 20 times, prior to the calculations of 'instantaneous' values (Bartholomew et al., 1981). Analyses such as smoothing, baseline and lag time corrections, 'instantaneous' transformation (the effective volume of the chamber, calculated as described in Chappell et al. (2004), was 17.0 l) and $\dot{V}_{O_2}$ calculations with Eq. 1 were performed using 'LabAnalyst' software (Warthog Systems, Riverside, CA, USA).

Maximum $\dot{V}_{O_2}$ during exercise

We used open-flow respirometry to determine maximum aerobic performance as maximum rate of oxygen consumption ($\dot{V}_{O2max}$). After voluntary activity trials on the wheels, we estimated $\dot{V}_{O2max}$ during both forced exercise and acute cold-exposure. Each individual was tested once a day and $\dot{V}_{O2max}$ was estimated twice with each method on consecutive days (Table 1). Hence, all values were obtained within a 4 day window, and less than 6 days apart from measurements of $\dot{V}_{O_2}$ in the wheel.

Mice were run in an enclosed motorized treadmill, as described previously (Hayes and Chappell, 1990; Chappell et al., 2003). The treadmill had an inclination of 25°, which has been reported to yield maximal values of $\dot{V}_{O_2}$ in laboratory house mice (Kemi et al., 2002). Mice were placed in the working section (6 cm wide, 7 cm high, 13.5 cm long), allowed a 1–2 min acclimation period, and then run at increasing speeds, starting at 0.15–0.2 m s^{-1} and raised in step increments of about 0.1 m s^{-1} every 45 s, until the mouse could no longer maintain position and $\dot{V}_{O_2}$ no longer increased. The maximum speed that each individual attained on the treadmill was recorded. Tests lasted from 6 to 17 min, and reference readings of incurrent gas were obtained at the start and end of measurements. Trial quality was also assessed using a subjective scale (five categories, from poor to excellent: Swallow et al., 1998b), and poor trials (a single trial for one C female where trial quality=1) were not included in the final analyses.

Changes in O_2 concentration were measured using an Ametek/Applied Electrochemistry S-3A analyzer (Pittsburg, Pennsylvania, USA), and recorded on a Macintosh computer equipped with National Instruments A-D converters and Warthog software. Gas flow (2100 ml min^{-1}) was regulated with Tylan mass flow controllers (Billerica, MA, USA) upstream from the treadmill. About 100 ml min^{-1} of excurrent gas was sampled, dried with magnesium perchloride, and scrubbed of CO_2 before going to the O_2 analyzer. Because of the short duration of treadmill tests, we applied the 'instantaneous' transformation (Bartholomew et al., 1981) to resolve rapid changes in metabolism. Effective volume of the treadmill was 903 ml. We calculated $\dot{V}_{O_2}$ using Eq. 2 and computed $\dot{V}_{O2max}$ as the highest instantaneous $\dot{V}_{O_2}$ (ml min^{-1}) averaged over continuous 1 min intervals, using LabAnalyst:

$$\dot{V}_{O_2} = \dot{V}(F_{iO_2}-F_{eO_2})\,/\,(1-F_{eO_2})\,, \quad (2)$$

where F_{iO_2} was 0.2095 and F_{eO_2} was always >0.204.

Maximum $\dot{V}_{O_2}$ during cold-exposure

We measured $\dot{V}_{O2max}$ in an atmosphere of heliox (79% He, 21% O_2), which is several-fold more conductive than air (Chappell and Bachman, 1995), using the same gas flow system described for treadmill trials. A Plexiglas metabolism chamber (volume 600 ml) was supplied with heliox at 1700 ml min^{-1}. An environmental cabinet controlled the temperature of the metabolic chamber. Animals were weighed (±0.05 g) and placed in the metabolism chamber at an ambient temperature of about –1°C, and recording began as soon as the system was completely flushed with heliox (about 1 min). Temperature declined around 0.5°C min^{-1}. We terminated measurements and removed animals when $\dot{V}_{O_2}$ declined below initial values for more than 1 min, or did not increase as temperature decreased more than 2°C. Trials lasted no longer than 15 min.

During these trials we also measured breathing frequency (f; breaths s^{-1}) and tidal volume (V_T; ml) using whole-body plethysmography (Withers, 1977; Chappell, 1985). Chamber pressure changes caused by warming and humidification of tidal air were recorded with a pressure transducer (PX 164-010, Omega Instruments, Stamford, Connecticut, USA) connected to the computer and sampled at 125 Hz. The system was calibrated after each trial by injecting a known volume of heliox (1.0 ml) into the chamber at rates matching the kinetics of inhalation cycles. Tidal volume was calculated from calibration data and pressure changes during inspiration according to Malan (1973); we assumed lung temperature was 37°C and that air in the respiratory tract was 100% saturated with water vapor. [Although body temperature T_b probably decreases during heliox trials, 2°C would affect V_T estimates by only about 2% when ambient temperature $T_a \approx 0$°C. Hence, 37°C was assumed for convenience because real T_b at $\dot{V}_{O2max}$ was unknown (Mortola and Frappell, 1998; Rezende et al., 2004b).] Oxygen extraction efficiency (O_2EE,%) was calculated as $100\dot{V}_{O2max}/(0.2095 \times \text{minute volume})$, where minute volume is ($f V_T$). Immediately after removing an animal from the chamber, T_b was determined to ±0.1°C using a rectal thermocouple connected to a BAT-12 thermometer (Sensortek, Lake Forest, California, USA). Mice were mildly hypothermic after trials (T_b=35.11±0.16°C, mean ± S.E.M.), indicating that $\dot{V}_{O2max}$ was probably achieved.

Eq. 2 was used to calculate $\dot{V}_{O_2}$, although in this case we did not employ the 'instantaneous' transformation because the chamber volume was small relative to the flow rate, mice usually remained still during heliox trials, and steady-state values of $\dot{V}_{O_2}$ were obtained.

Statistical analyses

Because we were interested in values of maximum performance for each individual, we selected the highest 1 min value of voluntary $\dot{V}_{O_2}$ obtained during either of the 2 days of wheel-running ($\dot{V}_{O2max,W}$), the two treadmill ($\dot{V}_{O2max,T}$) or the two heliox ($\dot{V}_{O2max,H}$) trials. We also selected the lowest 5 min average $\dot{V}_{O_2}$ throughout the 2 days of measurements in the wheels as an estimate of resting metabolic rates (RMR_W). For repeatability analyses, we selected highest or lowest values within days 1 and 2 separately (see below). Wheel data for one S female were discarded because of measurement problems (however, records on treadmill and heliox for this female were included).

Analyses were performed with SPSS for Windows version 11.5 (SPSS Inc., Chicago, IL, USA, 2002) and SAS version 8.02 (SAS Institute, Cary, NC, USA, 1996). First, to determine effects of selection, we estimated line-type effects (S *vs* C) using a one-way nested analysis of covariance (ANCOVA) with type III sums of squares, using SAS PROC MIXED (a program in the SAS statistical package that allows testing simultaneously for fixed- and random-effects; i.e. 'mixed models'). Line type was the grouping variable (fixed factor) and replicate lines (N=8) were nested within line type as random factors. Body mass and age were included as covariates, but many variables were also analyzed without mass in the model (e.g. maximum speeds, distances) as these were not correlated with mass. Because of differences in body mass among lines (see Discussion), head-to-rump length (*HRL*) was used as an additional indicator of size. Line random effects were determined by comparing the log likelihoods of the models with and without line (twice the difference in log likelihoods follows a χ^2 distribution with 1 d.f.). Adjusted least-squares means (and S.E.M.) were used to estimate the difference between S and C lines. Second, regular ANCOVAs were performed for S and C separately, including line as a random factor (4 lines) and using the same covariates of the nested model. Although we report P-values for two-tailed hypotheses for simplicity, we adopted directional hypothesis whenever it was possible (i.e. S mice are expected to run faster than C) to increase statistical power.

We also assessed how $\dot{V}_{O_2}$ changed between wheel, treadmill and heliox trials, compared running performance during voluntary *vs* forced exercise, and whether these changes differed between line types and lines, employing general linear model for repeated measures (GLM procedure in SPSS). Individuals were experimental units, type of measurement (treadmill, heliox or wheels) was the within-subjects factor, and selection history (S *vs* C) and lines were between-subjects factors. To determine how variables differed between trials, contrasts (i.e. the difference between successive values for each individual) were compared employing multivariate ANOVAs (test of within-subjects contrasts).

Repeatability was estimated in two different ways (Hayes and Jenkins, 1997). First, residuals from both nested and regular ANCOVAs obtained in the first and second trial were tested using Pearson product–moment correlation. Second, we obtained the intraclass correlation with a one-way ANOVA, employing again the residuals from nested and regular ANCOVA. The intraclass correlation coefficient τ was calculated as (groups MS–error MS)/[groups MS+(n–1)error MS], where MS is mean square, n (=2) is the number of repeated measures per individual, and 'groups' are the individual mice (Zar, 1999; p. 404). Differences between days

Table 2. *Repeatability indices obtained with Pearson's correlation (*r*) and intraclass coefficient (*τ*) from one-way ANOVA*

	Residuals from regular regression			Residuals from nested model		
	R	τ	one-tailed *P*	*R*	τ	one-tailed *P*
$\dot{V}_{O2max,W}$ (*N*=47)	0.844	0.982	**<0.0001**	0.750	0.969	**<0.0001**
$\dot{V}_{O2max,T}$ (*N*=48)	0.420	0.816	**0.0015**	0.342	0.719	**0.009**
$\dot{V}_{O2max,H}$ (*N*=47)	0.134	–0.098	0.185	0.095	–0.418	0.263
T_b,H (*N*=34)[a]	0.360	0.653	**0.018**	0.124	–0.336	0.243
$RMR_{5,W}$	0.490	0.868	**<0.001**	0.530	0.887	**<0.001**
$Speed_{Max,W}$	0.879	0.984	**<0.0001**	0.843	0.981	**<0.0001**
$Speed_{Max,T}$	0.677	0.950	**<0.0001**	0.700	0.956	**<0.0001**

[a]Two influential points (1 S and 1 C) were removed.

P values are mathematically identical because measurements were performed only twice.

Repeatability was calculated from residuals of regular multiple regressions of metabolism on body mass and age, or residuals obtained from the entire nested model (SAS PROC MIXED) with line type and line within line types as fixed and random factors, respectively, and body mass and age as covariates.

N=48 females approximately 8 weeks old. Values in bold indicate statistical significance ($P<0.05$). All values were the highest 1 min averages.

W, wheel (voluntary exercise); T, treadmill (forced exercise); H, heliox, except RMR (lowest 5 min period of $\dot{V}_{O_2}$ during wheel trial).

1 and 2 on wheels, or trials 1 and 2 for treadmill and heliox, were addressed using paired *t*-tests.

Results

As expected, S lines ran more (+ 74%) during days 5 and 6 than did C lines (adjusted means and standard errors were 18370±1015 m *vs* 10570±995 m, respectively; $P_{selection}$=0.0015). The difference in running distance was mainly because S ran faster than C (22.4±1.3 and 13.6±1.2 m min^{-1}, respectively; $P_{selection}$=0.0026), rather than for longer periods (388±26 *vs* 385±26 min day^{-1}). Body size, estimated using either mass (22.8±0.4 *vs* 26.8±0.4 g; $F_{1,6}$=39.83, $P<0.001$) or *HRL* (9.45±0.06 *vs* 9.80±0.06 cm; $F_{1,6}$=19.13, $P<0.005$), was significantly smaller in S lines.

Most variables were significantly repeatable, independent of whether residuals were obtained from simple linear regressions or from the complete nested ANCOVA model (Table 2). $\dot{V}_{O2max}$ obtained in heliox, however, was not significantly repeatable by either method. Absolute values of $\dot{V}_{O2max}$ in heliox (i.e. not residuals) were significantly correlated with each other, although the relationship was relatively weak (r=0.311, one-tailed P=0.017). Body temperature after heliox trials also was not repeatable (after removing two influential points from the nested model; Table 2), and was not correlated with body mass (P=0.374).

Considering all mice, $\dot{V}_{O2max}$ during first and second trials did not differ during wheel trials (t_{46}=–1.858, two-tailed P=0.070), treadmill trials (t_{47}=1.201, P=0.236) or in heliox (t_{46}=–1.679, P=0.100). No significant differences between days were detected in maximum voluntary running speeds (t_{46}=–0.801, P=0.426). Results remained unchanged when S and C were analyzed separately. Maximum speed attained during forced exercise tended to be higher during the first trial (t_{47}=1.943, P=0.059). In heliox, S mice had a significantly higher $\dot{V}_{O2max}$ during the second trial (6.27±0.83 *vs* 6.77±1.00, one-tailed $P<0.03$), which was not the case in C (P=0.498).

Metabolism, performance, and selection history

Selected and C lines did not differ significantly in either forced-exercise or cold-exposure $\dot{V}_{O2max}$ (Table 3). Because all metabolic variables increased significantly with body mass, we tested whether reduction in size in S lines had changed the interaction (i.e. slope) between mass and $\dot{V}_{O2max}$ (mass×selection history factor was tested over mass×line in SAS PROC MIXED). There were no significant differences between S and C in the slopes of mass *vs* $\dot{V}_{O2max}$ on the treadmill ($F_{1,6}$=0.69, P=0.44), in cold-exposure ($F_{1,6}$=4.12, P=0.09) or during wheel-running ($F_{1,6}$=1.24, P=0.31). Line effects within each line type, obtained with conventional ANCOVAs, were significant only for $\dot{V}_{O2max}$ on the treadmill in C mice ($P<0.04$), but not when *HRL* was included as an indicator of size instead of mass (P=0.36). As expected, size increased significantly with age (body mass: $F_{1,39}$=21.79, $P<0.001$; *HRL*: $F_{1,39}$=37.93, $P<0.001$). Maximum metabolism during voluntary exercise was significantly higher in S lines (Table 3). That conclusion was unchanged when $\dot{V}_{O2max}$ was expressed in absolute terms or on a per gram basis (always with body mass as a covariate).

Aerobic scope during cold exposure ($\dot{V}_{O2max}$ in heliox/RMR during wheel trials) was significantly higher in S lines only after removing body mass from the model (with mass in the model, $P_{selection}$=0.16, $P_{body\ mass}$=0.79). After accounting for body size, T_b following cold-exposure was marginally higher in S lines after trial 1 (34.3±0.3 *vs* 35.4±0.3°C for C and S, respectively, N=39, $P<0.033$), but not after trial 2 (35.0 *vs* 35.4°C, N=44, P=0.37).

Voluntary wheel-running speeds were significantly higher in S, whereas maximum speeds attained during forced exercise did not differ between line types (Table 3, Fig. 1). There were no significant differences in trial quality (see Materials and

Table 3. *Metabolic and locomotor variables for 48 female mice from lines selected for high wheel-running performance or bred randomly as controls*

	Selected	Control	S/C	$P_{selection}$	P_{lines}	$P_{body\ mass}$
$\dot{V}_{O2max,W}$ ($mlO_2\ min^{-1}$)	4.465±0.088	3.710±0.086	1.20	**0.002**	1	**<0.001**
$\dot{V}_{O2max,T}$ ($mlO_2\ min^{-1}$)	5.434±0.170	5.096±0.170	1.07	0.238	0.083	**<0.001**
$\dot{V}_{O2max,H}$ ($mlO_2\ min^{-1}$)	7.149±0.180	6.747±0.180	1.06	0.189	0.123	**<0.001**
$\dot{V}_{O2max,T}$ ($mlO_2\ g^{-1}\ min^{-1}$)[a]	0.222±0.006	0.207±0.006	1.07	0.175	0.332	**0.017**
$\dot{V}_{O2max,H}$ ($mlO_2\ g^{-1}\ min^{-1}$)[a]	0.289±0.007	0.272±0.007	1.06	0.160	0.286	**0.001**
$RMR_{5,W}$ ($mlO_2\ min^{-1}$)	0.845±0.026	0.836±0.026	1.01	0.831	0.653	**<0.001**
$Scope_W$	5.349±0.123	4.33±0.120	1.23	**0.001**	1	
$Scope_T$	6.556±0.250	5.985±0.250	1.10	0.157	0.243	
$Scope_H$	8.656±0.222	7.883±0.222	1.10	**0.049**	1	
$Speed_{Max,W}$ ($m\ min^{-1}$)	35.395±1.783	21.896±1.753	1.62	**0.002**	0.674	
$Speed_{Max,T}$ ($m\ min^{-1}$)	37.848±1.855	33.210±1.855	1.14	0.128	0.515	

[a]Mass-specific metabolic rates.

S, selected; C, controls.

Significance of the effects of selection history (S *vs* C), replicate line within line type, and body mass are from nested analysis of covariance (ANCOVA), including age as a covariate. Adjusted means were calculated from SAS PROC MIXED for a female of 25.3±0.2 g (mass indices were not the same for all traits, because measurements were performed on separate days) and 72 days of age (for scope and speed, mass was not included in the model; see Statistical analyses).

W, wheel (voluntary exercise); T, treadmill (forced exercise); H, heliox.

Values are adjusted means ± S.E.M.; values in bold indicate significant effects ($P \leqslant 0.05$).

$RMR_{5,W}$ (lowest 5 min $\dot{V}_{O_2}$ average during wheel trials); Scope, aerobic scope (=$\dot{V}_{O2max}$/RMR).

methods) on the treadmill between S and C (P=0.175), and this variable was never a significant predictor of $\dot{V}_{O2max}$. Effects of selection on maximum treadmill speeds were statistically significant, however, when size – either mass or *HRL* – was included as a covariate (S>C: one-tailed P<0.028 and P<0.040, respectively). In addition, repeated measures performed separately for each line type showed a highly significant effect of measurement (treadmill *vs* wheel) on running speeds in C ($F_{1,20}$=57.19, P<0.001), but no effect in S ($F_{1,19}$=0.61, P<0.416; Fig. 1B). All three estimates of $\dot{V}_{O2max}$ differed significantly from each other: effects of experimental protocol were significant in the 'pooled' analysis or when S and C were analyzed separately. $\dot{V}_{O2max}$ obtained in heliox was higher than on the treadmill, which was higher than maximum voluntary $\dot{V}_{O_2}$ (Fig. 2), with a significant trial×line type interaction (tested over trial×line type×line, $F_{2,12}$=4.323, P=0.038) probably due to higher voluntary $\dot{V}_{O2max}$ in S lines. The ratio between $\dot{V}_{O2max}$ during cold exposure and forced exercise ($\dot{V}_{O2max,H}/\dot{V}_{O2max,T}$) did not differ between S and C ($F_{1,6}$=0.02, P<0.90).

Breathing frequency measured during cold-exposure was 0.55 Hz higher in S mice. The difference was marginally significant with mass and $\dot{V}_{O2max}$ in the model, and clearly significant when mass, or mass and $\dot{V}_{O2max}$, were removed from the model (Table 4). Tidal and minute volumes did not differ between line types, but did increase with mass. None of these variables were significantly correlated with $\dot{V}_{O2max}$ in heliox.

Correlations between metabolism and performance

Many of the predicted correlations between metabolism and running performance were statistically significant (e.g. $\dot{V}_{O2max}$

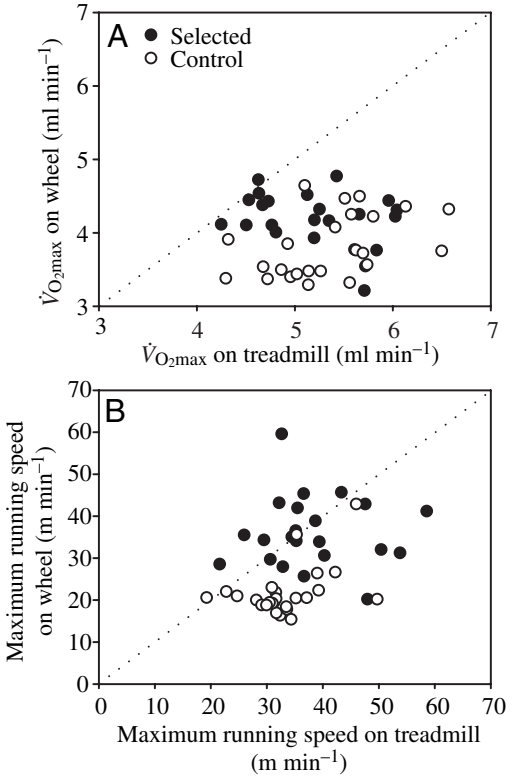

Fig. 1. Relationship between maximum rates of oxygen consumption ($\dot{V}_{O2max}$; A) and running speeds (B) during voluntary wheel-running and forced exercise on the treadmill. Each value represents one of 47 female mice measured over 4 consecutive trial days (see Materials and methods and Table 1). Dotted lines indicate equal values. Each point represents the mean highest 60 s recorded for each individual.

vs maximum running speed on treadmill and on wheels for 'pooled' data), although results differed slightly depending on selection history (Table 5). There was only a weak positive correlation between $\dot{V}_{O_2max}$ obtained in heliox and on the treadmill, observed when line types were pooled together (one-tailed P=0.045).

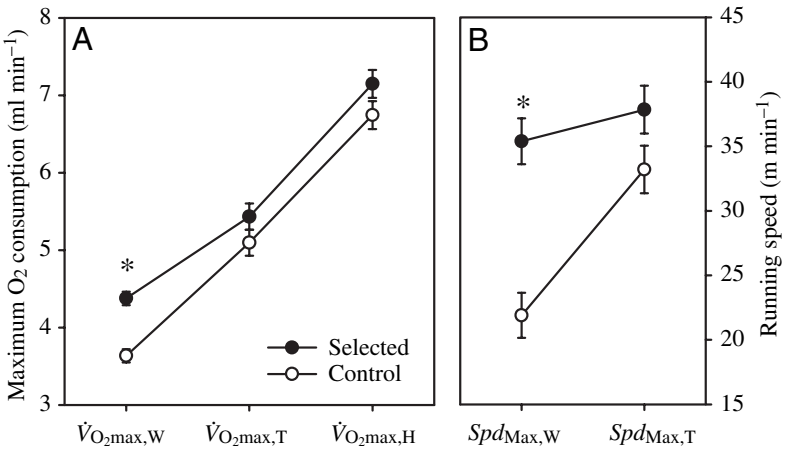

Fig. 2. (A) Maximum 1 min average $\dot{V}_{O_2}$ during voluntary exercise on the wheels ($\dot{V}_{O_2max,W}$), forced exercise on the treadmill ($\dot{V}_{O_2max,T}$), and acute cold exposure in heliox ($\dot{V}_{O_2max,H}$), in 47 females from S and C lines. (B) Maximum running speeds obtained on the treadmill ($Spd_{Max,T}$) and on the wheel ($Spd_{Max,W}$) for these same individuals. Values for oxygen consumption are adjusted means from SAS PROC MIXED for a female of 25.3 g; values for speed are not adjusted for mass (see text). Asterisks indicate significant differences between S and C lines (P<0.05, Table 3). Vertical bars indicate ± 1 S.E.M.

Discussion

Our values of $\dot{V}_{O_2max}$ on the treadmill are in agreement with previous values reported for generation 10 in these same lines (Swallow et al., 1998b). Their values in sedentary S and C lines (table 3 in their study) were 1.5% and 2.9% higher than in our mice, respectively. Due to the 25° inclination of the treadmill in the present study (necessitated because the present treadmill had a lower maximum speed than the one used by Swallow et al., 1998b), however, our mice attained lower maximum running speeds. This presumably also explains why some S individuals attained higher running speeds, but not higher $\dot{V}_{O_2}$, on the wheel (Fig. 1). The $\dot{V}_{O_2max}$ obtained during cold-exposure averaged 32% higher than on the treadmill, with a similar mean increase for both S and C (31.6% and 32.4% respectively; Table 3). The $\dot{V}_{O_2max}$ during voluntary wheel-running was significantly lower than during forced exercise, and the magnitude of the difference was dependent on selection history; during peak running (highest 1 min values) S mice ran voluntarily at an average of 80% of their $\dot{V}_{O_2max}$, whereas C mice ran at 71% (Table 3, Fig. 2). Chappell et al. (2004) reported that deer mice *Peromyscus maniculatus* measured with the same system at a similar temperature (22–25°C) ran voluntarily at around 72% of their $\dot{V}_{O_2max}$ during peak 1 min running, similar to our C lines.

Repeatability was high for most of the metabolic traits, which is consistent with previous studies of food consumption in these same lines under cold-exposure (Koteja et al., 2000), $\dot{V}_{O_2max}$ during forced exercise in deer mice and ground squirrels, *Spermophilus beldingi* (Hayes and Chappell, 1990; Chappell et al., 1995), and deer mice measured in the same enclosed wheels used here (Chappell et al., 2004). The low repeatability of $\dot{V}_{O_2max}$ in heliox between days was unexpected, however. Chappell et al. (1995) reported that repeatability measured over different test periods in *S. beldingi* was generally higher in exercise than thermogenic $\dot{V}_{O_2max}$, although the later was repeatable over relatively longer periods (e.g. several days). In addition, $\dot{V}_{O_2max}$ obtained during cold-exposure was repeatable in deer mice over long periods (over 8 weeks), and across different acclimation temperatures (Hayes and Chappell, 1990; Rezende et al., 2004b).

Table 4. *Ventilatory variables during heliox trials for 48 female mice from lines selected for high wheel-running performance or bred randomly as controls*

	Selected	Control	$P_{selection}$	P_{lines}	$P_{body\ mass}$	$P(\dot{V}_{O_2max})$
Breathing frequency (Hz)	6.861±0.147	6.306±0.147	0.052	0.294	0.690	0.518
Breathing frequency (Hz)[a]	6.878±0.143	6.289±0.143	**0.037**	0.308	0.375	
Breathing frequency (Hz)[b]	6.823±0.130	6.345±0.130	**0.041**	0.276		
Tidal volume (ml)	0.270±0.008	0.274±0.008	0.775	1	**0.017**	0.890
Minute volume (ml min^{-1})	111.4±3.8	103.7±3.8	0.253	1	**0.026**	0.828
O_2 extraction (%)	31.24±1.22	32.39±1.22	0.580	1	0.068	**0.0064**

[a]$\dot{V}_{O_2max}$ not included in the model.

[b]$\dot{V}_{O_2max}$ and body mass not included in the model.

S, selected; C, controls.

Significance of the effects of selection history, line, and body mass are from nested analysis of covariance (ANCOVA), including age as a covariate.

Adjusted means and S.E.M. were calculated from SAS PROC MIXED for a female of 25.0 g and 71 days of age. Values are means ± S.E.M.; values in bold indicate statistically significant effects (P≤0.05).

Table 5. *Correlations (r) between residuals of $\dot{V}_{O2max}$ and maximum speeds attained during forced and voluntary exercise, and during cold exposure in heliox, for 47 female mice*

		$\dot{V}_{O_2max,H}$	$\dot{V}_{O_2max,W}$	$RMR_{5,W}$	$Speed_{Max,T}$	$Speed_{Max,W}$
Pooled (S and C)	$\dot{V}_{O_2max,T}$	0.251*	–0.116	0.063	0.366**	–0.083
	$\dot{V}_{O_2max,H}$		0.054	0.003	–0.115	–0.120
	$\dot{V}_{O_2max,W}$			0.284*	0.077	0.356**
	$RMR_{5,W}$				0.077	–0.266*
	$Speed_{Max,T}$					0.118
Selected (S)	$\dot{V}_{O_2max,T}$	0.180	–0.378*	0.156	0.248	–0.355*
	$\dot{V}_{O_2max,H}$		–0.046	0.023	–0.151	–0.216
	$\dot{V}_{O_2max,W}$			–0.059	–0.056	0.423*
	$RMR_{5,W}$				0.264	–0.437*
	$Speed_{Max,T}$					–0.107
Control (C)	$\dot{V}_{O_2max,T}$	0.199	0.096	0.007	0.560**	0.205
	$\dot{V}_{O_2max,H}$		–0.060	–0.144	–0.149	–0.064
	$\dot{V}_{O_2max,W}$			0.512**	0.372*	0.373*
	$RMR_{5,W}$				–0.005	–0.128
	$Speed_{Max,T}$					0.292

T, forced (treadmill) and W, voluntary exercise (wheel); H, during cold exposure in heliox.

*P<0.05, **P<0.01 after Pearson's one-tailed correlation, not controlling for multiple comparisons.

Age and body mass were included as covariates for metabolic traits, but only age was considered for speeds. Residuals were obtained from nested ANCOVA (SAS PROC GLM) including line type and line as grouping factors (Pooled), or from regular ANCOVA performed for S or C separately (N=23 and 24, respectively), with line as a random factor.

Critical r values (P<0.05) were 0.238 for pooled data (N=47), 0.330 for C (N=24), and 0.337 for S (N=23).

Domestication might be a confounding factor, however. Richardson et al. (1994) showed that maximum nonshivering thermogenesis (NST) in response to norepinephrine injection tended to be higher in wild *Mus* than in mice from the same strain as used to found the selection experiment, and wild mice had significantly more interscapular brown adipose tissue than their laboratory counterparts. Thus, it is possible that laboratory mice cannot sustain high NST for an entire heliox trial. Another possibility is that overall repeatability of $\dot{V}_{O_2max}$ in heliox was lower because S and C lines differed in their 'training response' to heliox trials (S mice had higher $\dot{V}_{O_2max}$ in the second trial whereas C did not; see Results).

Lack of repeatability of $\dot{V}_{O_2max}$ in heliox is consistent with low or non-significant repeatabilities for final T_b, depending on the inclusion of two influential points. Everything else being equal, one would expect animals with higher $\dot{V}_{O_2}$ to have higher T_b in these trials. There was a significant but weak positive relationship between $\dot{V}_{O_2max}$ and final T_b in trial 1 ($F_{1,36}$=6.47, P=0.015), but not in trial 2 ($F_{1,41}$=1.74, P=0.19).

Performance on wheel, treadmill and heliox

Many correlations predicted *a priori* were significant according to pairwise tests of Pearson product-moment correlations (Table 5). As expected, maximum running speeds on the wheels were positively correlated with maximum voluntary $\dot{V}_{O_2}$ in both S and C (Pearson, P<0.05). There was a strong positive correlation between running speed and $\dot{V}_{O_2max}$ in C lines during forced exercise on the treadmill, but not in S mice (Table 5). Differences in the incremental cost of transport (defined as the slope of the linear regression of $\dot{V}_{O_2}$ on running speed) could explain why the strong correlation between $\dot{V}_{O_2max}$ and speed observed in C is not present in S (see below). Accordingly, S mice tended to attain lower $\dot{V}_{O_2max}$ at high speeds on the treadmill (Fig. 3), although the interaction of line type×speed was never significant in the nested model, either when tested over line or including an additional line×speed term in the model (P>0.20).

Interestingly, the correlation between RMR and $\dot{V}_{O_2max}$ on the wheels differed between S and C lines: the correlation was positive in C lines, but not in S (Table 5). We also found a significant positive correlation in deer mice at a similar temperature (25°C; Chappell et al., 2004). Taken together, these results suggest that in S lines the potential costs of attaining high $\dot{V}_{O_2}$ on the wheels were 'attenuated' (i.e. S lines do not have high maintenance costs in spite of their higher $\dot{V}_{O_2}$ during voluntary running). However, the apparent difference between S and C lines requires further study because in a full nested ANCOVA of wheel $\dot{V}_{O_2max}$ with RMR added as an additional covariate, the line type×RMR interaction was not significant (tested over the line×RMR term: $F_{1,6}$=2.10, $P_{linetype\times RMR}$=0.197). On the other hand, this ANCOVA probably had low power to detect such an interaction (Wahlsten, 1990).

Whether $\dot{V}_{O_2max}$ during forced exercise and cold-exposure are comparable and to what extent they can be considered physiologically and genetically 'the same trait' has been debated in the literature (e.g. Rezende et al., 2004a). If $\dot{V}_{O_2max}$ was ultimately restricted by pulmonary and cardiovascular systems (central limitation hypothesis), then $\dot{V}_{O_2max}$ obtained under different experimental conditions should be similar, and

maximum values on the treadmill and in heliox should be highly correlated (Hammond and Diamond, 1997; Bacigalupe and Bozinovic, 2002). In contrast, if the physiology underlying each index is not the same, then $\dot{V}_{O2max}$ in heliox and during forced exercise could be quite different, and they should not necessarily be correlated.

Our results provide partial support to both alternatives. There was a significant but weak correlation between $\dot{V}_{O2max}$ on the treadmill and in heliox in the complete nested model (Table 5), consistent with (some) common physiology underlying these traits. Significant correlations between these traits were reported for *Peromyscus* (Hayes and Chappell, 1990; Chappell et al., 2003) and *Spermophilus beldingi* (data from Chappell and Bachman, 1995; reanalyzed in Rezende et al., 2004a). However, $\dot{V}_{O2max}$ during cold-exposure was about 32% higher than during forced exercise, emphasizing that different tissues and processes are involved in attaining maximum values during forced exercise and acute cold-exposure. Recruitment of additional muscles during shivering, and brown adipose tissue for NST (Heldmaier, 1993; Nespolo et al., 2001), may explain why $\dot{V}_{O2max}$ was higher during cold-exposure, as has been described for some species of small mammals (Belding's ground squirrels, Chappell et al., 1995; deer mice, Chappell and Hammond, 2003). In addition, our results show that, at least on the treadmill, $\dot{V}_{O2max}$ is not centrally limited: pulmonary and cardiovascular systems could provide more oxygen than required by muscles while running at maximum aerobic levels (however, limitations at the vascular level might occur in localized regions of the body).

On these wheels, both C and S mice choose to run at speeds below their maximum aerobically sustainable speed, as is also true for *Peromyscus* (Chappell et al., 2004). This result seems to contradict the theoretical prediction that mice (S lines in particular) would run close to maximum aerobic levels on the wheels to maximize running economy (i.e. minimize costs of transport; see Chappell et al. (2004). All else being equal and given enough individual variation, one would expect that mice that ran voluntarily closer to their MAS eventually would have been favored in the selection experiment. Several non-exclusive explanations are possible. First, the protocol of forced exercise may overestimate maximum $\dot{V}_{O_2}$ during voluntary running. Recruitment of additional tissues because of stress responses, for instance, could account for a higher $\dot{V}_{O2max}$ during forced exercise (M.A.C., unpublished data). If that is the case, then perhaps S mice are indeed running at speeds close to their voluntary MAS, and aerobic performance could be constraining the evolution of higher running speeds on the wheels. Another possibility is that mice reach their lactate threshold before attaining $\dot{V}_{O2max}$, as is the case for human beings (Powers and Howley, 2001). Finally, if animals are energy-limited – i.e. if energy cannot be provided to cells at similar rates as O_2 – it may simply be impossible to sustain voluntary running at maximum aerobic levels (i.e. there may have been a trade-off between speed and endurance during voluntary running on wheels).

Nevertheless, S lines run voluntarily at speeds that come closer to eliciting their treadmill $\dot{V}_{O2max}$ (see above) – i.e. selection for longer running distances has produced individuals tending to run closer to their maximum aerobic capacities. Accordingly, because there were no differences in energy expenditure between S and C lines in spite of the 70% difference in running distances at generation 10, Koteja et al. (1999a) concluded that 'running distance over a given period of time (e.g. 24 h) could be increased substantially by increasing speed, with only a small increase in the total cost of activity', supporting the idea that running faster can have an important effect on running economy in these lines.

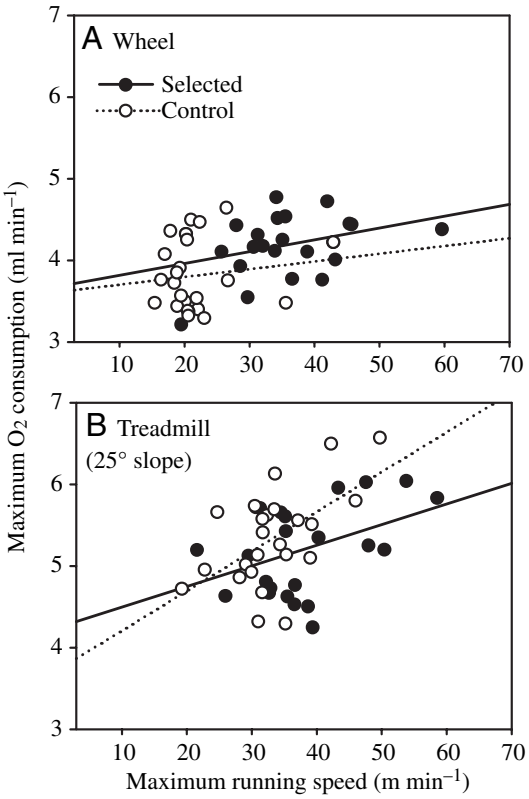

Fig. 3. Relationship between maximum running speeds and $\dot{V}_{O2max}$ during (A) voluntary activity in the wheel and (B) forced exercise on the treadmill, obtained for 47 female mice. Dotted lines show least-squares linear regressions for raw data (i.e. not corrected for covariates). Correlations between $\dot{V}_{O2max}$ and speed after accounting for body mass and age are listed in Table 5.

Effects of selection on $\dot{V}_{O2max}$

Our results show that $\dot{V}_{O2max}$ during forced exercise and cold exposure have not increased significantly in the S lines, despite of a 20.3% increase in maximum voluntary $\dot{V}_{O_2}$ attained during voluntary wheel-running. Although this seems to contradict the results obtained by Swallow et al. (1998b), who reported a small (6%) but significant increase in $\dot{V}_{O2max}$ during forced exercise in S mice, mean values in the present study were on average 6.6% higher in treadmill trials and 6.0% higher in heliox in S lines, although neither difference was significant in our study (Table 3). One possible explanation for the lack of

statistical significance in the present study is that differences among the replicate lines are now greater (P=0.083 for treadmill values; Table 3) than at generation 10 (significance of line effects was not reported in Swallow et al., 1998b). Another possibility is that the results are discrepant because we employed females whereas Swallow et al. (1998b) studied only males, and some responses to selection have been observed to be sex-specific (e.g. S females have evolved relatively higher running speeds than males, compared to C). Possible line differences and sex effects on aerobic capacity will be the subject of future studies with larger sample sizes.

Although mice from the S lines run faster than C on wheels (Koteja et al., 1999a,b; Rhodes et al., 2000; Girard et al., 2001; this study), they do not run faster during forced treadmill trials (see also Swallow et al., 1998b). These results emphasize the fact that S and C mice have similar aerobic capacities and that the major differences in running performance on wheels are apparently a consequence of neurological changes associated with 'willingness to run' (Rhodes et al., 2001, 2003, 2005; Rhodes and Garland, 2003), rather than the physiology – whole aerobic performance, at least – involved with 'being able to run'. Nevertheless, with long-term access to running-wheels, differential training effects related to contrasting running activity in S and C may eventually lead to important physiological differences between them (i.e. physiological plasticity in a genotype by environment interaction; Zhan et al., 1999; Houle-Leroy et al., 2000; Swallow et al., 2005).

Conclusions

Aerobic capacity has not coadapted with increased voluntary wheel-running in our selected lines of mice, and many explanations may account for this lack of correlated response. First, the base population may have lacked additive genetic variation for $\dot{V}_{O2max}$. Consistent with this possibility, Dohm et al. (2001) obtained very low estimates of the additive genetic contribution to individual differences in $\dot{V}_{O2max}$ in the base population, with heritabilities ranging between 0 and 0.64, depending on which quantitative genetic model was used. Second, even if substantial additive genetic variance existed in the base population, the genetic correlation between $\dot{V}_{O2max}$ and wheel-running behavior might be close to zero, suggesting that only few genes influencing wheel-running also affect $\dot{V}_{O2max}$ (Swallow et al., 1998b; Roff, 1997). We emphasize that the lack of response in $\dot{V}_{O2max}$ after 35 generations of selection may be strictly dependent on the genetic background of the base population; other studies have shown than $\dot{V}_{O2max}$ can be heritable and evolutionary labile (Rezende et al., 2004a; Nespolo et al., 2005; Sadowska et al., 2005), including in laboratory rats selected for treadmill running performance (Henderson et al., 2002).

Third, aerobic capacity in these lines could have been 'excessive' in the first place, i.e. animals can run at higher aerobic levels than they are willing to. Indeed, several lines of evidence indicate higher motivation to run in S lines (Rhodes et al., 2001, 2003, 2005; Rhodes and Garland, 2003). Fourth, the correlated increase in frequency of the 'mighty mini-muscle' allele (individuals that are homozygous have gastrocnemius about 50% lighter than normal) in two of the four S lines (Garland et al., 2002) suggests that running efficiency may have been a correlated target of selection. Individuals with the mini-muscle possess gastrocnemius with twice the aerobic capacity per gram of tissue than normal (Houle-Leroy et al., 2000, 2003), but evidence regarding the effect (if any) of the mini-muscle allele on $\dot{V}_{O2max}$ is not yet available. Thus, the mini-muscle allele might have increased in frequency because it reduces the overall cost of locomotion (e.g. Myers and Steudel, 1985; Steudel, 1990; Garland et al., 2002). We are currently testing whether costs of locomotion have evolved in S lines. Fifth, sex-specific responses to selection might be involved, and there also is the possibility that significant effects reported by Swallow et al. (1998b) could result from a type I error.

Although our results suggest that activity levels can evolve without a concomitant change in aerobic capacity in our lines of laboratory mice, extrapolation of these results to wild species should be performed with caution. Domestication of *Mus* strains has apparently led to major effects on behavior without compromising overall physiology to any great extent, at least based on limited comparisons of wild and laboratory house mice and their reciprocal crosses (Dohm et al., 1994; Richardson et al., 1994; Garland et al., 1995). Behavioral and/or whole-organism performance traits (e.g. maximum sprint speeds, $\dot{V}_{O_2}$ on the treadmill) differed considerably between laboratory and wild house mice, despite relatively minor differences in lower-level physiological traits. Hence, aerobic capacity may have been 'excessive' in these lines to the extent that it did not – and does not presently – constrain activity levels or running performance. Therefore, although we have shown that voluntary running performance and activity levels can evolve independently of aerobic capacity, results may depend on the animal model employed. It would be of considerable interest to perform interspecific comparative studies of home range area, voluntary wheel-running and maximal aerobic capacity in other rodents.

We would like to thank L. Karpinski, J. Sinclair and several undergraduates for their help with the mouse colony, and E. Hice and J. Urrutia for constructing the wheel enclosures, environmental cabinets and treadmill. This work was supported by NSF IBN-0212567 (T.G.) and NSF IBN-0111604 (K. A. Hammond and M.A.C.). E.L.R. is grateful to S. Kelly and P. del Agua for discussions, comments, and insights on early versions of the manuscript. All animal procedures are in compliance with the UCR Institutional Animal Care and Use Committee and US laws.

References

Autumn, K., Jindrich, D., DeNardo, D. and Mueller, R. (1999). Locomotor performance at low temperature and the evolution of nocturnality in geckos. *Evolution* **53**, 580-599.

Bacigalupe, L. D. and Bozinovic, F. (2002). Design, limitations and sustained metabolic rate: lessons from small mammals. *J. Exp. Biol.* **205**, 2963-2970.

Bartholomew, G. A., Vleck, D. and Vleck, C. M. (1981). Instantaneous measurements of oxygen consumption during preflight warm-up and postflight cooling in sphinged and saturniid moths. *J. Exp. Biol.* **90**, 17-32.

Bennett, A. F. (2003). Experimental evolution and the Krogh Principle: generating biological novelty for functional and genetic analyses. *Physiol. Biochem. Zool*. **76**, 1-11.

Bennett, A. F. and Huey, R. (1990). Studying the evolution of physiological performance. In *Surveys in Evolutionary Biology*, Vol. 7 (ed. D. J. Futuyma and J. Antonovics), pp. 251-284. Oxford UK: Oxford University Press.

Bennett, A. F. and Ruben, J. A. (1979). Endothermy and activity in vertebrates. *Science* **206**, 649-654.

Bradley, T. J. and Folk, D. G. (2004). Analyses of physiological evolutionary response. *Physiol. Biochem. Zool.* **77**, 1-9.

Chappell, M. A. (1985). Effects of ambient temperature and altitude on ventilation and gas exchange in deer mice (*Peromyscus maniculatus*). *J. Comp. Physiol. B* **155**, 751-758.

Chappell, M. A. and Bachman, G. C. (1995). Aerobic performance in Belding's ground squirrels (*Spermophilus beldingi*): variance, ontogeny, and the aerobic capacity model of endothermy. *Physiol. Zool.* **68**, 421-442.

Chappell, M. A., Bachman, G. C. and Odell, J. P. (1995). Repeatability of maximal aerobic performance in Belding's ground squirrels, *Spermophilus beldingi*. *Funct. Ecol.* **9**, 498-504.

Chappell, M. A., Rezende, E. L. and Hammond, K. A. (2003). Age and aerobic performance in deer mice. *J. Exp. Biol.* **206**, 1221-1231.

Chappell, M. A., Garland, T., Rezende, E. L. and Gomes, F. R. (2004). Voluntary running in deer mice: speed, distance, energy costs, and temperature effects. *J. Exp. Biol.* **207**, 3839-3854.

Dohm, M. R. (2002). Repeatability estimates do not always set an upper limit to heritability. *Funct. Ecol.* **16**, 273-280.

Dohm, M. R., Richardson, C. S. and Garland, T., Jr (1994). Exercise physiology of wild and random-bred laboratory house mice and their reciprocal hybrids. *Am. J. Physiol.* **267**, R1098-R1108.

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

Garland, T., Jr (1999). Laboratory endurance capacity predicts variation in field locomotor behaviour among lizard species. *Anim. Behav.* **57**, 77-83.

Garland, T., Jr (2003). Selection experiments: an under-utilized tool in biomechanics and organismal biology. In *Vertebrate Biomechanics and Evolution* (ed. V. L. Bels, J.-P. Gasc and A. Casinos), pp. 23-56. Oxford, UK: BIOS Scientific Publishers.

Garland, T., Jr and Carter, P. A. (1994). Evolutionary physiology. *Annu. Rev. Physiol.* **56**, 579-621.

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

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

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

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

Hammond, K. A. and Diamond, J. (1997). Maximal sustained energy budgets in humans and animals. *Nature* **386**, 457-462.

Hayes, J. P. and Chappell, M. A. (1990). Individual consistency of maximal oxygen consumption in deer mice. *Funct. Ecol.* **4**, 495-503.

Hayes, J. P. and Jenkins, S. H. (1997). Individual variation in mammals. *J. Mamm.* **78**, 274-293.

Heldmaier, G. (1993). Seasonal acclimatization of small mammals. *Verh. Deutsch. Zool. Gesell.* **86**, 67-77.

Henderson, K. K., Wagner, H., Favret, F., Britton, S. L., Koch, L. G., Wagner, P. D. and Gonzalez, N. C. (2002). Determinants of maximal O_2 uptake in rats selectively bred for endurance running capacity. *J. Appl. Physiol.* **93**, 1265-1274.

Houle-Leroy, P., Garland, T., Swallow, J. G. and Guderley, H. (2000). Effects of voluntary activity and genetic selection on muscle metabolic capacities in house mice *Mus domesticus*. *J. Appl. Physiol.* **89**, 1608-1616.

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

Irschick, D. J. and Garland, T., Jr (2001). Integrating function and ecology in studies of adaptation: investigations of locomotor capacity as a model system. *Ann. Rev. Ecol. Syst.* **32**, 367-396.

Kemi, O. J., Loennechen, J. P., Wisløff, U. and Ellingsen, Ø. (2002). Intensity-controlled treadmill running in mice: cardiac and skeletal muscle hypertrophy. *J. Appl. Physiol.* **93**, 1301-1309.

Koteja, P. (2000). Energy assimilation, parental care and the evolution of endothermy. *Proc. R. Soc. Lond. B* **267**, 479-484.

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

Koteja, P., Garland, T., Sax, J. N., Swallow, J. G. and Carter, P. A. (1999b). Behaviour of house mice artificially selected for high levels of voluntary wheel-running. *Anim. Behav.* **58**, 1307-1318.

Koteja, P., Swallow, J. G., Carter, P. A. and Garland, T., Jr (2000). Individual variation and repeatability of maximum cold-induced energy assimilation in house mice. *Acta Theriol.* **45**, 455-470.

Li, G., Rhodes, J. S., Girard, I., Gammie, S. C. and Garland, T., Jr (2004). Opioid-mediated pain sensitivity in mice bred for high voluntary wheel-running. *Physiol. Behav.* **83**, 515-524.

Malan, A. (1973). Ventilation measured by body plethysmograph in hibernating mammals and poikilotherms. *Resp. Physiol.* **17**, 32-44.

Mortola, J. P. and Frappell, P. B. (1998). On the barometric method for measurements of ventilation, and its use in small animals. *Can. J. Physiol. Pharmacol.* **76**, 937-944.

Myers, M. J. and Steudel, K. (1985). Effects of limb mass and its distribution on the energetic cost of running. *J. Exp. Biol.* **116**, 363-373.

Nespolo, R. F., Bacigalupe, L. D., Rezende, E. L. and Bozinovic, F. (2001). When non-shivering thermogenesis equals maximum metabolic rate: thermal acclimation and phenotypic plasticity of fossorial *Spalacopus cyanus* (Rodentia). *Physiol. Biochem. Zool.* **74**, 325-332.

Nespolo, R. F., Bustamante, D. M., Bacigalupe, L. D. and Bozinovic, F. (2005). Quantitative genetics of bioenergetics and life histories in the wild mammal, *Phyllotis darwini*. *Evolution*. In press.

Powers, S. K. and Howley, E. T. (2001). *Exercise Physiology: Theory and Application to Fitness and Performance*, 4th edn. Boston, USA: McGraw-Hill Publishers.

Rezende, E. L., Bozinovic, F. and Garland, T., Jr (2004a). Climatic adaptation and the evolution of maximum and basal rates of metabolism in rodents. *Evolution* **58**, 1361-1374.

Rezende, E. L., Chappell, M. A. and Hammond, K. A. (2004b). Cold-acclimation in *Peromyscus*: temporal effects and individual variation in maximum metabolism and ventilatory traits. *J. Exp. Biol.* **207**, 295-305.

Rhodes, J. S. and Garland, T., Jr (2003). Differential sensitivity to acute administration of Ritalin, apomorphine, SCH 23390, but not raclopride in mice selectively bred for hyperactive wheel-running behavior. *Psychopharmacol.* **167**, 242-250.

Rhodes, J. S., Koteja, P., Swallow, J. G., Carter, P. A. and Garland, T., Jr (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.

Rhodes, J. S., Hosack, G. R., Girard, I., Kelley, A. E., Mitchell, G. S. and Garland, T., Jr (2001). Differential sensitivity to acute administration of cocaine, GBR 12909, and fluoxetine in mice selectively bred for hyperactive wheel-running behavior. *Psychopharmacol.* **158**, 120-131.

Rhodes, J. S., Garland, T. and Gammie, S. C. (2003). Patterns of brain activity associated with variation in voluntary wheel-running behavior. *Behav. Neurosci.* **117**, 1243-1256.

Rhodes, J. S., Gammie, S. C. and Garland, T., Jr (2005). Neurobiology of mice selected for high voluntary wheel-running activity. *Int. Comp. Biol*. **45**, in press.

Richardson, C. S., Dohm, M. R. and Garland, T., Jr (1994). Metabolism and thermoregulation in crosses between wild and random-bred laboratory house mice (*Mus domesticus*). *Physiol. Zool.* **67**, 944-975.

Roff, D. A. (1997). *Evolutionary Quantitative Genetics*. Montreal: Chapman & Hall.

Sadowska, E. T., Labocha, M. K., Baliga, K., Stanisz, A., Wroblewska, A. K., Jagusiak, W. and Koteja, P. (2005). Genetic correlations between basal and maximum metabolic rates in a wild rodent: consequences for evolution of endothermy. *Evolution* **59**, 672-681.

Schwarz, E. and Schwarz, H. K. (1943). The wild and commercial stocks of the house mouse, *Mus musculus* Linnaeus. *J. Mammal.* **24**, 59-72.

Steudel, K. L. (1990). The work and energetic cost of locomotion. I. The effects of limb mass distribution in quadrupeds. *J. Exp. Biol.* **154**, 273-285.

Swallow, J. G. and Garland, T., Jr (2005). Selection experiments as a tool in evolutionary and comparative physiology: insights into complex traits–An introduction to the symposium. *Int. Comp. Biol.* **45**, in press.

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

Swallow, J. G., Garland, T., Carter, P. A., Zhan, W. Z. and Sieck, G. C. (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., Koteja, P., Carter, P. A. and Garland, T., 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.

Swallow, J. G., Rhodes, J. S. and Garland, T., Jr (2005). Phenotypic and evolutionary plasticity of organ masses in response to voluntary exercise in house mice. *Int. Comp. Biol.* **45**, in press.

Taigen, T. L. and Beuchat, C. A. (1984). Anaerobic threshold of anuran amphibians. *Physiol. Zool.* **57**, 641-647.

Wahlsten, D. (1990). Insensitivity of the analysis of variance to heredity-environment interaction. *Behav. Brain Sci.* **13**, 109-161.

Weibel, E. R., Bacigalupe, L. D., Schmitt, B. and Hoppeler, H. (2004). Allometric scaling of maximal metabolic rate in mammals: muscle aerobic capacity as determinant factor. *Respir. Physiol. Neurobiol.* **140**, 115-132.

Withers, P. C. (1977). Measurements of metabolic rate, VCO_2, and evaporative water loss with a flow through mask. *J. Appl. Physiol.* **42**, 120-123.

Zar, J. H. (1999). *Biostatistical Analysis*, 4th edn. London: Prentice-Hall International.

Zhan, W.-Z, Swallow, J. G., Garland, T., Proctor, D. N., Carter, P. A. and Sieck, G. C. (1999). Effects of genetic selection and voluntary activity on the medial gastrocnemius muscle in house mice. *J. Appl. Physiol.* **87**, 2326-2333.