ORIGINAL RESEARCH

Are Voluntary Wheel Running and Open-Field Behavior Correlated in Mice? Different Answers from Comparative and Artificial Selection Approaches

Vincent Careau · Olaf R. P. Bininda-Emonds · Genesis Ordonez · Theodore Garland Jr.

Abstract Voluntary wheel running and open-field behavior are probably the two most widely used measures of locomotion in laboratory rodents. We tested whether these two behaviors are correlated in mice using two approaches: the phylogenetic comparative method using inbred strains of mice and an ongoing artificial selection experiment on voluntary wheel running. After taking into account the measurement error and phylogenetic relationships among inbred strains, we obtained a significant positive correlation between distance run on wheels and distance moved in the open-field for both sexes. Thigmotaxis was negatively correlated with distance run on wheels in females but not in males. By contrast, mice from four replicate lines bred for high wheel running did not differ in either distance covered or thigmotaxis in the open field as compared with mice from four non-selected control lines. Overall, results obtained in the selection experiment were generally opposite to those observed among inbred strains. Possible reasons for this discrepancy are discussed.

Keywords Anxiety · $Cdh23^{ahl}$ mutation · Comparative method · Experimental evolution · Exploration · Inbred strains

Edited by Stephen Maxson.

V. Careau (✉) · G. Ordonez · T. Garland Jr.
Department of Biology, University of California, Riverside, CA 92521, USA
e-mail: vcareau@ucr.edu

O. R. P. Bininda-Emonds
AG Systematik und Evolutionsbiologie, IBU—Fakultät V, Carl von Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany

Introduction

Two of the most common measures of locomotor behavior in laboratory rodents are voluntary wheel running (Sherwin 1998; Lerman et al. 2002; Garland et al. 2011a) and open-field behavior (Archer 1973; Walsh and Cummins 1976; Gould et al. 2009). In spite of their frequency of study, we do not have a good idea of what either behavior represents at the levels of general locomotor activity or neurobiology.

The open-field test was initially developed as a measure of "emotionality" (Hall 1934), but interpretations of this test diversified over time and it has been viewed as a measure of anxiety, exploratory drive, and even general activity (Archer 1973; Walsh and Cummins 1976; Gould et al. 2009). In comparing different species of rodents, Wilson et al. (1976) suggested that the open-field test should simply be viewed as an analytical tool for assessing individual-specific behavioral reactions when introduced into a novel environment. Clearly, the open-field test does not give a simple index of the status of locomotor output (Stanford 2007; see also Garland et al. 2011b). Behavioral ecologists first used a type of open-field test as a measure of "exploratory activity" and showed that variation in population growth rate among populations of *Microtus ochrogaster* was associated with variation in average latency to enter the testing arena (from an entry tunnel) and the number of fecal pellets deposited in the arena (Krebs 1970). More recently, the open-field and other novel-environment tests have gained popularity in behavioral ecology because it is believed that they may provide insight concerning the behavior of animals in nature, including their space use (Boon et al. 2008; Boyer et al. 2010; van Overveld and Matthysen 2010) and dispersal (Fraser et al. 2001; Dingemanse et al. 2003). Moreover, an individual's behavior in the open-field and other novel-environment

tests, if it represents exploration and/or anxiety, may correlate with its ability to find new food sources and mates under natural conditions, which in turn may affect Darwinian fitness (Dingemanse et al. 2004; Boon et al. 2007).

One of the earliest studies of wheel running (e.g., Stewart 1898) viewed it as a measure of general activity (although the "wheels" were actually rotating drum-type cages). Later studies of activity in wheels housed inside or attached to cages suggested that wheel running may indicate general activity, exploration, migration, stereotypic activity, escape, play, and/or level of deprivation (e.g., food, water, conspecifics; Mather 1981; Sherwin 1998). Owing to the irreconcilable nature of dissimilar findings over a century of research, Sherwin (1998) suggested that "wheel running may be an artifact of captive environments or of the running-wheel itself, possibly resulting from feedback dysfunction." In any case, many species of rodents will readily use running wheels promptly after they become available, and this behavior can occupy a substantial proportion of their time budget (Dewsbury 1980; Mather 1981; Sherwin 1998; Garland et al. 2011b). Although behavioral ecologists have paid little if any attention to wheel running, some workers have speculated that it may reflect aspects of daily movement distance or home range size that would occur in nature (e.g., Swallow et al. 2009; Feder et al. 2010).

Given that both wheel running and open-field behavior involve locomotor activity, it is of obvious interest to ask whether they are related, and a number of previous studies have addressed this question. Dewsbury (1980), Wilson et al. (1976) and Webster et al. (1979) measured both open-field and wheel running in 13 species of muroid rodents and found that the number of squares entered in the open field did not correlate significantly (among species) with either the mean number of revolutions in running wheels or the percentage of revolutions in darkness (see also reanalysis in Bronikowski et al. 2001). Previous artificial selection studies that targeted either voluntary wheel running or open-field behavior did not find correlated responses in the other trait (DeFries et al. 1970; Bronikowski et al. 2001). The reciprocal results from these two selection experiments offer strong evidence that these behavioral traits are not influenced by common genes and neural circuits (see also Zombeck et al. 2011). However, genetic correlations can change across generations during the course of artificial or natural selection, and more recent work using three of the eight lines in the wheel-running selection experiment suggests that some components of open-field behavior may have changed as a result of selective breeding for high wheel running (Jonas et al. 2010). The first objective of the present study was, therefore, to re-evaluate whether open-field behavior changed as a correlated response to artificial selection on wheel

running at a much later generation than studied by Bronikowski et al. (2001).

Another approach for exploring the relationships of traits is by use of the large number of inbred mouse strains currently available (Swallow et al. 2009). Inbred strains of mice are generally created by repeated, successive brother-sister matings that eventually result in >98 % homozygosity at any given locus (Silver 1995), such that individual differences *within* strains are caused by environmental variation whereas the differences *among* strains are caused by the environment, allelic differences, and G × E interactions. Therefore, the general interpretation of observed correlations among strain means for two or more traits is that common genes are influencing their expression (Hegmann and Possidente 1981; Crabbe et al. 1990). However, an overlooked aspect of most among-strain comparisons is the potential violation of the fundamental statistical assumption that the data points under analysis can be considered independent and identically distributed. Given that inbred strains of mice are related through their evolutionary history (Atchley and Fitch 1991; Petkov et al. 2004), comparative analyses should take into account their phylogenetic relationships (Garland et al. 2005; Rhodes et al. 2007). The second objective of the present study was to use the phylogenetically informed statistical method to estimate correlations among strain means for wheel running and open-field behavior. We assembled a comparative dataset from the Mouse Phenome Database (www.jax.org/phenome) and analyzed an available molecular dataset comprising sites with single nucleotide polymorphism (SNPs) under a maximum likelihood framework to develop a phylogeny that captures the degree of molecular divergence among strains, including some wild-derived strains (CAST/EiJ, MOLF/EiJ, SPRET/EiJ). This allowed us to evaluate the extent of phylogenetic signal (Blomberg et al. 2003) in wheel running and open-field behaviors and to test whether the inclusion of phylogenetic information changes the correlation results.

Methods

Mouse inbred strains: dataset on wheel running

The dataset on wheel running originated from a study of exercise physiological genomics conducted in a single laboratory using a solid-surface wheel (circumference = 45.0 cm, diameter = 14.3 cm, and width = 3.5 cm) (Lightfoot et al. 2010). Briefly, wheel running was monitored for 21 consecutive days on mice 9–21 weeks old. Body mass was recorded weekly. The complete dataset comprised a total of 448 individuals of 41 different strains (mean individuals per strain = 11, range 4–64) purchased from

Jackson Laboratories at ~6 week of age, then shipped to the testing location. For each sex, we extracted strain averages across the monitoring period and standard errors (SE) for three aspects of wheel-running behavior (daily distance run, duration of running in the wheel, and average running speed). Although strain-mean correlations between the sexes for each of distance run, duration, and speed are all highly significant (distance: $r = 0.72$, duration: $r = 0.68$, speed: $r = 0.67$; $n = 41$, all $P < 0.0001$), we decided to keep the data separate for males and females because Lightfoot et al. (2010) reported significant sex-by-strain interaction (i.e., in some strains, females run more than males whereas the opposite is true in some others). In some strains included in this study, wheel running was monitored over a period of 7 days (PL/J and SM/J) or a mix of 7- and 21-days (C57BL/6J, CAST/EiJ, and NZB/BINJ), but because Lightfoot et al. (2010) found no difference in the average daily activity measures, we followed them and pooled all the data (see also Turner et al. 2005).

Mouse inbred strains: dataset on open-field behavior

The dataset on open-field behavior originated from a study conducted in two laboratories using a similar apparatus (a 40 L × 40 W × 30 H cm arena) equipped with a video-tracking system (Wahlsten et al. 2006). Briefly, a total of 398 open-field tests were performed on mice 9 weeks old from 21 laboratory strains (average number of individual per strain = 19, range 9–24) purchased from Jackson Laboratories at 4–6 weeks of age, then shipped to the testing location. In any given session, all the mice were transported in their home cage(s) to (and from) the testing room at once on a small rack, where they were kept and acclimated for at least 30 min (up to 1 h) before testing begun. Open-field tests lasted 5 min and were performed once per individual. Body mass was recorded at the end of testing. For each sex, we extracted the average and standard error for three aspects of open-field behavior: distance covered (in cm), thigmotaxis (percent of time spent within 10 cm of the walls), and number of fecal pellets (boli). Similarly to wheel-running behaviors, correlations between the sexes were highly significant for all three variables (distance: $r = 0.92$, thigmotaxis: $r = 0.83$, number of boli: $r = 0.68$; $n = 41$, all $P < 0.001$). Although there were no statistical differences between the sexes in terms of distance covered and thigmotaxis (two-tailed paired t-tests, distance covered: $t_{19} = -1.36$, $P = 0.19$; thigmotaxis: $t_{19} = 0.15$, $P = 0.88$), males excreted more boli than females ($t_{19} = 4.19$, $P < 0.0001$). The strain-specific scores for distance moved in the open-field test produced by Wahlsten et al. (2006) are highly correlated with those more recently reported by Tarantino et al. (2010; 17 strains in common: females, $r = 0.89$, males $r = 0.88$) and Miller et al. (2010; 16 strains in common: males only, $r = 0.88$).

Mouse inbred strains: phylogeny

Our molecular dataset was derived from the haplotype diversity database of the Center of Genome Dynamics of the Jackson Laboratory (http://cgd.jax.org/datasets/). The base dataset comprises the nucleotide information for all known SNP positions in the mouse genome for a variety of different mouse strains and subspecies. From these data, we extracted the sequence information for our 20 focal strains and concatenated it into a single dataset of 581,672 base pairs. In so doing, we retained information regarding the chromosomes from which each data point was obtained. Thereafter, we constructed a phylogeny from this "supermatrix" under a maximum likelihood framework using RAxML v7.2.8 (Stamatakis 2006) under a GTR + GAMMA model of sequence evolution. Analysis used RAxML's rapid bootstrap option (Felsenstein 1985; Stamatakis et al. 2008) with 1,000 replicates, which simultaneously obtains the optimal topology (within the limits of the heuristic search) and the confidence limits on it. The strain SPRET/EiJ was used as the outgroup to root the tree. The results were topologically identical regardless if separate models of evolution were optimized for each partition (= chromosome of origin) or not. Our preferred tree and branch lengths, however, derive from the unpartitioned analysis, given that it is unclear if the chromosomes represent true molecular partitions of the genome, each evolving according to a different evolutionary model. The resulting phylogeny captures the amount of molecular divergence among the 20 strains in common to both datasets (Fig. 1).
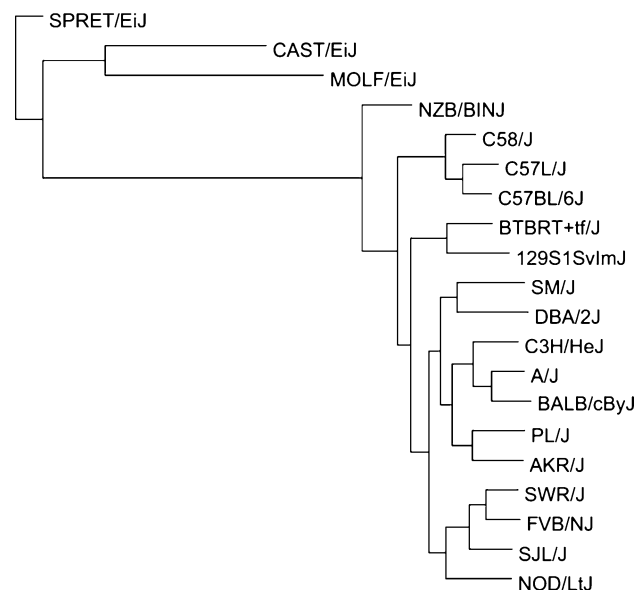

Fig. 1 Phylogeny used for statistical analyses in the comparative study of inbred mouse strains

Mice selectively bred for wheel running

This study was conducted on generation 61 of an ongoing selection experiment on voluntary wheel running (Swallow et al. 1998). The base population was the outbred Hsd:ICR strain, which is albino. The selection experiment includes four replicate lines of mice bred for high levels of voluntary wheel running (high runner or HR lines) and four non-selected control (C) lines (Swallow et al. 1998). Voluntary wheel running is measured on stainless-steel and plexiglas, Wahman-type activity wheels (circumference = 112 cm, diameter = 35.7 cm, and width = 10 cm; Lafayette Instruments, Lafayette, IN). Each line (four HR and four C) is maintained with at least 10 families per generation, routinely housed in same-sex groups of four per cage (except during breeding and wheel-running measurements). Mice are maintained on a 12-h light/12-h dark cycle (lights on 0700), which is also maintained during the wheel-running trials. In C lines, a male and female from within each family are chosen randomly to obtain breeders. In HR lines, breeders are chosen based on the average number of revolutions run on days 5 and 6 of a 6-day test at approximately 6–8 weeks of age. Sib-matings are disallowed in all lines. Although the number of revolutions are monitored continuously over days 5 and 6 (i.e., for 48 h), they occur primarily during the dark phase in both HR and C lines (e.g., see Malisch et al. 2009). Substantial response to selection has occurred; at an apparent selection limit, HR mice typically run 2.5- to 3.0-fold farther than C mice on a daily basis and the increased distance is mainly accomplished by higher running speeds (Garland et al. 2011a and references therein). As a group, the four replicate HR lines show a diverse suite of morphological, biochemical, physiological, and behavioral differences compared to the four non-selected C lines (reviewed in Swallow et al. 2009), including increased predatory aggression in the HR lines (Gammie et al. 2003).

The 154 mice ($\sim$10 of each sex of each line) included in this study had their wheel running measured in four different batches. We excluded two females that appeared as outliers when comparing wheel running on day 5 versus 6. For the open-field experiments, for each batch, 40 mice were chosen on the morning of the 6th day of wheel-exposure, weighed, individually placed in a clean cage with water and food, and immediately transported to the testing room on a small rack, where they were kept and acclimated for at least 4 h before open-field testing began. Although this is a slightly longer acclimation period than used by Wahlsten et al. (2006), we believe this was required to lessen any effect that prior wheel running could have on open-field tests (Burghardt et al. 2004; Duman et al. 2008; Leasure and Jones 2008; Fuss et al. 2010). All open-field tests were done between 1400 and 1830 (i.e., during light phase, sensu Wahlsten et al. 2006; personal communication). The cage of the mouse to be tested was gently taken from the rack and moved onto a counter 1–1.5 m away from the test apparatus. As the cage was carried next to the apparatus, the mouse was gently grasped by the tail and placed in the corner of the test apparatus facing the center. This process was done gently, smoothly, and continuously. A light, continuous background noise was present, and potentially distressing noises from any source were avoided as far as possible. The open-field apparatus was one of those used by Wahlsten et al. (2006)—a 40 × 40 × 30 cm high clear acrylic box with no top, situated within a 50 × 50 × 60 cm testing cubicle. A clean sheet of black Trovicel plastic was placed under the open field before each trial. Dim lighting ($\sim$30 Lux) was maintained by using opaque white filters that covered all four fluorescent light bulbs attached to the cubicle lid. Each testing session lasted 5 min. Fecal boli were counted and the presence of urine recorded. The open-field chamber was prepared for the next mouse by cleaning the lower inside surface (using an enzyme stain and odor remover, Nutri-Vet®) and changing the plastic floor for a clean and dry one.

A camera (Logitech webcam model C500) was situated directly above the center of the apparatus. Videos were analyzed with a computerized tracking system (TopScan Lite V2). We defined 2 zones in the arena: periphery (within 10 cm of the walls) and center (a concentric square of 20 × 20 cm). The tracking system followed the center of the mouse body and its nose. Rearing was counted when the nose of the mouse exited the outer periphery zone, which happened when the mouse put both of its front legs on the wall and stretched its back legs. An experimenter subsequently watched all videos and manually counted rearings in the center when the mouse briefly stopped and both front paws left the ground but did not touch the wall. In a subset of 32 videos, the experimenter also counted rearings along the walls to evaluate the accuracy of video-tracking system in this respect. Although the rearings along the walls counted by the tracking system were slightly lower (mean ± SD = 51.2 ± 11.4, $n = 32$) than manually counted rearings (mean ± SD = 60.1 ± 18.2, $n = 32$), the correlation between the two estimates was highly significant ($r = 0.92$, $n = 32$, $P < 0.0001$). More importantly, the bias (manually counted rearings minus counted by the tracking system) did not differ between HR versus C lines ($t = -0.07$, $P = 0.94$). The resulting dataset includes the following traits: total distance moved (in cm), thigmotaxis (percent of time spent within 10 cm of the walls), number of rearings along the walls and in the center, latency to enter center from the start of the experiment (in s), and number of fecal boli. Latencies for two individuals that never entered the small center zone were scored as 300 s.

Statistical analyses

Phylogenetic signal is the tendency for related species (or populations) to resemble each other, and its presence has implications for understanding how traits evolve and how comparative data are best analyzed in the context of a phylogeny (Blomberg et al. 2003; Revell et al. 2008). The presence of phylogenetic signal was tested as described in Blomberg et al. (2003), with a randomization procedure for the mean-squared error (MSE) using MATLAB program PHYSIG_LL.m. We calculated the amount of phylogenetic signal relative to a Brownian motion expectation of 1.00, as indicated by the K-statistic (Blomberg et al. 2003).

For each sex, we first applied conventional analysis (i.e., ignoring phylogenetic information) to test whether aspects of wheel running (distance, duration, and speed) were correlated with behaviors recorded in the open-field (distance, thigmotaxis, and number of boli). We then used the MECorrPHYSIG.m package in MATLAB to fit estimated generalized least squares (EGLS) models while taking measurement error into account (Ives et al. 2007). The statistical significance of correlations was estimated using parametric bootstrapping to simulate 2,000 new datasets from which new correlation estimates were derived with the distribution of these values being used to calculate approximate P-values (Ives et al. 2007). We ran two sets of EGLS models: a first set assuming a star phylogeny (i.e., no phylogenetic information) and a second set that incorporated phylogenetic information. We also analyzed body mass recorded in both studies for comparison purposes.

Because we performed many statistical tests on closely related data, our Type I error rate may exceed the nominal 5 % alpha level (Benjamini et al. 2001). We stress, however, that we are specifically interested in certain correlations within the correlation matrices; nevertheless, we provide all correlations and significance at different alpha levels (0.05, 0.01, and 0.001) for the sake of completeness and comparison. In the analysis of inbred mouse mice, we present nominal significance levels for all 28 possible correlations, although we are specifically interested in only nine of them (those between all aspects of wheel-running and open-field behavior).

We examined strain descriptions on the Jackson Labs website and searched for any obvious mutations that might influence behavior on the wheel or in an open-field situation. Eight strains (NOD/LtJ, SJL/J, FVB/NJ, SWR/J, AKR/J, PL/J, BALB/cByJ, and A/J) are homozygous for the albinism mutation of the tyrosinase gene (Tyr^c). Six of the 20 strains included in this study (C3H/HeJ, FVB/NJ, MOLF/EiJ, PL/J, SJL/J, SWR/J) are homozygous for the retinal degeneration 1 mutation ($Pde6b^{rd1}$), which causes blindness by weaning age (Sidman and Green 1965; Drager and Hubel 1978). Two strains (A/J and SJL/J) suffer from muscular

dystrophy as a result of mutations of the dysferlin gene ($Dysf^{prmd}$ and $Dysf^{im}$, respectively). Finally, six strains (BALB/cByJ, C57BL/6J, C57L/J, C58/J, DBA/2J, NOD/ShiLtJ) carry the $Cdh23^{ahl}$ mutation, which results in progressive hearing loss that could be severe by just 3 months of age (Johnson et al. 2000). We tested for any effect of these mutations on wheel running and open-field behavior using REGRESSIONv2.m (Lavin et al. 2008). We used a phylogenetic generalized least squares model in which the regression coefficients and the strength of phylogenetic signal in the residuals were estimated (Restricted Maximum Likelihood) simultaneously assuming an Ornstein–Uhlenbeck (OU) evolutionary process along the phylogenetic tree (sensu Swanson and Garland 2009; Gartner et al. 2010). The OU model is often used to model stabilizing selection (Garland et al. 1993; Hansen 1997; Butler and King 2004). Upon finding a significant effect at $P < 0.1$ for any of these mutations for *both* wheel running and open-field behaviour, we re-tested the correlations in subsets of the strains with and without the mutation. We also estimated the partial correlations with the presence of the mutation coded as a dummy variable (0 = absence, 1 = presence). For this, we only present analyses ignoring measurement error and phylogeny, as results from a multiple regression analysis (using MERegPHYSIG) taking both measurement error and phylogeny yielded similar conclusions.

All 156 mice used in the selection experiment had their voluntary wheel running and open-field behavior recorded. For each sex, we analyzed each trait independently using mixed models to test for the effect of selection history or linetype (HR vs. C). Mixed models were performed using ASReml-R (Butler et al. 2007) [analyses yielded the same results as those conducted in SAS Procedure Mixed, as used in our previous studies (e.g., Garland et al. 2011a)]. In these models, we included linetype and appropriate covariates (age in days, z^2-transformed age, and batch) as fixed effects, with line nested within selection history as a random effect (i.e., four HR lines and four C lines). For wheel-running traits, wheel freeness (Garland et al. 2011a) was included as a covariate, whereas time of day was included for open-field traits. P-values for the effect of selection history (i.e., linetype) were estimated with a conditional Wald F-statistic using 1 and 6 for the numerator and denominator degrees of freedom, respectively.

Results

Mouse inbred strains: conventional analyses

Table 1 lists all pairwise strain-mean correlations between body mass, wheel-running, and open-field traits for each sex while ignoring and considering SE into account. In all

Table 1 Pairwise correlations estimated without consideration of measurement error (conventional), with consideration of measurement error (conventional + SE), and with consideration of measurement error and phylogenetic information (phylogenetic + SE) among 20 strain means for voluntary wheel running (distance, duration, and average speed, diameter = 143 mm, over either 7 or 21 days of exposure, data from Lightfoot et al. 2010) and open-field behavior (distance, thigmotaxis, and number of boli, test duration = 5 min, arena = 40 L × 40 W × 30 H cm, data from Wahlsten et al. 2006) for females (above diagonal) and males (below diagonal) separately, using MECorrPhysig (Ives et al. 2007)

Trait	Analysis	Body mass at wheel test	Distance on wheel	Duration on wheel	Speed on wheel	Body mass at OF test	Distance in OF	Thigmotaxis in OF	Boli in OF
Body mass at wheel test	Conventional		−0.243	−0.283	−0.265	0.894***	0.062	−0.438	−0.145
	Conventional + SE		−0.266	−0.300	−0.284	0.923***	0.065	−0.478*	−0.189
	Phylogenetic + SE		−0.228	−0.410	0.114	0.854***	−0.024	−0.197	0.069
Distance on wheel	Conventional	0.038		0.864***	0.645**	−0.254	*0.547**	*−0.460**	*0.161*
	Conventional + SE	0.039		0.979**	0.743*	−0.279	*0.617**	*−0.540**	*0.236*
	Phylogenetic + SE	0.151		0.947**	0.594	−0.245	*0.598**	*−0.734**	*0.511*
Duration on wheel	Conventional	0.071	0.852***		0.260	−0.212	*0.495**	*−0.442*	*0.109*
	Conventional + SE	0.074	0.883***		0.289	−0.227	*0.540**	*−0.500**	*0.148*
	Phylogenetic + SE	0.069	0.856***		−0.120	−0.204	*0.517**	*−0.634**	*0.455*
Speed on wheel	Conventional	−0.104	0.661**	0.237		−0.327	*0.303*	*−0.090*	*0.104*
	Conventional + SE	−0.107	0.685**	0.246		−0.352	*0.336*	*−0.107*	*0.156*
	Phylogenetic + SE	0.096	0.620**	0.061		−0.100	*0.180*	*−0.117*	*−0.072*
Body mass at OF test	Conventional	0.851***	−0.060	−0.049	−0.069		0.155	−0.277	−0.214
	Conventional + SE	0.870***	−0.062	−0.051	−0.071		0.166	−0.306	−0.277
	Phylogenetic + SE	0.819***	0.069	−0.105	0.286		0.233	0.030	−0.103
Distance in OF	Conventional	0.372	*0.706****	*0.453**	*0.597***	0.192		−0.675**	0.009
	Conventional + SE	0.386	*0.734***	*0.473**	*0.622***	0.198		−0.761**	0.020
	Phylogenetic + SE	0.549*	*0.737***	*0.394*	*0.630***	0.399		−0.710**	−0.149
Thigmotaxis in OF	Conventional	−0.456*	*−0.408*	−0.277	−0.264	−0.432	−0.547*		0.033
	Conventional + SE	−0.506*	*−0.454*	−0.309	−0.294	−0.474*	−0.612*		0.031
	Phylogenetic + SE	−0.410	*−0.421*	−0.094	−0.413	−0.315	−0.737*		−0.218
Boli in OF	Conventional	0.101	*0.017*	0.168	*−0.262*	−0.103	0.047	−0.008	
	Conventional + SE	0.118	*0.022*	0.199	*−0.307*	−0.120	0.055	−0.012	
	Phylogenetic + SE	0.236	*0.203*	0.460	*−0.409*	−0.206	0.179	−0.240	

Italicized values of the table indicate the correlations of interest

Conventional analyses assume a star phylogeny whereas phylogenetic analyses used the tree shown in Fig. 1

Significance level is indicated as * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$

cases, incorporation of SE increased the magnitude of the estimated correlation slightly, but in only three case did it lead to a change in statistical significance relative to an α level of 0.05.

As would be expected, the average body masses in the wheel-running and open-field datasets were highly correlated in both sexes (Table 1). Total distance run on wheels was positively related to both duration and speed in both sexes. Similarly, distance covered in the open-field was negatively correlated with thigmotaxis in both males and females.

Distance in the open-field was positively correlated with wheel-running distance in both sexes (Table 1; Fig. 2a, b). In males, distance in the open-field was also correlated with both duration and speed on the wheel, whereas in females it was significantly correlated with duration but not speed (Table 1). Thigmotaxis was significantly negatively correlated with wheel-running distance in females but not males (Table 1; Fig. 2c, d). The three wild-derived strains did not appear to overly contribute strongly to variation or covariation in wheel-running distance and open-field behaviors (distance and thigmotaxis; see grey symbols in Fig. 2).

Mouse inbred strains: phylogenetic analyses

The randomization test indicated relatively high and statistically significant phylogenetic signal in body mass in both datasets, except for males in the wheel-running dataset (Table 2). In general, the randomization test revealed low and non-significant phylogenetic signal in behavioral traits, except for thigmotaxis in females (Table 2).

Fig. 2 Average (±SE) distance run on wheels (km/day) against (**a**, **b**) distance covered in an open-field test (cm, total over 5-min) and (**c**, **d**) thigmotaxis (% of time spent in periphery) among 20 inbred strains of mice for females (*upper panels*) and males (*lower panels*). See Table 1 for estimates of correlation coefficients based on all 20 strains. The three wild-derived strains are shown in *grey*. The six strains carrying the age-related hearing-loss mutation ($Cdh23^{ahl}$) are shown as *solid circles*. See "Results" section for correlation estimates using subsets with and without this and two other mutations
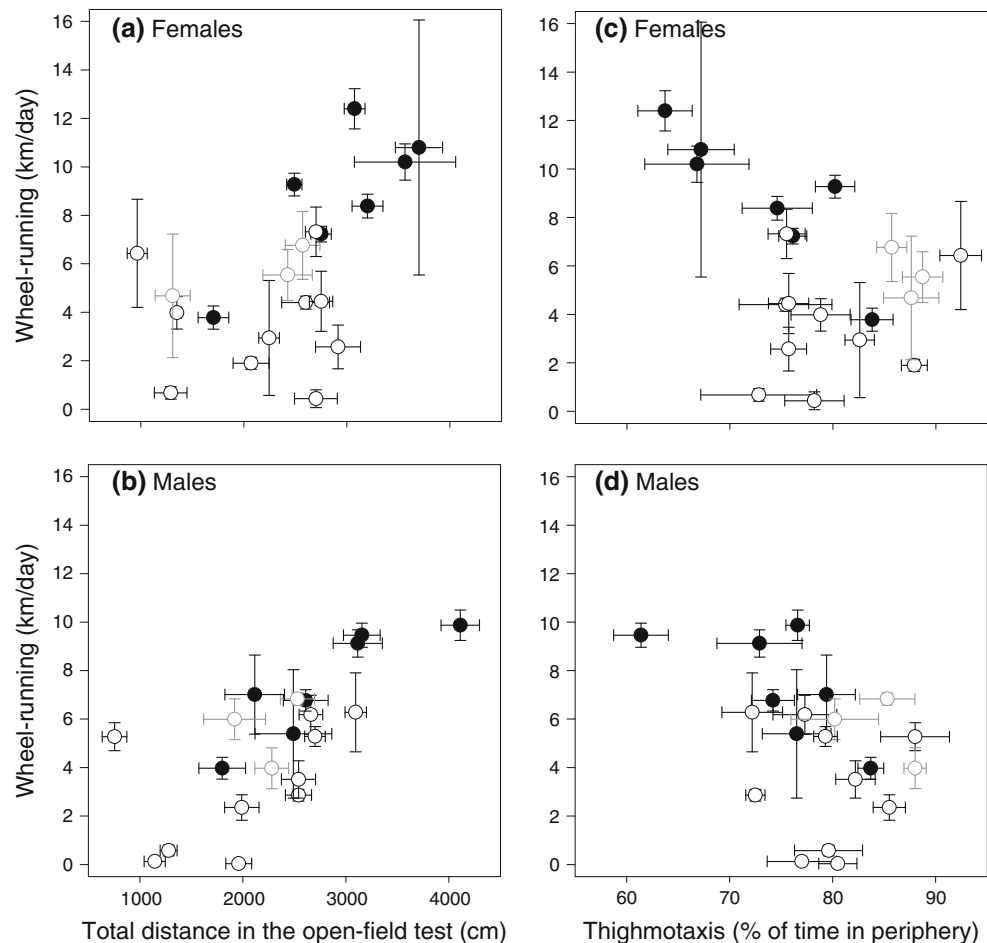

Table 2 Statistics (natural logarithm of maximum likelihood: lnML, and *P*-values) for randomization tests for significance of phylogenetic signal (*K*) for eight measures using the phylogenetic tree shown in Fig. 1

Traits	Females				Males			
	K	lnML star	lnML phylo	*P*	*K*	lnML star	lnML phylo	*P*
M_b at wheel test	**0.845**	**−49.54**	**−45.62**	**<0.001**	0.748	−56.86	−57.26	0.065
Distance on wheel	0.230	−52.05	−52.47	0.064	0.181	−49.13	−51.87	0.353
Duration on wheel	0.203	−125.50	−127.97	0.340	0.138	−124.30	−129.33	0.771
Speed on wheel	0.193	−62.09	−65.13	0.387	0.152	−65.73	−70.42	0.741
M_b at OF test	**0.983**	**−56.75**	**−53.77**	**0.003**	**1.490**	**−58.48**	**−53.16**	**<0.001**
Distance in OF	0.480	−160.65	−162.17	0.161	0.198	−160.98	−163.88	0.416
Thigmotaxis	**0.535**	**−68.81**	**−68.58**	**0.036**	0.218	−64.69	−66.32	0.222
Boli in OF	0.150	−27.38	−32.15	0.750	0.172	−35.61	−40.63	0.774

Bold indicate statistical significance at *P* < 0.05

Table 1 also lists all pairwise correlations between body mass, wheel-running, and open-field traits for each sex while taking measurement error and phylogeny into account. Some correlation coefficients lost or gained significance after the inclusion of phylogenetic information into the analysis. For example, distance and speed on wheel were no longer significantly correlated in females (Table 1). Only one correlation between wheel-running and open-field behaviors changed as a result of the inclusion of phylogeny (the correlation between distance in the open-field and duration on wheel became non-significant in male; Table 1). Finally, the correlations between wheel-running distance and distance in the open-field remained significant in both sexes (Table 1).

Mouse inbred strains: effect of four mutations

In phylogenetic regression models, the albinism mutation positively affected number of boli in the open-field in both males and females, but had no effect on wheel running (Table 3). The blindness mutation had no statistical effect on either voluntary wheel running or open-field behaviour in either sex (Table 3). The dystrophy mutation negatively affected distance in the open-field in both males and females, whereas it positively affected thigmotaxis in females only and negatively affected speed on the wheel in males only (Table 3). The hearing mutation, however, had a significant effect in both sexes on wheel running distance, duration on the wheel, and distance moved in the open-field test (Table 3). This mutation also had marginally non-significant effects on thigmotaxis in both sexes (Table 3). The correlations between wheel-running and distance in the open field are not significant when restricted to the 14 strains without the $Cdh23^{ahl}$ hearing mutation in both females ($r = 0.11$, $P = 0.72$; Fig. 2a) and males ($r = 0.47$, $P = 0.09$; Fig. 2b). However, for the 6 strains carrying $Cdh23^{ahl}$, the correlation is high and marginally non-significant in females ($r = 0.79$, $P = 0.06$; Fig. 2a) and high and significant in males ($r = 0.92$, $P = 0.01$; Fig. 2b). The partial correlation between wheel-running and distance in the open field was non-significant in females ($r = 0.35$, $P = 0.13$), but still significant in males ($r = 0.60$, $P = 0.01$). Similarly, the correlations between wheel-running and thigmotaxis in the open field are not significant when restricted to the 14 strains without $Cdh23^{ahl}$ in both females ($r = 0.25$, $P = 0.39$; Fig. 2c) and males ($r = 0.05$, $P = 0.86$; Fig. 2d), but strong and significant in females ($r = -0.87$, $P = 0.02$; Fig. 2c) and non-significant in males ($r = -0.68$, $P = 0.14$; Fig. 2c) for the six afflicted strains. The partial correlations between wheel-running and thigmotaxis in the open field were non-significant in both females ($r = -0.19$, $P = 0.42$) and males ($r = -0.20$, $P = 0.40$).

Mice from lines selected for wheel running: effect of selection history

Some traits were affected by different covariates included in the full models, such as age, batch, and time of day (see Table S1 in Supplementary material for P-values associated with covariates). In this paper, however, our focus is not on the effect of these covariates, but rather on the potential differences between HR and C lines after having accounted for these sources of phenotypic variation.

The least squares means, F-, and P-values for the main effect of selection history from the full mixed models are reported in Table 4. All components of wheel running were significantly higher in HR than C mice in both sexes

Table 3 Effect of mutations causing retinal degeneration ($Pde6b^{rd1}$), blindness ($Pde6b^{rd1}$), muscular dystrophy ($Dysf^{prmd}$ and $Dysf^{im}$), progressive hearing loss ($Cdh23^{ahl}$), and albinism (Tyr^c) on wheel running and open-field (OF) behavior in 20 inbred mouse strains, using phylogenetic regressions in regressionv2.m (Lavin et al. 2008) with an OU evolutionary process along the phylogenetic tree

Trait/sex	$Pde6b^{rd1}$			$Dysf^{prmd}$ and $Dysf^{im}$			$Cdh23^{ahl}$			Tyr^c		
	Est ± SE	t	P	Est ± SE	t	P	Est ± SE	t	P	Est ± SE	t	P
Females												
Distance on wheel	-1.758 ± 1.639	1.07	0.298	-1.509 ± 2.118	0.71	0.485	**4.496 ± 1.293**	**3.48**	**0.003**	1.515 ± 1.752	0.86	0.399
Duration on wheel	-35.94 ± 67.42	0.53	0.600	-51.53 ± 90.99	0.57	0.578	**168.45 ± 53.23**	**3.16**	**0.005**	9.558 ± 68.551	0.14	0.891
Speed on wheel	-2.061 ± 2.942	0.70	0.493	-3.968 ± 0.643	0.41	0.685	4.469 ± 2.730	1.64	0.119	2.851 ± 2.996	0.95	0.354
Distance in OF	42.2 ± 401.4	0.11	0.917	**-931.6 ± 436.4**	**2.13**	**0.047**	**804.9 ± 353.1**	**2.28**	**0.035**	446.2 ± 400.5	1.11	0.280
Thigmotaxis in OF	-0.607 ± 3.882	0.16	0.877	**10.899 ± 3.509**	**3.11**	**0.006**	**-7.275 ± 3.455**	**2.11**	**0.050**	-0.514 ± 4.218	0.12	0.904
Boli in OF	0.596 ± 0.469	1.27	0.220	1.089 ± 0.702	1.55	0.138	-0.610 ± 0.468	1.30	0.208	**0.875 ± 0.409**	**2.14**	**0.046**
Males												
Distance on wheel	-1.954 ± 1.451	1.35	0.195	-2.384 ± 2.017	1.18	0.253	**3.396 ± 1.214**	**2.80**	**0.012**	0.276 ± 1.518	0.18	0.858
Duration on wheel	-93.40 ± 59.92	1.56	0.137	-36.80 ± 93.71	0.39	0.699	**131.48 ± 54.01**	**2.43**	**0.026**	1.900 ± 61.339	0.03	0.976
Speed on wheel	-2.052 ± 3.340	0.61	0.547	**-11.429 ± 4.311**	**2.65**	**0.016**	3.731 ± 3.234	1.15	0.264	-2.108 ± 3.097	0.68	0.505
Distance in OF	-86.8 ± 406.3	0.21	0.833	**-1168.2 ± 464.9**	**2.51**	**0.022**	**921.2 ± 348.7**	**2.64**	**0.017**	415.9 ± 398.9	1.04	0.311
Thigmotaxis in OF	1.512 ± 3.254	0.46	0.648	6.969 ± 3.768	1.85	0.081	-5.741 ± 2.927	1.96	0.065	-0.941 ± 3.372	0.28	0.783
Boli in OF	0.242 ± 0.736	0.33	0.746	1.548 ± 1.067	1.45	0.164	0.099 ± 0.751	0.13	0.897	**1.461 ± 0.599**	**2.44**	**0.025**

Bold indicate statistical significance at $P < 0.05$

Table 4 Least-square means (LSmean), pooled SE, *F*- and *P*-values in a series of univariate mixed models testing the effect of selection history (linetype: selected HR or control) on components of wheel running (distance, time, average speed, and maximum speed) and open-field behavior (OF; distance, thigmotaxis, number of boli, latency to enter center, and rearings at walls and in the center) in (a) 76 females and (b) 78 males of generation 61 of an ongoing selection experiment on voluntary wheel running (Swallow et al. 1998)

Trait	Units	Females				Males			
		LSmean ± SE		Linetype effect		LSmean ± SE		Linetype effect	
		C	HR	$F_{1,6}$	P	C	HR	$F_{1,6}$	P
Distance on wheel	Rev/day	4,076 ± 435	13,262 ± 438	211.90	<0.0001	3,282 ± 350	12,194 ± 351	293.30	<0.0001
Time on wheel	Min/day	411.3 ± 11.2	527.6 ± 11.3	36.90	0.0009	329.0 ± 22.6	519.8 ± 22.6	35.13	0.0010
Speed on wheel	Rev/min	9.8 ± 1.3	24.8 ± 1.3	65.47	0.0002	9.4 ± 0.6	23.2 ± 0.6	284.00	<0.0001
Max speed on wheel	Rev/min	19.9 ± 1.3	39.2 ± 1.3	104.50	0.0001	19.2 ± 0.9	37.8 ± 0.9	194.60	<0.0001
Distance in OF	Cm	2,539 ± 177	2,258 ± 177	1.26	0.3053	2,519 ± 137	2,133 ± 137	3.95	0.0941
Thigmotaxis	%	280.4 ± 4.0	283.1 ± 4.1	0.22	0.6555	281.1 ± 3.2	282.1 ± 3.2	0.04	0.8402
Number of boli	Pellet	2.1 ± 0.3	1.4 ± 0.3	2.14	0.1937	1.8 ± 0.3	1.3 ± 0.3	1.42	0.2784
Latency center	s	94.2 ± 30.4	124.4 ± 30.5	3.59	0.1069	93.6 ± 20.4	78.8 ± 20.5	0.25	0.6318
Rearing at walls	Number	53.7 ± 3.1	45.3 ± 3.1	0.49	0.5111	53.6 ± 2.7	45.5 ± 2.7	4.39	0.0810
Rearings in center	Number	7.2 ± 1.9	4.3 ± 1.9	1.13	0.3283	7.5 ± 1.8	2.5 ± 1.8	4.03	0.0913

(Table 4). HR females ran 3.3× more revolutions per day, 1.3× longer, 2.6× faster, and reached 2.0× higher maximum speed than control females (Table S2 in Supplementary material). HR males ran 3.9× more revolutions per day, 1.6× longer, 2.5× faster, and reached 2.0× higher maximum speed than control males (Table S2 in Supplementary material). No component of open-field behavior differed significantly between HR versus C mice in both sexes (Table 4). The only marginally non-significant trends detected indicated *less* distance moved and fewer rearings in HR than C mice (Table 4).

Discussion

Our analysis of mean values for 20 inbred strains of mice indicated a *positive* among-strain correlation between distance run on the wheel and distance moved in the open-field in both sexes (Table 1; Fig. 2). This correlation remained statistically significant after controlling for the non-independence potentially introduced by the phylogenetic relationships among strains (Fig. 1). By contrast, the behavior of selectively bred HR and C lines of mice did not differ statistically in the open-field. In fact, the tendencies were in the opposite direction: HR mice tended to show less locomotion and fewer rearings than C mice. Before discussing the potential reasons for this discrepancy in detail, it is appropriate to comment on the effect of phylogeny and place our results within the context of previous work on the HR lines of mice (Bronikowski et al. 2001; Jonas et al. 2010).

Phylogenetic analyses

Blomberg et al. (2003) reported that the mean *K* was 0.83 (95 % confidence interval: 0.63–1.07) for 24 measures of adult body size in studies that considered different species (including plants). Hence, the K values we obtained for mouse body mass (Table 2: range = 0.75–1.49) are typical of what has been found for both inter- and intraspecific datasets (Blomberg et al. 2003; Ashton 2004), including another study on inbred strains of mice (Rhodes et al. 2007). By contrast, of the six behavioral traits analyzed in the present study, only thigmotaxis showed a statistically significant signal, and this was in females only. It must be noted that our power to detect a significant signal may have been relatively good, potentially on the order of 0.8 (see Fig. 2 in Blomberg et al. 2003), assuming that the phylogenetic topology and branch lengths have been estimated without error (see below). Our results support the observation that behavioral traits exhibit significantly lower phylogenetic signal than body size, perhaps because behavior is relatively labile evolutionarily (Blomberg et al. 2003), even as inbred strains diverge.

In the only previous phylogenetic comparative study of inbred mouse strains, the inclusion of phylogenetic information in the analysis did not change the overall results and it was concluded that previous studies may not have been mislead by ignoring the history of strain development (Rhodes et al. 2007). However, Rhodes et al. (2007) noted that their study did not include very divergent, wild-derived inbred strains, and raised the question of whether or not the inclusion of phylogenetic information under such scenarios

Fig. 3 Relationship between estimates of pairwise correlation coefficients for eight traits estimated with and without taking phylogenetic relationships into account, in males and females separately (see Table 1). All correlation coefficients were estimated while taking measurement error into account. *Dashed lines* show equality (1:1)
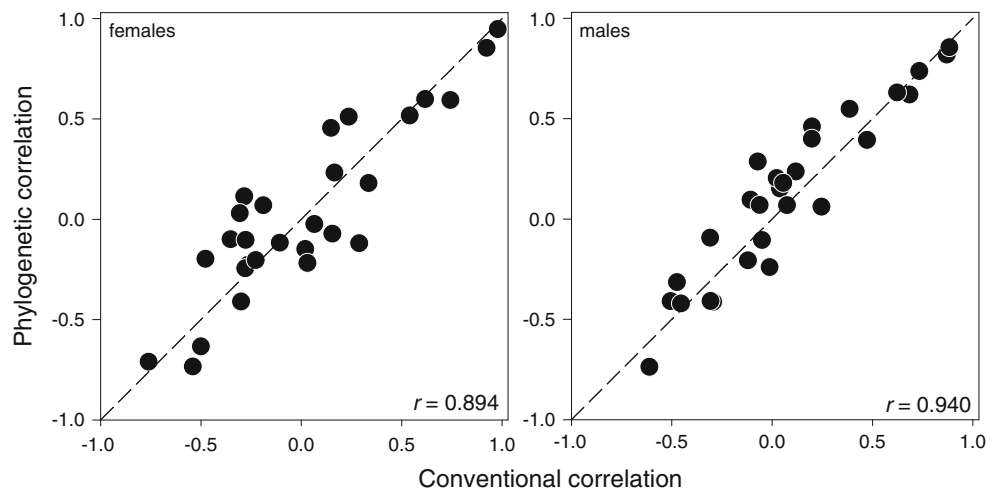
would change statistical outcomes. Our analysis included some phylogenetically very divergent wild-derived inbred strains, and incorporation of phylogenetic information in our analyses did indeed change some results, as some correlations lost statistical significance (e.g., duration on the wheel and distance in the open-field in males), whereas others gained significance (e.g., duration on the wheel and thigmotaxis in females). These changes, however, were relatively small (Table 1; Fig. 3).

Although the origins of most of the inbred strains included in our study are well known (Atchley and Fitch 1991), the early history of some is unclear (e.g., AKR; http://www.informatics.jax.org/), and some strains might have been genetically contaminated by unknown sources (e.g., 129/SvJ, C57BL, C3H; see http://www.informatics. jax.org/). It is therefore likely that our bifurcating tree is an inadequate representation of the reticulate nature of the evolution of these hybrid strains. Similarly, some breeds of dogs were developed by mixing two or more breeds, and a phylogenetic analysis did not reveal significant phylogenetic structure for 76 of 85 breeds (Parker et al. 2004). Thus, it is unsurprising that the inclusion of phylogenetic information did not change results in a dog study examining associations between personality, longevity, and energy expenditure (Careau et al. 2010) (see also discussion in Galis et al. 2007). In any case, future studies of strains and breeds could aim at controlling for both phylogenetic non-independence and gene flow (i.e., hybridization), although this will be challenging (Stone et al. 2011).

Differences with previous work on the high runner lines

Jonas et al. (2010) studied two of the HR lines and one C line. Using a circular open-field, they found that HR line #8 (lab designation) showed more locomotion (in terms of number of rings visited) and less thigmotaxis than C line

#2, and that HR line #7 showed more rearings than C line #2. Because these results were based on only two HR and one C line, their generality needed to be confirmed using all lines of our selection experiment. In contrast to the results of Jonas et al. (2010), we obtained no clear behavioral difference between the four HR and four C lines during the open-field test. Our results are more consistent with those previously obtained by Bronikowski et al. (2001), showing few statistically significant differences between HR and C lines during a 3-min open-field test in a larger arena. In addition to the fact that Jonas et al. (2010) studied only three of eight lines, many differences in our protocols could account for differences between the two studies, including several factors known to affect open-field behavior [arena configuration (circular vs. square), dimensions, housing conditions, prior exposure to wheels, handling, time of day; recently reviewed by Gould et al. 2009].

Genetic correlations and correlated responses to selection

Given that correlations among strain means are generally taken as evidence for genetic correlations (e.g., Hegmann and Possidente 1981; Crabbe et al. 1990), our results for strain correlations would suggest that voluntary wheel running and open-field behavior are positively genetically correlated and hence that these tests are measuring biological constructs with some underlying similarities. According to artificial selection studies conducted so far, however, selective breeding for high voluntary wheel running has not produced correlated changes in open-field behavior (this study; Bronikowski et al. 2001) and vice versa (DeFries et al. 1970). In fact, the differences between HR and C lines we obtained, if any, were in the opposite direction than those obtained by comparing inbred strains. At least four non-mutually exclusive explanations could

account for this discrepancy between comparing inbred strains and selection experiments (see also Klomberg et al. 2002).

First, our wheel-running and/or open-field measures are not identical to those of Lightfoot et al. (2010) and Wahlsten et al. (2006), and both of these behaviors can be highly sensitive to various environmental factors. Notwithstanding differences in the wheel dimensions, we exposed our mice to the wheel over six consecutive days and took the average of days 5 and 6 (as in the routine selective-breeding protocol), whereas Lightfoot et al. (2010) exposed their mice to the wheels over 7 or 21 consecutive days and took the average of all days. To evaluate the extent to which our wheel-running data can be compared with those of Lightfoot et al. (2010), we can use a partial standard provided by data from HR mice and the inbred strain C57BL/6J obtained in a different laboratory using the same type of wheels that we used. Insofar as the results can be compared, Nehrenberg et al. (2009) and Lightfoot et al. (2010) appear to have generated similar data for C57BL/6J in terms of distance run on the wheel in both sexes (Table 5). Differences in speed and time spent on the wheel are probably related to differences in wheel dimensions, with Lightfoot et al. (2010) having used smaller wheels. Given that Nehrenberg et al. (2009) produced wheel-running data that are similar to ours in HR mice (considering specifically our data for HR line #8, as they used), we can indirectly conclude that wheel-running distance is qualitatively comparable to what Lightfoot et al. (2010) measured for inbred strains. The only way to be entirely sure, however, would be to raise several inbred strains and measure wheel running over 6 days in our laboratory at the same time that we test the HR and C lines (see also Crabbe et al. 1999; Wahlsten et al. 2006).

In light of behaviour in the open-field test being highly sensitive to various factors (see Gould et al. 2009), we made efforts to match our protocol with that used by Wahlsten et al. (2006): we used the same apparatus (shipped to us by Dr. Wahlsten) and tested for the same length of time (5 min) and during the photophase. The first difference was the slightly different acclimation period, as Wahlsten et al. (2006) tested mice 30–60 min after transfer to testing room whereas we let mice rest for at least 4 h to reduce potential effects of prior wheel running on open-field tests (see below). The only other difference was the floor, as we used a black sheet of Trovicel plastic whereas Wahlsten et al. (2006) used pink butcher paper. Although it has been shown that various floor textures (soil, bedding, Astroturf, and metal) can affect open-field behavior, they do not appear to affect the rank order of the different individuals/groups measured (Dixon and Van Mayeda 1976); this result, however, has not been evaluated for the black Trovicel versus pink butcher paper contrast. The average distance moved by our female and male HR and C mice (2,564 and 2,484 cm, respectively) is close to the average across the inbred strains (2,418 and 2,320 cm, respectively), which suggests that our protocol captured the same trait as measured by Wahlsten et al. (2006).

A second potential reason for the discrepancy between comparing inbred strains and selection experiments is that we tested HR and C mice on the day following their 6th night of wheel-running, whereas the mice tested by Wahlsten et al. (2006) were never exposed to a wheel nor had the opportunity to exercise at high intensity. It has been shown that intense physical activity subsequently alters behavior in the open-field, but the results are inconsistent, ranging from reduced (Duman et al. 2008; Fuss et al. 2010), unaffected (Leasure and Jones 2008), to increased locomotion (Burghardt et al. 2004) in rodents after voluntary wheel running. Several environmental factors might have contributed to differences among those studies (e.g., the level of physical activity and housing conditions). Duman et al. (2008) speculated that reduced locomotion in the open-field that was possibly related to immediate fatigue after running (see also Fuss et al. 2010). This seems unlikely to us, however, as even our HR mice do not run voluntarily on wheels at speeds that exceed their maximal aerobic capacity (Girard et al. 2001; Rezende et al. 2005), nor do they show obvious signs of fatigue during or following wheel running (Meek et al. 2009; personal observations). Alternatively, voluntary wheel running induces a significant elevation in circulating corticosterone in both HR and C mice (Girard and Garland 2002) and in C57BL/6N mice (Droste et al. 2003). This elevation could have a carry-over effect and alter open-field behavior that was measured some hours later. Therefore, it is possible that a positive genetic correlation between voluntary wheel-running and open-field behavior could potentially be masked by a negative environmental correlation when mice are given the opportunity to exercise at high intensity shortly before the open-field test. For example, although there was no significant phenotypic correlation between distance moved in a novel environment and resting metabolic rate in deer mice (*Peromyscus maniculatus*), a quantitative genetic analysis revealed a positive genetic correlation between these traits (Careau et al. 2011). In the current study, we obtained negative correlations between the intensity of wheel running (average speed) and distance moved in the open-field at the levels of replicate lines in females ($r = -0.798$; $N = 8$; $P = 0.0176$; Table S3 in Supplementary material) and at the level of individuals in males ($r = -0.309$; $N = 78$; $P = 0.006$; Table S3 in Supplementary material), suggesting that environmental factors might affect these traits (see Table S3 in Supplementary material for all possible correlations between all pair-wise correlations estimated at the level of replicate line means and

Table 5 Average estimates for voluntary wheel-running traits (distance, duration, and speed; females and males separately) recorded in Nehrenberg et al. (2009), Lightfoot et al. (2010), and in this study for (a) C57BL/6J and (b) mice selectively bred for high voluntary wheel running (HR line #8)

	Females		Males	
	Nehrenberg	Lightfoot	Nehrenberg	Lightfoot
(a) C57BL/6J				
Distance (km/day)	7.85	8.38	6.63	6.77
Duration (min/day)	530	258	463	213
Speed (m/min)	14.6	32.1	13.8	31.4
	Females		Males	
	Nehrenberg	This study	Nehrenberg	This study
(b) HR line #8				
Distance (km/day)	15.06	14.42	13.53	14.10
Duration (min/day)	645	548	565	556
Speed (m/min)	23.9	26.2	23.9	25.2

individuals for components of wheel running and open-field behavior).

A third possible reason for the discrepancy is that, just as for heritability, genetic correlations can vary among populations (Visscher et al. 2008) and change over the course of an artificial selection study (e.g., Bult and Lynch 2000). Therefore, two reciprocal artificial selection experiments can yield somewhat different correlated responses. Indeed, correlated responses to selection not only depend on the genetic and phenotypic variance and covariance between the traits, but may also be influenced by population size, selection intensity, number of loci influencing the traits, allele frequencies, linkage disequilibrium, and pleiotropy (Falconer and Mackay 1996; Roff 1997). In fact, an asymmetrical correlated response to selection is a common evolutionary outcome for physiologically complex traits (Bohren et al. 1966; Shiotsugu et al. 1997; Czesak et al. 2006 and references therein). For example, a recent selection experiment on home-cage activity in Hsd:ICR mice did not obtain a selection response (Zombeck et al. 2011), whereas our selection experiment for wheel running (also using a base population of Hsd:ICR mice) produced a strong response in wheel running and a correlated response in home-cage activity (when housed without wheels: Rhodes et al. 2005; Malisch et al. 2009). As outlined by Crabbe et al. (1990), the same reasons as to why two reciprocal selection studies may produce different correlated responses can also explain why a genetic correlation suggested by comparing strain means may not be supported by artificial selection studies.

Finally, it may be that some inbred strains have experienced fixation of (deleterious) recessive alleles that simultaneously decreased (or increased) their performance in both wheel-running and open-field tests. Interestingly, the strains carrying the mutation for progressive hearing loss ($Cdh23^{ahl}$) ran *more* distance and spent more time on wheels and covered more distance and showed less thigmotaxis in the open-field test (Fig. 2). After we removed the strains carrying the $Cdh23^{ahl}$ mutation from the analysis, there was no statistically significant correlation remaining between wheel running and open-field behaviour and the partial correlations (with the presence of mutation as a covariate) were lower and/or not statistically significant. It remains unclear as to why hearing-impaired mice run more and are more active in the open-field test, but perhaps it is related to a reduced sensitivity to environmental disturbances. Many other inbreeding-related pathologies occur in inbred strains that may or may not be related to their low scores in one or both behavioral tests, such as muscular dystrophy (see Table 3), mammary tumors, lung tumors, hemolytic anemia, reticulum cell sarcomas, etc. Beyond these possibilities, our observation supports the contention that care must be taken when interpreting correlations across inbred strains. Rose (1984) reached a similar conclusion upon finding that inbreeding can produce artifactual positive correlations among fitness components in *Drosophila melanogaster*.

Conclusion

We have considered four non-mutually exclusive explanations for the discrepancy obtained by comparing inbred strains with the results of artificial selection experiments (see also Klomberg et al. 2002): different traits measured in different studies, effect of prior wheel-running exercise on subsequent behavior in the open-field, population-specificity of genetic correlations, and effects of fixation of

(deleterious) recessive mutations on behaviour. Although we cannot confidently rule out any of these scenarios, we consider that the last scenario is the most likely, as suggested by our analysis (see Fig. 2).

Stanford (2007) stated that the "only reliable way to measure animals' spontaneous locomotor activity is to record their movements while they are in their home cage". We concur with Stanford (2007) that behavior measured in a standard (novel environment) open-field test should not be equated with general locomotor activity (see also Zombeck et al. 2011). Interestingly, when wheels are not provided, HR mice are more active in their home cages than C mice (Rhodes et al. 2005; Malisch et al. 2009), but this difference is nonexistent when wheels are accessible (unpublished results). These results suggests that, at least in our population, home-cage activity and voluntary wheel running are positively genetically correlated, but in a way that depends on environmental context, and that our wheel-running protocol measures a biological construct that shares some underlying similarities with home-cage activity. In any case, open-field behavior does not differ between HR and C mice, indicating no apparent relation with either wheel running or (when wheels are absent) home-cage activity. Taken together, results on home-cage activity, wheel running, and open-field behavior suggest that it is simply erroneous to use the open-field test as a proxy for measuring general locomotor activity (e.g., as in Kirsten et al. 2010; Padilla et al. 2010). Yet, it may also be erroneous to assume that voluntary wheel running is a measure of general locomotor activity, especially given that laboratory mice are routinely housed without wheel access. Moreover, the relation of wheel running to home-cage activity may well depend on the size and configuration of the wheel and the home cage. In any case, it is crucial to maintain clear distinctions among different measures of locomotor behavior in addition to whether they occur in familiar versus novel environments (Stanford 2007; Garland et al. 2011b).

References

Archer J (1973) Test for emotionality in rats and mice: a review. Anim Behav 21:205–235

Ashton KG (2004) Comparing phylogenetic signal in intraspecific and interspecific body size datasets. J Evol Biol 17:1157–1161

Atchley WR, Fitch WM (1991) Gene trees and the origins of inbred strains of mice. Science 254:554–558

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125:279–284

Blomberg SP, Garland T Jr, Ives AR (2003) Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717–745

Bohren BB, Hill WG, Robertso A (1966) Some observations on asymmetrical correlated responses to selection. Genet Res 7:44–57

Boon AK, Réale D, Boutin S (2007) The interaction between personality, offspring fitness, and food abundance in North American red squirrel. Ecol Lett 10:1094–1104

Boon AK, Réale D, Boutin S (2008) Personality, habitat use, and their consequences for survival in North American red squirrels *Tamiasciurus hudsonicus*. Oikos 117:1321–1328

Boyer N, Réale D, Marmet J, Pisanu B, Chapuis JL (2010) Personality, space use and tick load in an introduced population of Siberian chipmunks *Tamias sibiricus*. J Anim Ecol 79:538–547

Bronikowski AM, Carter PA, Swallow JG, Girard IA, Rhodes JS, Garland T Jr (2001) Open-field behavior of house mice selectively bred for high voluntary wheel-running. Behav Genet 31:309–316

Bult A, Lynch CB (2000) Breaking through artificial selection limits of an adaptive behavior in mice and the consequences for correlated responses. Behav Genet 30:193–206

Burghardt PR, Fulk LJ, Hand GA, Wilson MA (2004) The effects of chronic treadmill and wheel running on behavior in rats. Brain Res 1019:84–96

Butler MA, King AA (2004) Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am Nat 164:683–695

Butler D, Cullis BR, Gilmour AR, Gogel DJ (2007) ASReml-R reference manual Release 2.0. VSN International Ltd., Hemel Hempstead

Careau V, Réale D, Humphries MM, Thomas D (2010) The pace of life under artificial selection: personality, energy expenditure and longevity are correlated in domestic dogs. Am Nat 175:753–758

Careau V, Thomas D, Pelletier F, Turki L, Landry F, Garant D, Réale D (2011) Genetic correlation between resting metabolic rate and exploratory behaviour environment in deer mice (*Peromyscus maniculatus*). J Evol Biol 24:2153–2163

Crabbe JC, Phillips TJ, Kosobud A, Belknap JK (1990) Estimation of genetic correlation: interpretation of experiments using selectively bred and inbred animals. Alcoholism 14:141–151

Crabbe JC, Wahlsten D, Dudek BC (1999) Genetics of mouse behavior: interactions with laboratory environment. Science 284:1670–1672

Czesak ME, Fox CW, Wolf JB (2006) Experimental evolution of phenotypic plasticity: how predictive are cross-environment genetic correlations? Am Nat 168:323–335

DeFries JC, Wilson JR, McClearn GE (1970) Open-field behavior in mice: selection response and situational generality. Behav Genet 1:195–211

Dewsbury DA (1980) Wheel-running behavior in 12 species of muroid rodents. Behav Proces 5:271–280

Dingemanse NJ, Both C, van Noordwijk AJ, Rutten AL, Drent PJ (2003) Natal dispersal and personalities in great tits (*Parus major*). Proc R Soc Lond B 270:741–747

Dingemanse NJ, Both C, Drent PJ, Tinbergen JM (2004) Fitness consequences of avian personalities in a fluctuating environment. Proc R Soc Lond B 271:847–852

Dixon LK, Van Mayeda D (1976) Effects of floor textures on open-field behavior in selected lines of mice. Behav Genet 6:87–92

Drager UC, Hubel DH (1978) Studies of visual function and its decay in mice with hereditary retinal degeneration. J Comp Neurol 180:85–114

Droste SK, Gesing A, Ulbricht S, Muller MB, Linthorst ACE, Reul JMHM (2003) Effects of long-term voluntary exercise on the mouse hypothalamic-pituitary-adrenocortical axis. Endocrinology 144:3012–3023

Duman CH, Schlesinger L, Russell DS, Duman RS (2008) Voluntary exercise produces antidepressant and anxiolytic behavioral effects in mice. Brain Res 1199:148–158

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, Harlow

Feder ME, Garland T Jr, Marden JH, Zera AJ (2010) Locomotion in response to shifting climate zone: not so fast. Annu Rev Physiol 72:167–190

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fraser DF, Gilliam JF, Daley MJ, Le An N, Garrick TS (2001) Explaining leptokurtic movement distributions: intrapopulation variation in boldness and exploration. Am Nat 158:124–135

Fuss J, Ben Abdallah NMB, Vogt MA, Touma C, Pacifici PG, Palme R, Witzemann V, Hellweg R, Gass P (2010) Voluntary exercise induces anxiety-like behavior in adult C57BL/6J mice correlating with hippocampal neurogenesis. Hippocampus 20:364–376

Galis F, Van Der Sluijs I, Van Dooren TJM, Metz JAJ, Nussbaumer M (2007) Do large dogs die young? J Exp Zool (Mol Dev Evol) 308:119–126

Gammie SC, Hasen NS, Rhodes JS, Girard I, Garland T Jr (2003) Predatory aggression, but not maternal or intermale aggression, is associated with high voluntary wheel-running behavior in mice. Horm Behav 44:209–221

Garland T Jr, Dickerman AW, Janis CM, Jones JA (1993) Phylogenetic analysis of covariance by computer-simulation. Syst Biol 42:265–292

Garland T Jr, Bennett AF, Rezende EL (2005) Phylogenetic approaches in comparative physiology. J Exp Biol 208:3015–3035

Garland T Jr, Kelly SA, Malisch JL, Kolb EM, Hannon RM, Keeney BK, Van Cleave SL, Middleton KM (2011a) How to run far: multiple solutions and sex-specific responses to selective breeding for high voluntary activity levels. Proc R Soc Lond B 278:574–581

Garland T Jr, Schutz H, Chappell MA, Keeney BK, Meek TH, Copes LE, Acosta W, Drenowatz C, Maciel RC, van Dijk G, Kotz CM, Eisenmann JC (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

Gartner GEA, Hicks JW, Manzani PR, Andrade DV, Abe AS, Wang T, Secor SM, Garland T Jr (2010) Phylogeny, ecology, and heart position in snakes. Physiol Biochem Zool 83:43–54

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

Gould TD, Dao DT, Kovacsics CE (2009) The open field test. In: Gould TD (ed) Mood and anxiety related phenotypes in mice: characterization using behavioral tests. Humana Press, New York, pp 1–20

Hall CS (1934) Emotional behavior in the rat: defecation and urination as measures of individual differences in emotionality. J Comp Psychol 18:385–403

Hansen TF (1997) Stabilizing selection and the comparative analysis of adaptation. Evolution 51:1341–1351

Hegmann JP, Possidente B (1981) Estimating genetic correlations from inbred strains. Behav Genet 11:103–114

Ives AR, Midford PE, Garland T Jr (2007) Within-species variation and measurement error in phylogenetic comparative methods. Syst Biol 56:252–270

Johnson KR, Zheng QY, Erway LC (2000) A major gene affecting age-related hearing loss is common to at least ten inbred strains of mice. Genomics 70:171–180

Jonas I, Schubert KA, Reijne F, Scholte J, Garland T Jr, Gerkema MP, Scheurink AJW, Nyakas C, van Dijk G (2010) Behavioral traits are affected by selective breeding for increased wheel-running behavior in mice. Behav Genet 40:542–550

Kirsten TB, Taricano M, Florio JC, Palermo-Neto J, Bernardi MM (2010) Prenatal lipopolysaccharide reduces motor activity after an immune challenge in adult male offspring. Behav Brain Res 211:77–82

Klomberg KF, Garland T Jr, Swallow JG, Carter PA (2002) Dominance, plasma testosterone levels, and testis size in house mice artificially selected for high activity levels. Physiol Behav 77:27–38

Krebs CJ (1970) Microtus population biology: behavioral changes associated with the population cycle in M. ochrogaster and M. pennsylvanicus. Ecology 51:34–52

Lavin SR, Karasov WH, Ives AR, Middleton KM, Garland T Jr (2008) Morphometrics of the avian small intestine, compared with non-flying mammals: a phylogenetic perspective. Physiol Biochem Zool 81:526–550

Leasure JL, Jones M (2008) Forced and voluntary exercise differentially affect brain and behavior. Neuroscience 156:456–465

Lerman I, Harrison BC, Freeman K, Hewett TE, Allen DL, Robbins J, Leinwand LA (2002) Genetic variability in forced and voluntary endurance exercise performance in seven inbred mouse strains. J Appl Physiol 92:2245–2255

Lightfoot JT, Leamy L, Pomp D, Turner MJ, Fodor AA, Knab A, Bowen RS, Ferguson D, Moore-Harrison T, Hamilton A (2010) Strain screen and haplotype association mapping of wheel running in inbred mouse strains. J Appl Physiol 109:623–634

Malisch JL, Breuner CW, Kolb EM, Wada H, Hannon RM, Chappell MA, Middleton KM, Garland T Jr (2009) Behavioral despair and home-cage activity in mice with chronically elevated baseline corticosterone concentrations. Behav Genet 39:192–201

Mather JG (1981) Wheel-running activity: a new interpretation. Mammal Rev 11:41–51

Meek TH, Lonquich BP, Hannon RM, Garland T Jr (2009) Endurance capacity of mice selectively bred for high voluntary wheel running. J Exp Biol 212:2908–2917

Miller BH, Schultz LE, Gulati A, Su AI, Pletcher MT (2010) Phenotypic characterization of a genetically diverse panel of mice for behavioral despair and anxiety. PLoS ONE 5:e14458

Nehrenberg DL, Hua K, Estrada-Smith D, Garland T Jr, Pomp D (2009) Voluntary exercise and its effects on body composition depend on genetic selection history. Obesity 17:1402–1409

Padilla E, Shumake J, Barrett DW, Holmes G, Sheridan EC, Gonzalez-Lima F (2010) Novelty-evoked activity in open field predicts susceptibility to helpless behavior. Physiol Behav 101:746–754

Parker HG, Kim LV, Sutter NB, Carlson S, Lorentzen TD, Malek TB, Johnson GS, DeFrance HB, Ostrander EA, Kruglyak L (2004) Genetic structure of the purebred domestic dog. Science 304:1160–1164

Petkov PM, Ding Y, Cassell MA, Zhang W, Wagner G, Sargent EE, Asquith S, Crew V, Johnson KA, Robinson P, Scott VE, Wiles MV (2004) An efficient SNP system for mouse genome scanning and elucidating strain relationships. Genome Res 14:1806–1811

Revell LJ, Harmon LJ, Collar DC (2008) Phylogenetic signal, evolutionary process, and rate. Syst Biol 57:591–601

Rezende EL, Chappell MA, Gomes FR, Malisch JL, Garland T Jr (2005) Maximal metabolic rates during voluntary exercise, forced exercise, and cold exposure in house mice selectively bred for high wheel-running. J Exp Biol 208:2447–2458

Rhodes JS, Gammie SC, Garland T (2005) Neurobiology of mice selected for high voluntary wheel-running activity. Integr Comp Biol 45:438–455

Rhodes JS, Ford MM, Yu CH, Brown LL, Finn DA, Garland T Jr, Crabbe JC (2007) Mouse inbred strain differences in ethanol drinking to intoxication. Genes Brain Behav 6:1–18

Roff DA (1997) Evolutionary quantitative genetics. Chapman and Hall, New York

Rose MR (1984) Genetic covariation in *Drosophila* life history: untangling the data. Am Nat 123:565–569

Sherwin CM (1998) Voluntary wheel running: a review and novel interpretation. Anim Behav 56:11–27

Shiotsugu J, Leroi AM, Yashiro H, Rose MR, Mueller LD (1997) The symmetry of correlated selection responses in adaptive evolution: an experimental study using *Drosophila*. Evolution 51:163–172

Sidman RL, Green MC (1965) Retinal degeneration in the mouse. Location of the *rd* locus in linkage group XVII. J Hered 56:23–29

Silver LM (1995) Mouse genetics: concepts and applications. Oxford University Press, New York

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. Syst Biol 57:758–771

Stanford SC (2007) The open field test: reinventing the wheel. J Psychopharm 21:134–135

Stewart CC (1898) Variations in daily activity produced by barometric pressure and diet, with a description alcohol and by changes in of recording methods. Am J Physiol 1:40–56

Stone GN, Nee S, Felsenstein J (2011) Controlling for non-independence in comparative analysis of patterns across populations within species. Philos Trans R Soc B 366:1410–1424

Swallow JG, Carter PA, Garland T Jr (1998) Artificial selection for increased wheel-running behavior in house mice. Behav Genet 28:227–237

Swallow JG, Hayes JP, Koteja P, Garland T Jr (2009) Selection experiments and experimental evolution of performance and physiology. In: Garland T, Rose MR (eds) Experimental evolution: concepts, methods, and applications of selection experiments. University of California Press, Berkeley, pp 301–351

Swanson DL, Garland T Jr (2009) The evolution of high summit metabolism and cold tolerance in birds and its impact on present-day distributions. Evolution 63:184–194

Tarantino LM, Eisener-Dorman AF, Grabowski-Boase L, Steffy BM, Wiltshire T (2010) Quantitative trait locus and haplotype mapping in closely related inbred strains identifies a locus for open field behavior. Mamm Genome 21:231–246

Turner MJ, Kleeberger SR, Lightfoot JT (2005) Influence of genetic background on daily running-wheel activity differs with aging. Physiol Genomics 22:76–85

van Overveld T, Matthysen E (2010) Personality predicts spatial responses to food manipulations in free ranging Great Tits (*Parus major*). Biol Lett 6:187–190

Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era—concepts and misconceptions. Nat Rev Genet 9:255–266

Wahlsten D, Bachmanov A, Finn DA, Crabbe JC (2006) Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. Proc Natl Acad Sci USA 103:16364–16369

Walsh RN, Cummins RA (1976) The open-field test: a critical review. Psychol Bull 83:482–504

Webster DG, Baumgardner DJ, Dewsbury DA (1979) Open-field behavior in eight taxa of muroid rodents. Bull Psychonom Soc 13:90–92

Wilson RC, Thomas V, Lanier DL, Dewsbury DA (1976) Open-field behavior in Muroid rodents. Behav Biol 17:495–506

Zombeck JA, DeYoung EK, Brzezinska WJ, Rhodes JS (2011) Selective breeding for increased home cage physical activity in collaborative cross and Hsd:ICR mice. Behav Genet 41:571–582