doi:10.1111/evo.12200

LIMITS TO BEHAVIORAL EVOLUTION: THE QUANTITATIVE GENETICS OF A COMPLEX TRAIT UNDER DIRECTIONAL SELECTION

Vincent Careau,[1,2] Matthew E. Wolak,[1] Patrick A. Carter,[3] and Theodore Garland Jr.[1,4]

[1]*Department of Biology, University of California, Riverside, California 92521*

[2]*Current address: Centre for Integrative Ecology, Deakin University, 75 Pigdons Road, Waurn Ponds, Victoria 3216, Australia*

[3]*School of Biological Sciences, Washington State University, Pullman, Washington 99164*

[4]*E-mail: tgarland@ucr.edu*

Received November 30, 2012
Accepted May 31, 2013
Data Archived: Dryad doi:10.5061/dryad.37b28

Replicated selection experiments provide a powerful way to study how "multiple adaptive solutions" may lead to differences in the quantitative–genetic architecture of selected traits and whether this may translate into differences in the timing at which evolutionary limits are reached. We analyze data from 31 generations (n = 17,988) of selection on voluntary wheel running in house mice. The rate of initial response, timing of selection limit, and height of the plateau varied significantly between sexes and among the four selected lines. Analyses of litter size and realized selection differentials seem to rule out counterposing natural selection as a cause of the selection limits. Animal-model analyses showed that although the additive genetic variance was significantly lower in selected than control lines, both before and after the limits, the decrease was not sufficient to explain the limits. Moreover, directional selection promoted a negative covariance between additive and maternal genetic variance over the first 10 generations. These results stress the importance of replication in selection studies of higher-level traits and highlight the fact that long-term predictions of response to selection are not necessarily expected to be linear because of the variable effects of selection on additive genetic variance and maternal effects.

KEY WORDS: **Animal model, Bulmer effect, experimental evolution, heritability, selection limit, voluntary exercise.**

Introduction

Artificial selection has been practiced for centuries, originally during domestication to increase expression of desirable traits, and more recently as a way to understand processes underlying evolutionary changes at both phenotypic and genetic levels (Garland and Rose 2009). In artificial selection, in contrast with studies of evolution in natural populations, the target and intensity of selection are known relatively precisely, so results can suggest what would happen if the same type of selection was applied to a similar population in nature (Flux and Flux 1982; Hill and Caballero 1992; Postma et al. 2007). As it turns out, however, replicate lines experiencing apparently identical directional selection often respond at different rates and with different correlated traits (Mayr 1961). Therefore, well-replicated artificial selection experiments can be useful for studying the consistency of the response to selection and the frequency of "multiple adaptive solutions" (Garland 2003; Garland et al. 2011).

Selection experiments also provide a powerful way to explore possible limits to evolutionary change (Kirkpatrick 2009). Although empirical evidence suggests that most traits are capable of responding to directional artificial selection (but see Coyne 1987, for example), it also suggests that limits to phenotypic change are commonly reached in selection experiments (Falconer and Mackay 1996; Barton and Partridge 2000; Garland and Rose

INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

2009; but see Weber 1996). One possible cause of these "selection limits" is a decrease in additive genetic variance (V_A) and hence narrow-sense heritability (h^2). If the trait under selection is determined by a finite number of genes (or if selection is strong and/or populations are small; Weber 1996), then an evolutionary limit may occur because of fixation of favorable alleles in combination with loss of alleles caused by random genetic drift (Falconer and Mackay 1996). However, the prediction that V_A should decline or even be eliminated in the face of (strong) continued directional selection turns out to be less reliable than intuition might suggest (Bulmer 1971; Barton and Turelli 1987; Heath et al. 1995; Blows and Higgie 2003). Nevertheless, Blows and Hoffman (2005) point out that many critical experiments still need to be conducted to elucidate the importance of an elimination of V_A in limiting evolutionary change.

The few long-term selection experiments that have been designed and analyzed specifically to determine how V_A evolves under directional selection have not yielded consistent results (Hine et al. 2011). In studies using laboratory house mice, directional selection was shown to induce variable changes in V_A, ranging from a 65% to 83% decrease (Meyer and Hill 1991; Beniwal et al. 1992; but see Hill and Mbaga 1998) to no change (Martinez et al. 2000) and even a 99% increase (Heath et al. 1995). Although phenotypically, the selected lines diverged substantially from the control lines in all of these studies, none had clearly reached a plateau by the generation(s) for which V_A was estimated, and as such, they tell us little about whether a lack of V_A may ultimately cause a selection limit. Furthermore, all but one targeted morphological traits (e.g., body mass, lean mass), which generally are less closely associated with Darwinian fitness than higher-level traits (e.g., life-history and behavioral traits; Careau and Garland 2012), and generally have higher h^2 (Mousseau and Roff 1987; Kruuk et al. 2000; Walsh and Blows 2009). Hence, it would be informative to test whether a lack of V_A is responsible for an apparent limit to evolutionary change in replicate lines selected for a higher-level trait (Hine et al. 2011).

Counterposing natural selection that limits trait evolution during artificial selection is well documented in a number of cases (Lerner and Dempster 1951; Falconer 1955; Roberts 1966; Hine et al. 2011). For example, Falconer (1955) showed that counterposing natural selection impeded the response to bidirectional selection on body mass after ~22 generations; specifically, the proportion of fertile matings and postnatal survival declined in both high and low lines, and mothering ability and litter size declined in the low line only (see also Roberts 1966). Ultimately, counterposing natural selection occurs as a consequence of adverse pleiotropic effects (e.g., negative COV_A) between the alleles underlying the selection response and fitness as defined in the absence of the artificial selection (Barton and Turelli 1989; Hill and Mbaga 1998). Other forms of antagonistic pleiotropy can also limit the response to selection and maintain V_A. For example, maternal genetic effects arise when an individual's phenotype is influenced by the maternal genotype independently of the inherited genetic effects (Mousseau and Fox 1998). When the covariance between maternal genetic variance (V_M) and V_A (COV_{AM}) is negative, as is commonly observed in domesticated animals, then antagonistic pleiotropy may maintain V_A and limit responses to selection (Wilson and Réale 2006).

The purpose of this study is to test specific hypotheses regarding the quantitative–genetic limits to selection in four replicate lines of laboratory house mice bred for high voluntary wheel running (Swallow et al. 1998). Voluntary wheel running varies widely among species of rodents (Dewsbury 1980) and shows "almost unbelievable individual differences" (Rundquist 1933, p. 416), yet is highly repeatable (Swallow et al. 1998). The extent to which wheel running reflects variation in one or more behaviors under natural conditions is unclear, but Mather (1981) argues that it reflects "exploratory migration" related to "the search for potential resources" (or the need to "monitor" resources; Perrigo and Bronson 1985; but see Careau et al. 2012). In any case, innate differences in capacities or motivation for wheel running (as an index of locomotor capacities) may have important implications for foraging, patrolling a territory, or dispersal.

Substantial response to selection occurred over the first 15–20 generations, at which point the four high runner (HR) lines were running ~2.5–3.0-fold more revolutions per day as compared with four nonselected control (C) lines (Kolb et al. 2010; Fig. 1). The replicate HR lines show a number of interesting sex and line differences that can be viewed as "multiple adaptive solutions." For example, although females and males showed a similar fold increase in running distance, the increased daily running distance was accomplished almost entirely by increased speed in females, but by both increased speed and duration of running in HR males (Garland et al. 2011 and references therein). Moreover, this experiment led to the discovery of a Mendelian recessive "mini-muscle" allele that causes a 50% reduction in hind limb muscle mass and many pleiotropic effects apparently conducive to supporting high levels of endurance exercise (Garland et al. 2002; Middleton et al. 2008). This gene of major effect increased in frequency in two of the four HR lines, and eventually became fixed in one (Garland 2003; Hannon et al. 2008). Despite these differences, a common feature to the HR lines is that they all seem to have reached a plateau at approximately generation 20, after which no further increase in wheel running has occurred despite continued directional selection (Fig. 1). A study conducted at generations 21–22 showed that the HR lines did not differ statistically

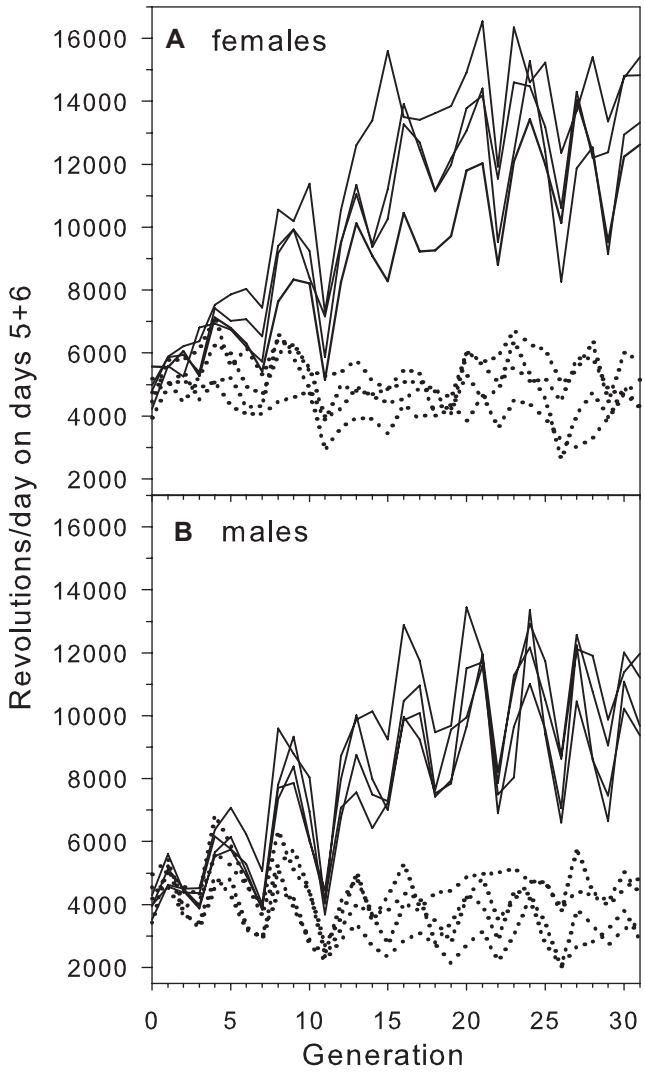

Figure 1. Unadjusted average running distance in (A) female and (B) male mice from each of the four nonselected control lines (dotted lines) and four selected lines (solid lines) plotted against generation. Values are the mean number of revolutions run on days 5 and 6 (the selected trait) of a 6-day period of wheel access for young adults. In control lines, average sample size per generation per line was 19.88 (range 12–31) for females and 19.63 (range 11–34) for males, for a total of 5057 mice. In selected lines, average sample size per generation per line was 50.16 (range 31–70) for females and 49.39 (range 22–69) for males, for a total of 12,743 mice (see Appendix S2). (Note that the apparent cyclical variation is probably attributable to seasonal effects; see Fig. S4 in Appendix S5.)

from C lines in terms of fertility, litter size, mean offspring mass, or most aspects of maternal-care behavior (Girard et al. 2002), suggesting that counterposing natural selection did not cause the selection limits.

We first use segmented regressions to formally test for the presence of a selection limit and to estimate its timing and the height of the plateau in each of the HR lines. Second, we compare the four HR lines with respect to their response to selection before reaching a limit (realized h^2). Third, we evaluate whether counterposing natural selection could explain the selection limits by testing for a reduction in litter size, a reduction in the *realized* selection differential (Roberts 1966), or a build-up of a negative COV_A between wheel running and litter size. Finally, we used "animal models" (Henderson 1973; Thompson 1973) to evaluate the effect of directional selection on key quantitative–genetic parameters, including, V_A, V_M, and COV_{AM}. In all these steps, we first analyze the data for both sexes averaged and then tested for sex differences because (i) females run ∼30% more than males (Fig. 1); (ii) sex-specific heterosis has been observed in line crosses (Hannon et al. 2011); and (iii) sex-specific analyses may elucidate the relative importance of each sex for the selection response in each line as a whole (Reeve and Fairbairn 1996).

Materials and Methods

ANIMALS AND DATA

The base population, housing conditions, wheel-running protocol, and selection procedures are described in previous publications (Swallow et al. 1998; Girard et al. 2002; Garland et al. 2011) and Appendix S1 (see also Fig. 1). The pedigree and wheel-running data from the base population through the first 31 generations of the experiment were compiled from raw data files, checked for outliers (see Appendix S2), for a total sample size of 17,988 individuals. The inbreeding coefficient (F) of each individual was calculated using ASReml-R version 3.0 (Butler et al. 2007). The population size and breeding scheme resulted in an increase in F from 0 (assumed) to a mean ($\pm$ SD) of 0.379 ± 0.064 at generation 31 ($n = 556$).

Standardized selection differentials were calculated for each generation within each line, as the difference between the mean wheel running (average of days 5 and 6) of the mice selected for breeding and the mean of the entire line divided by the phenotypic standard deviation of the line in that generation. This was done for each sex separately for sex-specific analyses and then averaged for the analysis on the mean. To remove the effects of temporal fluctuations of the line means that commonly occur during long-term selection experiments (e.g., see fig. 13.1 in Lynch 1994; Bünger et al. 1998; Keightley 1998; and Fig. 1), response to selection was calculated as the deviation between the mean of each HR line and the average of the means of the four C lines (Falconer and Mackay 1996; Hine et al. 2011). The pedigree, raw data, and descriptive statistics for each sex separately and their average are available in Dryad (doi:10.5061/dryad.37b28; for description see Appendix S2).

DATA ANALYSIS–SELECTION LIMIT

The presence of a selection limit was examined in each of the four replicate HR lines by using segmented regression in the "segmented" package (version 0.2.9.2; Muggeo 2008) in R (R Development Core Team 2006). Segmented regressions are typically used in data sets where the relationship between two variables exhibits an abrupt change past some threshold (breakpoint) of the independent variable and it is of interest to assess the threshold value where the effect of the independent variable changes. In this approach, the relationship between two variables is represented by two (or more) straight lines connected by breakpoints with an estimate of uncertainty (95% confidence interval [CI]) around the location of this breakpoint. We used a segmented regression to identify where the change occurred (breakpoint) in the relationship between cumulative response to selection and cumulative selection differential. By fixing the slope of the second regression line to 0 (i.e., a null right slope; Muggeo 2008), this approach allowed us to identify the breakpoint at which there was no more selective gain despite continued selection. We compared the fit of the simple linear model (three parameters: intercept, slope, and residuals) versus the segmented regression (four parameters: intercept, difference in slopes, breakpoint, and residuals) by calculating the difference in Akaike Information Criterion corrected for small sample size (ΔAICc).

Analyzing time series data introduces the potential for statistical nonindependence of data points across generations, hence autocorrelation among error terms should be evaluated (Hendry and Kinnison 1999). We extracted the residuals from the segmented regression models and used the corAR1 function in the "nlme" package in R to test for the presence of an autocorrelation structure of order 1. This analysis showed that the residuals were free of autocorrelation in all HR lines (all $P > 0.49$).

DATA ANALYSIS—REALIZED HERITABILITY

Following Lynch (1980), estimates of realized heritability (h^2_{w}, for within-family selection experiment) were obtained for each of the four replicate HR lines as the regression of cumulative response to selection (in units of phenotypic standard deviations: 2050 revolutions/day, average of days 5 and 6; Swallow et al. 1998, p. 235) on cumulative selection differential. Estimates of realized h^2_{w} were also converted to h^2 expected from mass selection or a randomly mating population (h^2_{r}) with the following formula: $h^2_{\mathrm{r}} = h^2_{\mathrm{w}}(1-t)/(1-r)$, where r is the coefficient of relationship of full sibs (0.5), and t is the intraclass correlation of full sibs for voluntary wheel running scores (Lynch 1980). We used the same intraclass correlation ($t = 0.246$) as in Swallow et al. (1998), calculated in generation 0 separately for males and females and then averaged. Realized h^2_{w} estimates were based on: (i) data from generations 0 to 10 (mainly to match Swallow et al. 1998); (ii) generations 0 through the selection limit (breakpoint in the segmented regressions); and (iii) after the selection limit (up to generation 31). In each set of generations, we tested for heterogeneity in h^2_{w} among replicate lines using analysis of covariances (ANCOVAs) with cumulative selection response as the dependent variable and cumulative selection differential and line as the independent variables in addition to an interaction between the two (i.e., test if slopes differ). We tested whether the height of the selection plateau differed among lines by restricting the data to generations after the limit and using an analysis of variance (ANOVA) with cumulative selection response as the dependent variable and line as the independent variable.

DATA ANALYSIS—COUNTERPOSING NATURAL SELECTION

We addressed whether counterposing natural selection is causing the observed selection limits by looking at how litter size and realized selection differentials changed in HR lines across generations using linear regressions (treating generation as a continuous variable). We calculated litter size as the number of pups attributed to an individual at weaning. This implies that a litter size could not be quantified in individuals that were not paired, did not give birth to a litter (rare), lost their entire litter before weaning (rare), or whose litter was discarded at weaning (extremely rare, see also Girard et al. 2002). We calculated the realized selection differential by weighting wheel running of each selected individual by the number of pups that it produced. Thus, if there was a negative correlation between wheel running and litter size, then the realized selection differential would be lower than the expected selection differential (based on the average of wheel running in selected individuals; Falconer and Mackay 1996). Finally, we also tested for the presence of a negative $\mathrm{COV}_{\mathrm{A}}$ between wheel running and litter size using a bivariate animal model.

DATA ANALYSIS—ANIMAL MODEL

We estimated quantitative–genetic parameters using linear mixed effects models, commonly known as the "animal model," with restricted maximum likelihood estimation in ASReml-R version 3.0 (Butler et al. 2007). For sake of comparison, we also estimated h^2 using offspring-on-parent regressions (see Fig. S1 and Table S1 in Appendix S3). For the purpose of estimating the actual V_{A} in any particular generation of selection, the animal model has the "undesirable" property of making inference back to animals in the base population (assumed to be unrelated, unselected, and noninbred; Lynch and Walsh 1998). We followed the heuristic approach developed by Sorensen and Kennedy (1984; but see van der Werf and de Boer 1990), which treats animals in generation t as unrelated and includes data for all animals in generations $\geq t$. Thus, we are making inference to generation t instead of the base population. We estimated V_{A} before the selection limit by

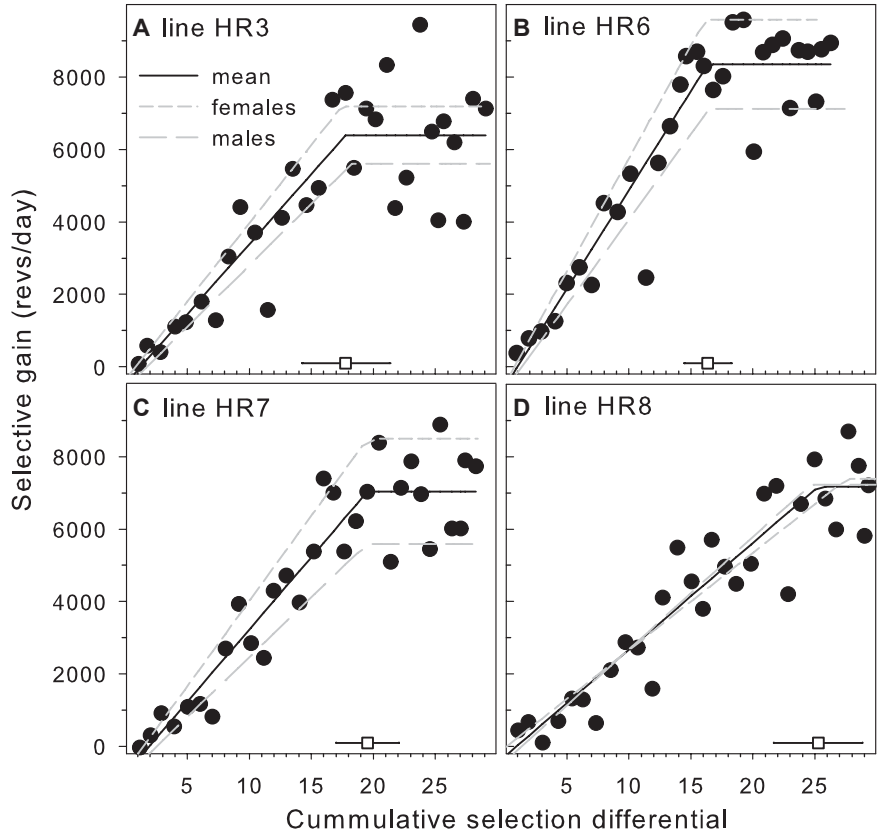

Figure 2. **Cumulative response to directional selection (measured as the deviation of each selected line from the mean of the four control lines) as a function of the cumulative selection differential (in units of standard phenotypic deviation) for four lines (A–D) of mice artificially selected for high voluntary wheel running (revolutions per day). Shown are fitted lines from the segmented regression analysis with the second slope fixed to 0 (solid line: average females and males; short-dashed line: females; long-dashed line: males; data points are for the mean of females and males). The 95% confidence intervals for the position of the breakpoints (open squares) are shown as horizontal lines along the x-axis. These graphs demonstrate that each line (except possibly line HR8) reached a selection limit and also that the height of the plateaus differed among the four replicate lines (see Results).**

analyzing the data from generation 0 up to and including generation 10. We estimated V_A after the selection limit by analyzing the data from generations 21–31 only, assuming that animals of generation 21 were unrelated. This interval was set to generations 25–31 for line HR8 that reached a limit later than others (Fig. 2D). Although the pedigree starts from generation 21 or 25 in these analyses, we specified the known F of each individual (as calculated from the entire pedigree back to the base population) to obtain an appropriate additive genetic relationship matrix in which the diagonal elements are equal to $1 + F$. See Figure S2 in Appendix S4 for a comparison of V_A estimates obtained by this procedure with those obtained using (i) the complete pedigree back to the base population and (ii) assuming that animals of generation 21 were unrelated but with an F of 0.

In all animal-model analyses, wheel running on days 5 and 6 was treated as a repeated measure (as opposed to the average of days 5 and 6) to facilitate separation of genetic and nongenetic sources of among-individual variation. Wheel-running data were standardized (mean = 0, variance = 1) for each line in each generation to account for changes in variance due to occasional large fluctuations in phenotypic means across generations (Fig. 1; see also Fig. S3 in Appendix S5 for control-adjusted means). Moreover, the total phenotypic variance increased with the gradual increase of trait means across generations in the HR lines (the correlation between the mean and variance from generation 0 to 31 ranged from 0.68 to 0.82; Appendix S5). Genetic variances pertaining to each line were estimated before and after the limit under the repeated measure animal model, by partitioning phenotypic variance (V_P) as: V_P = fixed effects + V_{PE} + V_C + V_A + V_R, where fixed effects were fitted within generation (sex, age, F, line, and measurement block [batches 1–3, rooms 1–2]) and V_{PE}, V_C, V_A, and V_R correspond to permanent environmental variance (individual identity), common environmental variance (dam identity), additive genetic variance, and residual variance, respectively. Initially, we estimated dominance genetic variance (Wolak 2012), but it was never statistically significant (see Table S2 in

Appendix S6) and subsequent analyses were conducted without this term in the model.

We tested for the statistical significance of the random-effect variance components by comparing the log-likelihoods of a full model that included the variance component and a reduced model that excluded it (or fixed it to 1e−08). The likelihood ratio test (LRT) statistic is equal to twice the difference in log-likelihoods between the two nested models and is assumed to follow a χ^2-distribution with degrees of freedom (df) equal to the difference in the number of parameters estimated. When testing a single component against a boundary of its parameter space (e.g., $V_A > 0$), the χ^2 statistic is distributed as an equally weighted mixture of χ^2 distributions with 1 and 0 df ($\chi^2_{0:1}$; Self and Liang 1987). In practice, this is equivalent to halving *P*-values obtained from a χ^2-distribution with 1 df (Dominicus et al. 2006).

We quantified variance components, before and after the limit separately, by running three sets of models. For each set, we included all the data, but allowed different levels of heterogeneity in the variance components (V_{PE}, V_C, V_A, and V_R) according to selection group (i.e., HR vs. C), sex, and line. First, we allowed variance components to be estimated separately between selection groups. Hence, there were eight components estimated in this model (V_{PE}, V_C, V_A, and V_R estimated separately for C and HR selection groups). This model allowed us to test whether V_{PE}, V_C, V_A, or V_R were significantly different between selection groups (i.e., in C vs. HR lines) using a LRT comparing the full model versus a model where the variance component of interest was held constant in C and HR selection groups (one less parameter = χ^2_1). Second, we allowed variance components to be estimated separately between sexes in each selection group. Hence, there were 16 components estimated in this model (V_{PE}, V_C, V_A, and V_R estimated for females and males separately in C and HR groups). This model allowed us to test whether V_{PE}, V_C, V_A, or V_R were statistically different between sexes within selection groups using a LRT. Finally, we allowed variance components to be estimated separately between sexes in all eight lines. Hence, there were 64 components estimated in this model (V_{PE}, V_C, V_A, and V_R estimated for females and males of each line). As the line estimates produced by this model can be treated as independent replicates (given the experimental design), we tested whether V_A was different in C versus HR lines before and after the limit using *t*-tests (separately in each sex).

In each model with sex-specific variance estimated earlier, the cross-sex genetic correlation (r_{MF}) was assumed to be equal to 1, which may not be appropriate. We therefore ran an additional model to estimate r_{MF} and test whether it was significantly smaller than 0.999 (dividing *P*-values by 2) and significantly larger than 0. We also ran another model in which V_P was partitioned as: V_P = fixed effects + V_{PE} + V_M + V_A + COV_{AM} + V_R, where fixed effects were fitted as earlier and V_M corresponds to maternal genetic variance (dam identity linked to the pedigree) and COV_{AM} corresponds to the covariance between V_A and V_M. The model for heterogeneous variance among lines did not converge and we therefore only present results for models with heterogeneous variance between the two selection groups. In addition, attempting to include V_C in this model yielded convergence problems when fixing COV_{AM} to be equal in C and HR lines, so V_C was left out. As a standardized measure of COV_{AM}, we calculated the direct-maternal genetic correlation (r_{AM}) by dividing COV_{AM} by the square root of the product V_A and V_M.

If natural selection is counteracting artificial selection, then we should detect a negative COV_A between wheel running and litter size after the limit. We therefore ran a bivariate animal model in which wheel running on days 5 and 6 was modeled as a repeated measure (with the same fixed effects as earlier) and litter size was modeled as a singly measured trait [standardized (mean = 0, variance = 1) for each line in each generation, as for wheel running]. We included a COV_A between litter size and wheel running. The residual variance component for litter size was constrained to be 0 (sensu Morrissey et al. 2012). Thus, V_R in litter size is represented as V_{PE} to allow estimation of the covariance ($COV_{PE\text{-}R}$) between litter size and the nongenetic component of individual differences in wheel running. As pointed out by Morissey et al. (2012), this does not imply that we can estimate V_{PE} for a singly-measured trait (i.e., litter size), but this constraint in the mixed model allows the statistical estimation of the biologically interesting relationship between the residuals of litter size and the nongenetic but repeatable component of variation in wheel running. The model for heterogeneous variance among lines did not converge and we therefore only present results for the model allowing for heterogeneous variance between selection groups. We additionally allowed for separate estimation in males and females. We estimated the phenotypic correlation (r_P) between wheel running and litter size by removing all random effects (i.e., V_{PE}, V_C, and V_A) and including a covariance among the residuals (while retaining fixed effects). Approximate standard errors for these correlations (r_P, r_A, and r_{AM}) were obtained using the delta method in R (see appendix 1 in Lynch and Walsh 1998).

Results

SELECTION LIMIT

Using segmented regressions on the mean selection response (males and females) for each of the HR lines separately, we found that the cumulative selection response increased until the average (± SD) cumulative selection differential across the four selected lines reached a value of 19.8 ± 3.9 (n = 4; Table 1; Fig. 2A–D). In terms of number of generations, this corresponds

Table 1. Estimated timing of selection limit in four replicate lines of mice artificially selected for high levels of voluntary wheel running (mean revolutions/day on days 5 and 6 of a 6-day period of wheel access). A segmented regression analysis with the second slope fixed to 0 (i.e., when a limit is reached) was performed on cumulative response to directional selection (measured as the deviation of each selected line from the mean of the four control lines) as a function of the cumulative selection differential. The analyses were done separately for females (f), males (m), as well as for the average (mean). Breakpoints with 95% confidence intervals (CI) are shown with the corresponding generation (Gen) along with estimates of the left slopes (± SE), the residual standard error of the left slope, the height of the plateau (i.e., intercept of the null right slope in wheel revolutions/day), and the residual standard error of the null right slope. Also shown are the log-likelihoods of the linear (log L linear, three parameters) and segmented regressions (log L seg, four parameters) and the difference in Akaike Information Criterion corrected for small sample size (ΔAICc) between a simple linear model and the segmented regression.

		Breakpoint		Left slope		Null right slope		Linear vs. segmented regression		
Line	Sex	Estimate ±95% CI	Gen	Estimate ±SE	Residual SE	Plateau ±SE	Residual SE	Log L linear	Log L seg	ΔAICc
HR3	f	17.44±3.27	18	433.3±58.3	1015	7188±481	1,798	−280.6	−273.9	10.73
HR3	m	18.27±4.58	17	339.0±60.4	963	5600±451	1,746	−278.9	−273.5	8.32
HR3	mean	17.78±3.55	17	386.2±54.8	965	6389±463	1,732	−279.0	−272.6	10.21
HR6	f	16.45±1.92	18	610.5±51.2	1192	9635±332	1,198	−279.1	−268.8	17.94
HR6	m	16.53±2.54	18	469.2±51.9	1131	7122±371	1,386	−278.8	−270.2	14.60
HR6	mean	16.39±1.92	18	543.8±46.0	1008	8353±296	1,107	−276.6	−264.7	21.02
HR7	f	19.53±2.26	21	470.1±37.0	917	8491±404	1,339	−272.0	−265.1	11.01
HR7	m	19.45±3.55	20	330.2±42.2	836	5588±446	1,544	−272.4	−267.1	8.07
HR7	mean	19.60±2.54	21	398.4±35.3	797	7047±420	1,393	−270.7	−263.7	11.38
HR8	f	27.59±4.57	27	269.0±24.3	941	7386±949	2,122	−266.3	−265.6	−1.11
HR8	m	24.67±3.58	24	310.1±28.3	1040	7229±384	1,086	−266.0	−264.7	0.02
HR8	mean	25.25±3.57	25	295.1±25.2	939	7173±485	1,282	−264.2	−263.1	−0.46

to an average of 20.3 ± 3.6 generations before reaching a limit (Table 1). According to the differences in AICc, a segmented regression described the data significantly better than a simple linear regression in three of the four lines (Table 1). One of the lines (HR8) reached a limit (plateau) later (at generation 25) than the other three HR lines (range = 17–21 generations; see Fig. 2D), and for this line the AICc for the simple linear regression was higher than for the segmented regression (Table 1). Comparing the 95% CI around the breakpoint estimates reveal that line HR8 (lower CI = 21.68) reached the limit significantly later than lines HR3 (upper CI = 21.33) and HR6 (upper CI = 18.31), but not line HR7 (upper CI = 22.14).

Analyzing data for each sex separately showed that the timing of the selection limit was similar in females (generation 21.0 ± 4.2) and males (generation 19.8 ± 3.1; Table 1; Fig. 2). The rate of initial response to selection (i.e., before a limit was reached), however, was 23% faster in females than in males (Table 1; Fig. 2). The height of the selection plateau (relative to C) was also 28% higher in females (8175 ± 1130 revolutions per day) than males (6385 ± 914.3 revolutions per day; Table 1; Fig. 2). Note, however, that there was no statistical difference between females and males in line HR8 in terms of initial response and height of the plateau (Table 1; Fig. 2D).

REALIZED HERITABILITY

The h^2_w, averaged between the sexes, differed among the replicate lines, as indicated by a significant interaction term (cumulative selection differential by line) in the ANCOVAs up to 10 generations ($F_{3,36} = 5.367$, $P = 0.004$) and up to the selection limit ($F_{3,77} = 10.551$, $P < 0.001$). A close look at the ANCOVA up to the selection limit revealed that h^2_w was significantly lower in line HR8 than line HR3 ($t_{77} = 2.351$; $P = 0.021$; Table 2), line HR6 ($t_{77} = 5.344$; $P < 0.001$; Table 2), and line HR7 ($t_{77} = 3.151$; $P = 0.002$; Table 2). The h^2_w was also significantly higher in line HR6 than line HR3 ($t_{77} = 2.463$; $P = 0.016$; Table 2) and line HR7 ($t_{77} = 2.418$; $P = 0.016$; Table 2). The only nonsignificant difference in h^2_w was between line HR3 and HR7 ($t_{77} = 0.306$; $P = 0.760$; Table 2).

Using only the data for generations after the limit (Table 2), the height of the plateau was significantly different among the HR lines (ANOVA with line as a fixed effect: $F_{3,39} = 5.901$, $P = 0.002$). A close look at the ANOVA indicated that the selection plateau was significantly higher in line HR6 than line HR3 ($t_{39} = 4.110$; $P = 0.0001$; Table 1) and line HR7 ($t_{39} = 2.743$; $P = 0.009$; Table 1). There was no other significant difference in the height of the selection plateau among HR lines (see Table S3d in Appendix S7).

Table 2. Realized heritabilities (h^2_w; for within-family selection) for wheel running (mean revolutions/day on days 5 and 6 of a 6-day period of wheel access) estimated (A) through generation 10, (B) through the selection limit, and (C) after the selection limit in each of the four replicate lines of mice artificially selected for high levels of voluntary wheel running. The analyses were done separately for females, males, as well as for the average (mean). Also shown are the generations over which h^2_w was estimated (gens) and the realized heritability expected in a mass selection experiment (h^2_r).

	Mean			Females			Males		
Line	Gens	$h^2_w \pm$SE	$h^2_r \pm$SE	Gens	$h^2_w \pm$SE	$h^2_r \pm$SE	Gens	$h^2_w \pm$SE	$h^2_r \pm$SE
A. Through generation 10									
HR3	0–10	0.191 ± 0.027	0.288 ± 0.032	0–10	0.192 ± 0.024	0.289 ± 0.028	0–10	0.192 ± 0.034	0.289 ± 0.040
HR6	0–10	0.253 ± 0.023	0.381 ± 0.027	0–10	0.276 ± 0.023	0.416 ± 0.027	0–10	0.231 ± 0.025	0.348 ± 0.030
HR7	0–10	0.178 ± 0.029	0.269 ± 0.034	0–10	0.184 ± 0.031	0.278 ± 0.037	0–10	0.171 ± 0.031	0.258 ± 0.036
HR8	0–10	0.112 ± 0.022	0.168 ± 0.026	0–10	0.098 ± 0.022	0.147 ± 0.026	0–10	0.127 ± 0.028	0.192 ± 0.032
Mean		0.183 ± 0.029	0.277 ± 0.044		0.188 ± 0.036	0.283 ± 0.043		0.180 ± 0.022	0.272 ± 0.032
B. Through limit									
HR3	0–17	0.193 ± 0.020	0.291 ± 0.023	0–18	0.206 ± 0.019	0.311 ± 0.023	0–17	0.177 ± 0.020	0.266 ± 0.024
HR6	0–18	0.259 ± 0.020	0.390 ± 0.024	0–18	0.299 ± 0.026	0.451 ± 0.030	0–18	0.221 ± 0.022	0.333 ± 0.026
HR7	0–21	0.200 ± 0.013	0.302 ± 0.015	0–21	0.231 ± 0.015	0.348 ± 0.017	0–20	0.160 ± 0.014	0.241 ± 0.017
HR8	0–25	0.142 ± 0.011	0.214 ± 0.013	0–26	0.138 ± 0.011	0.208 ± 0.013	0–24	0.153 ± 0.013	0.230 ± 0.015
Mean		0.199 ± 0.024	0.299 ± 0.028		0.219 ± 0.033	0.330 ± 0.039		0.178 ± 0.015	0.268 ± 0.023
C. After limit									
HR3	18–31	-0.011 ± 0.067	-0.016 ± 0.079	19–31	-0.028 ± 0.080	-0.043 ± 0.095	18–31	-0.042 ± 0.078	-0.063 ± 0.092
HR6	19–31	-0.004 ± 0.054	-0.006 ± 0.064	19–31	0.010 ± 0.063	0.015 ± 0.075	19–31	-0.038 ± 0.071	-0.058 ± 0.084
HR7	22–31	0.074 ± 0.087	0.112 ± 0.103	22–31	0.110 ± 0.080	0.166 ± 0.094	21–31	0.035 ± 0.104	0.053 ± 0.122
HR8	26–31	0.095 ± 0.224	0.143 ± 0.264	27–31	-0.293 ± 0.472	-0.442 ± 0.556	25–31	0.182 ± 0.173	0.274 ± 0.204
Mean		0.039 ± 0.027	0.058 ± 0.032		-0.050 ± 0.086	-0.076 ± 0.101		0.034 ± 0.052	0.052 ± 0.079

Analyzing data for each sex separately showed that estimates of h^2_w were slightly higher in females than males in all lines except line HR8 (Table 2). The h^2_w up to the selection limit also differed among the replicate lines in both sexes (Table S3a in Appendix S7), with slightly more heterogeneity among lines in females than males (Table S3b). The height of the plateau was also significantly different among the HR lines in both sexes (Table S3c; see also Fig. 2 and Table 1).

COUNTERPOSING NATURAL SELECTION

Least squares linear regression analysis indicated that litter size did not decrease significantly across generations in any of the HR lines ($t_{30} < 1.45$; $P > 0.16$; Fig. S5 in Appendix S8). The expected and realized selection differentials were highly correlated in all HR lines (range: $r = 0.92$–0.98) and did not differ significantly in any line (paired t-test; $t < 0.34$, $P > 0.74$). The realized selection differential decreased across generations in all HR lines in both females ($t_{30} > 2.29$; $P < 0.03$) and males ($t_{30} > 1.93$; $P < 0.08$; Fig. 3A–D). However, the average (±SE) realized selection differential after the limit was still far from being 0 in both females and males (Fig. 3A–D).

Before the limit, the r_P between wheel running and litter size was -0.003 ± 0.049 (LRT $\chi^2_1 = 0.003$, $P = 0.95$) in females and 0.025 ± 0.047 (LRT $\chi^2_1 = 0.28$, $P = 0.59$) in males. After the limit, the r_P was -0.153 ± 0.054 (LRT $\chi^2_1 = 7.44$, $P = 0.006$) in females and -0.005 ± 0.056 (LRT $\chi^2_1 = 0.007$, $P = 0.93$) in males. The bivariate animal model indicated that the COV_A between litter size and wheel running was positive in both sexes before the limit, and significantly so in males (Table 3A). After the limit, the COV_A was not significantly different from 0 in females and positive and marginally nonsignificant in males (Table 3B). The covariance between the residuals of litter size and the nongenetic component of individual differences in wheel running ($COV_{PE\text{-}R}$) was always negative in males and females, both before and after the limit (Table 3A, B).

ANIMAL MODELS—BASE POPULATION

Before selection was applied, wheel running was measured in 188 and 557 individuals of generations −1 and 0, respectively. Animal-model analysis of wheel running in the base population is presented in Table 4A for sexes pooled. Wheel running was significantly higher in females (mean ± SE = 5190 ± 91 revolutions per day) than males (4107 ± 74 revolutions per day; $F_{1,638} = 51.31$, $P < 0.001$). As expected, V_A was significant (Table 4A). A separate model in which V_C was replaced by V_M in addition to a COV_{AM} term revealed the presence of a negative but nonsignificant r_{AM} (Table 5A).

Modeling heterogeneous variance estimates for sexes in the base population did not reveal significant sex differences in any variance component (LRT $\chi^2_1 < 0.99$, $P > 0.32$; Table 4B). V_A was marginally nonsignificant in females ($V_A = 0.17 \pm 0.11$), but slightly higher and significantly different from 0 in males ($V_A = 0.25 \pm 0.09$; Table 4B). The cross-sex genetic correlation was not significantly smaller than unity in the base population ($r_{MF} \pm SE = 0.95 \pm 0.32$; see Table S4a in Appendix S9).

ANIMAL MODELS—BEFORE AND AFTER THE LIMIT

The first models were run under the assumption of homogeneous variance between sexes and heterogeneous variance between selection groups (Table 4C, D). V_A accounted for a moderate portion of the variance (range of h^2: 0.13–0.24) and was significantly lower in HR lines compared to C lines before the limit (LRT $\chi^2_1 = 4.80$, $P = 0.028$; Table 4C), but this difference became marginally nonsignificant after the limit (LRT $\chi^2_1 = 2.59$, $P = 0.10$; Fig. 4C; Table 4D).

Allowing heterogeneous variances between sexes did not reveal any significant sex differences in V_{PE}, V_C, and V_A in C lines either before or after the limit (LRT $\chi^2_1 < 0.88$, $P > 0.35$; Table 4E, F), except that V_R seemed higher in females than males (Fig. 4D). In HR lines, however, V_{PE} was significantly higher in females than males both before (LRT $\chi^2_1 = 13.77$, $P < 0.001$) and after (LRT $\chi^2_1 = 10.02$, $P = 0.002$; Table 4E, F; Fig. 4A) the limit. V_C was also significantly higher in females than males in HR lines before the limit (LRT $\chi^2_1 = 5.91$, $P = 0.015$), but this difference vanished after the limit (LRT $\chi^2_1 = 2.05$, $P = 0.15$; Table 4E, F; Fig. 4B). V_A was not significantly different in HR males versus females either before or after the limit (LRT $\chi^2_1 < 0.66$, $P > 0.42$; Table 4E, F). The sex-specific V_A estimates were slightly higher in models allowing cross-sex genetic correlations, which were not significantly smaller than unity before the limit in both C ($r_{MF} \pm SE = 0.95 \pm 0.08$) and HR ($r_{MF} \pm SE = 0.95 \pm 0.07$; Table S4b) lines. After the limit, the r_{MF} was still not significantly smaller than unity in C lines ($r_{MF} \pm SE = 0.92 \pm 0.10$), but was in HR lines ($r_{MF} \pm SE = 0.84 \pm 0.08$; LRT $\chi^2_{0:1} = 7.46$, $P = 0.003$; Table S4c).

Allowing sex-specific variances to be heterogeneous among all eight lines revealed that, before the limit, V_A was significantly greater than 0 in all but one line in females (line C4; Table S5a in Appendix S11), and all but two lines in males (lines C4 and HR6; Table S5a). After the limit, V_A was significantly greater than 0 in all but one line in females (line HR8; Table S5b) and all but two lines in males (lines C1 and HR 7; Table S5b). Contrary to the significant differences in V_A in C versus HR lines detected with LRT, using a t-test on the independent replicate line estimates indicated no statistical difference in V_A between C versus HR lines either before ($t_6 < 1.683$, $P > 0.143$) or after the limit ($t_6 < 1.528$, $P > 0.177$).

Replacing V_C by V_M and adding a COV_{AM} to the model with heterogeneous variances between selection groups indicated that COV_{AM} was not significantly different from 0 in C lines before or

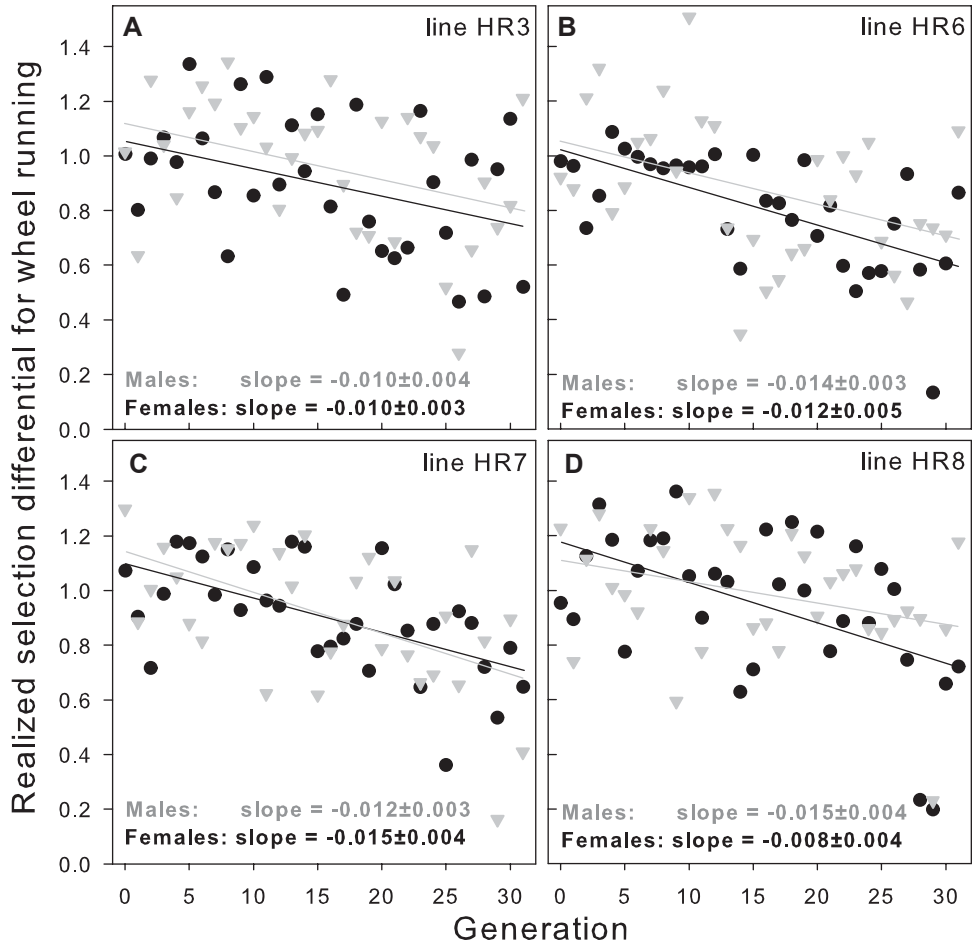

Figure 3. **Realized selection differentials (in units of standard phenotypic deviation) for voluntary wheel running in four lines (A–D) of mice artificially selected for high voluntary wheel running over 31 generations, shown for females (black dots) and males (gray triangles) separately. These graphs demonstrate that although the realized selection differentials declined, it never approached 0 and so the declines cannot explain the selection limits illustrated in Figure 2.**

after the limit (Table 5B). In HR lines, however, V_M was larger than in C lines and the COV_{AM} significantly different from 0 before the limit ($r_{AM} = -0.68 \pm 0.09$; Table 5B). After the limit, V_M and COV_{AM} were smaller than before the limit and the COV_{AM} was no longer significantly different from 0 ($r_{AM} = -0.35 \pm 0.22$; Table 5C). The COV_{AM} was significantly different between C and HR lines before the limit (LRT $\chi^2_1 = 4.85$, $P = 0.028$), but not after the limit (LRT $\chi^2_1 = 0.17$, $P = 0.38$). Adding sex-specific variance estimates revealed that, before the limit, the COV_{AM} was significantly different from 0 in females ($r_{AM} = -0.75 \pm 0.09$) but not in males ($r_{AM} = -0.57 \pm 0.19$; Table 5D). After the limit, none of the COV_{AM} estimates were significantly different from 0 (Table 5E).

Discussion

We found that directional selection yielded statistically different results among four replicate selected lines in terms of initial rate of response to selection (h^2_w, range: 0.14–0.25), timing at which a selection limit was reached (range = 17–25 generations), and height of the plateau (range = 6394–8344 revolutions/day). This heterogeneity among lines stresses the importance of replication in selection studies (see also Garland 2003; Garland and Rose 2009), as our conclusions about the response to selection and selection limits in wheel running could differ substantially depending on which line they would be based on.

Whether these differences reflect underlying variation in the physiological mechanisms of the response to selection (i.e., alternate adaptive solutions), or random genetic processes, remains unclear. Given that the lines were established with 10 male and 10 female parents, founder effects may have initially caused small divergences among the HR lines by altering the relative frequencies of alleles which contribute to voluntary wheel running (see also Garland et al. 2002). Further divergence among the HR lines may have been caused by genetic drift changing the frequencies of alleles whose effects are neutral or counter to what selection

Table 3. Components of variance (± SE; V_{PE}, permanent environmental; V_C, common environmental; V_A, additive genetic; V_R, environmental variance) in bivariate quantitative–genetic analyses of voluntary wheel running (RUN; treating revolutions run on days 5 and 6 as repeated measures of the same trait) and litter size (number of pups at weaning) in four lines artificially selected for high levels of voluntary wheel running (A) before and (B) after the selection limit in females (left columns) and males (right columns). In all models, a covariance was included between V_{PE} in RUN and V_R in litter size ($COV_{PE\text{-}R}$) in addition to covariance between V_A in both traits (COV_A) and their statistical significance was tested by comparing log-likelihoods of the full model versus a reduced model that excluded the covariance of interest.

	Females			Males		
Component	Estimate±SE	χ^2	*P*	Estimate±SE	χ^2	*P*
A. Through generation 10						
$V_{PE\text{-}RUN}$	0.092 ± 0.070			0.081 ± 0.052		
$COV_{PE\text{-}R}$	-0.076 ± 0.057	1.72	0.1899	-0.094 ± 0.049	3.89	0.0486
$V_{R\text{-litter size}}$	0.962 ± 0.093			0.972 ± 0.095		
$V_{C\text{-}RUN}$	0.269 ± 0.076			0.160 ± 0.052		
$V_{A\text{-}RUN}$	0.274 ± 0.103			0.228 ± 0.076		
COV_A	0.081 ± 0.066	1.51	0.2190	0.122 ± 0.058	5.35	0.0207
$V_{A\text{-litter size}}$	0.093 ± 0.085			0.087 ± 0.089		
$V_{R\text{-}RUN}$	0.381 ± 0.024			0.226 ± 0.014		
B. After limit						
$V_{PE\text{-}RUN}$	0.054 ± 0.061			0.013 ± 0.045		
$COV_{PE\text{-}R}$	-0.110 ± 0.060	6.62	0.0101	-0.098 ± 0.053	3.15	0.0761
$V_{R\text{-litter size}}$	1.048 ± 0.079			0.897 ± 0.115		
$V_{C\text{-}RUN}$	0.222 ± 0.064			0.262 ± 0.056		
$V_{A\text{-}RUN}$	0.152 ± 0.071			0.113 ± 0.056		
COV_A	-0.003 ± 0.051	0.00	0.9999	0.098 ± 0.050	3.42	0.0643
$V_{A\text{-litter size}}$	$0.001 \pm$ NA			0.123 ± 0.091		
$V_{R\text{-}RUN}$	0.252 ± 0.018			0.180 ± 0.013		

favors. Interestingly, across HR lines, the initial rate of response to selection tended to be negatively correlated with the timing at which a limit was reached, but positively correlated with the height of the plateau (see Fig. S7 in Appendix S12).

Previously, Garland et al. (2011, and references therein) reported heterogeneity in the correlated responses of wheel-running components (average running speed and duration), as well as in some lower-level traits related to exercise abilities or motivation for running (i.e., "multiple adaptive solutions"; see Introduction). Male and female HR mice have converged on the same amount of time spent running per day, whereas females run more minutes per day in C and the base population. Our results suggest that these sex-specific responses to selection in speed and duration of wheel running did not translate into sex-specific changes in V_A in response to directional selection on revolutions/day. Interestingly, the r_{MF} became significantly different from unity after the limit in HR but not in C lines. Future quantitative–genetic analyses of subordinate traits may reveal clearer sex differences in quantitative–genetic architecture.

Our results suggest that natural selection is not strongly counteracting artificial selection in this experiment (see also Girard et al. 2002). First, we did not detect a general reduction in litter size across generations (Appendix S8). Second, the phenotypic correlation between litter size and wheel running was small, and although the selection intensity declined significantly across generations in all four HR lines, it was still substantial after the selection limit (Fig. 3). Third, after the limit the COV_A between wheel running and litter size was positive in males and not significantly different from 0 in females. Therefore, at the genetic level, males that run more tend to sire larger litters, whereas wheel running and litter size appear genetically independent in females (Table 3; see also fig. 4d in Girard et al. 2002).

Interestingly, there appears to be a negative covariance between the environmental component (V_R) of litter size and the nongenetic but repeatable component (V_{PE}) of wheel running, in both males and females before and after the limit in HR lines. Although it is unlikely that this negative covariance can cause the limit because these components are mainly of environmental nature (and the covariance between the two was also negative *before* the limit), one may still wonder how it could arise. One possibility is that wheel running induces permanent environmental effects that have negative effects on subsequent reproduction. For example, mice that run more tend to lose more body mass (Kelly et al. 2011), which could lower

Table 4. Components of variance (± SE; V_{PE}, permanent environmental; V_C, common environmental; V_A, additive genetic; V_R, environmental variance) in univariate quantitative–genetic analyses of voluntary wheel running (treating revolutions run on days 5 and 6 as repeated measures of the same trait) in four nonselected control (C) lines of house mice and four lines artificially selected for high levels of voluntary wheel running (high runner or HR lines). Models were run for (A) the base population (generations −2 to 0) assuming homogeneous variance in females and males, (B) the base population allowing with heterogeneous variance components for each, (C) before and (D) after the selection limit with heterogeneous variance between selection groups (C vs. HR lines), (E) before and (F) after the limit with heterogeneous variance between sexes and selection groups. Models were run separately on different intervals of generations (gen), from either 0 to 10 (before the limit) or generations 21 to 31 (after the limit, except for line HR8 for which the interval was 25 to 31[a]). Statistical significance of each variance component was tested using a log-likelihood ratio test of the full model versus a reduced model that excluded the variance component of interest. Wheel-running data were standardized (mean = 0, variance = 1) within each generation for each line, such that they are relatively comparable (depending on the proportion of variance explained by the fixed effects, which were fitted within generation). Also shown are the ratios of (proportion of variance explained by) V_{PE}, V_C, and V_A over the sum of V_{PE}, V_C, V_A, and V_R (respectively, p^2, c^2, and h^2; see Fig. S6 for a graphical representation of these ratios).

			Permanent environment			Common environment			Additive genetic			Residual	Ratios		
Gens	Sex	Group	V_{PE}±SE	χ^2	P	V_C±SE	χ^2	P	V_A±SE	χ^2	P	V_R±SE	p^2±SE	c^2±SE	h^2±SE
A. Base population, sexes pooled															
−2–0		NA	0.458±0.049	60.92	<0.0001	0.037±0.032	1.76	0.0926	0.171±0.073	7.31	0.0034	0.179±0.009	0.54±0.06	0.04±0.04	0.20±0.08
B. Base population, heterogenous variance for sexes															
−2–0	f	NA	0.473±0.083	24.98	<0.0001	0.057±0.064	0.99	0.1599	0.168±0.111	2.64	0.0521	0.245±0.018	0.50±0.09	0.06±0.07	0.18±0.12
−2–0	m	NA	0.378±0.072	19.26	<0.0001	0.000±NA	0.00	0.5000	0.253±0.087	8.75	0.0015	0.112±0.008	0.51±0.10	0.00±0.00	0.34±0.11
C. Before the limit, sexes pooled, heterogeneous variance for selection group															
0–10		C	0.489±0.034	193.05	<0.0001	0.020±0.020	0.83	0.1812	0.224±0.046	53.92	<0.0001	0.208±0.007	0.52±0.04	0.02±0.02	0.24±0.04
0–10		HR	0.517±0.020	273.56	<0.0001	0.057±0.013	32.59	<0.0001	0.117±0.030	29.59	<0.0001	0.186±0.004	0.59±0.03	0.06±0.01	0.13±0.03
D. After the limit, sexes pooled, heterogeneous variance for selection group															
21–31		C	0.441±0.035	145.52	<0.0001	0.077±0.024	14.10	0.0001	0.183±0.040	35.90	<0.0001	0.204±0.007	0.49±0.04	0.09±0.03	0.20±0.04
21–31[a]		HR	0.449±0.022	200.77	<0.0001	0.054±0.013	32.54	<0.0001	0.107±0.026	32.13	<0.0001	0.170±0.004	0.58±0.03	0.07±0.02	0.14±0.03
E. Before the limit, heterogeneous variance for selection group and sexes															
0–10	f	C	0.488±0.061	62.65	<0.0001	0.041±0.045	0.84	0.1790	0.244±0.072	20.67	<0.0001	0.270±0.013	0.47±0.06	0.04±0.04	0.23±0.06
0–10	f	HR	0.563±0.030	222.13	<0.0001	0.100±0.023	28.08	<0.0001	0.108±0.039	14.55	<0.0001	0.239±0.007	0.56±0.03	0.10±0.02	0.11±0.04
0–10	m	C	0.494±0.049	80.04	<0.0001	0.000±NA	0.00	0.5000	0.173±0.056	15.29	<0.0001	0.146±0.007	0.61±0.06	0.00±0.00	0.21±0.06
0–10	m	HR	0.431±0.024	189.02	<0.0001	0.034±0.016	5.97	0.0073	0.131±0.035	27.13	<0.0001	0.135±0.004	0.59±0.04	0.05±0.02	0.18±0.05
F. After the limit, heterogeneous variance for selection group and sexes															
21–31	f	C	0.413±0.056	51.56	<0.0001	0.135±0.045	10.31	0.0007	0.180±0.059	15.59	<0.0001	0.269±0.013	0.41±0.05	0.14±0.04	0.18±0.06
21–31	f	HR	0.487±0.034	132.34	<0.0001	0.077±0.022	17.91	<0.0001	0.126±0.037	22.04	<0.0001	0.196±0.006	0.55±0.04	0.09±0.02	0.14±0.04
21–31[a]	m	C	0.344±0.049	48.67	<0.0001	0.105±0.039	8.15	0.0022	0.199±0.055	20.20	<0.0001	0.136±0.007	0.44±0.06	0.13±0.05	0.25±0.07
21–31[a]	m	HR	0.360±0.025	141.73	<0.0001	0.040±0.015	9.22	0.0012	0.093±0.026	25.32	<0.0001	0.144±0.005	0.57±0.04	0.06±0.02	0.15±0.04

[a]The generation interval for line HR8 was 25 to 31 (see Table 1 and Fig. 2D).

Table 5. Components of variance (± SE; sensu Table 4; V_M, maternal genetic and covariance between V_A and V_M [COV_{AM}]) in univariate quantitative–genetic analyses of voluntary wheel running (treating revolutions run on days 5 and 6 as repeated measures of the same trait) in four nonselected control (C) lines and four lines artificially selected for high levels of voluntary wheel running (high runner or HR lines). Models were run for (A) the base population (generations −2 to 0) assuming homogeneous variance in females and males, (B) before and (C) after the selection limit (generations 21 to 31, except for line HR8,[a] see Table 4) with heterogeneous variance between selection groups (C vs. HR lines), and (C) before and (D) after the limit with heterogeneous variances for sexes and selection groups. The model allowing heterogeneous variance components for each sex did not converge in the base population. Also shown are the direct-maternal genetic correlation (r_{AM}) and total heritability (h_t^2; calculated as $h_t^2 = [V_A + 1.5 * COV_{AM} + 0.5 * V_M]/[V_{PE} + V_M + V_A + V_R]$, following Willham 1972).

			(co)Variance components							Ratios	
Gens	Sex	Group	V_{PE}±SE	V_A±SE	V_M±SE	COV_{AM}±SE	χ^2	P	V_R±SE	r_{AM}±SE	h_t^2±SE
A. Base population, pooled sexes											
−2–0			0.417±0.070	0.258±0.118	0.058±0.053	−0.066±0.077	0.79	0.3756	0.179±0.009	−0.537±0.381	0.207±0.070
B. Before the limit, sexes pooled, heterogeneous variance for selection group											
0–10		C	0.481±0.039	0.239±0.061	0.003±0.028	−0.002±0.035	0.00	0.9552	0.206±0.007	−0.071±1.057	0.256±0.043
0–10		HR	0.458±0.027	0.249±0.049	0.104±0.025	−0.108±0.030	17.28	0.0000	0.186±0.004	−0.675±0.093	0.138±0.030
C. After the limit, sexes pooled, heterogeneous variance for selection group											
21–31		C	0.452±0.042	0.206±0.056	0.051±0.029	−0.012±0.033	0.14	0.7090	0.197±0.007	−0.115±0.291	0.236±0.046
21–31[a]		HR	0.430±0.028	0.163±0.040	0.040±0.017	−0.028±0.023	1.60	0.2061	0.163±0.004	−0.346±0.215	0.177±0.035
D. Before the limit, heterogeneous variance for selection group and sexes											
0–10	f	C	0.447±0.091	0.325±0.149	0.079±0.064	−0.070±0.077	0.88	0.3482	0.270±0.013	−0.442±0.315	0.230±0.063
0–10	f	HR	0.463±0.048	0.340±0.086	0.018±0.057	−0.169±0.050	16.18	0.0001	0.239±0.007	−0.746±0.094	0.136±0.035
0–10	m	C	0.477±0.058	0.208±0.082	0.151±0.040	−0.036±0.066	0.34	0.5583	0.146±0.007	−0.580±0.506	0.192±0.075
0–10	m	HR	0.397±0.030	0.210±0.050	0.060±0.030	−0.064±0.037	3.31	0.0690	0.135±0.004	−0.569±0.186	0.180±0.046
E. After the limit, heterogeneous variance for selection group and sexes											
21–31	f	C	0.403±0.090	0.250±0.122	0.141±0.060	−0.062±0.064	1.00	0.3181	0.257±0.013	−0.331±0.251	0.216±0.062
21–31[a]	f	HR	0.485±0.051	0.166±0.071	0.080±0.055	−0.018±0.038	0.23	0.6328	0.190±0.006	−0.209±0.366	0.182±0.046
21–31	m	C	0.373±0.056	0.205±0.068	0.047±0.027	−0.031±0.058	0.30	0.5815	0.133±0.007	−0.243±0.372	0.251±0.085
21–31[a]	m	HR	0.360±0.030	0.123±0.038	0.017±0.023	−0.008±0.029	0.08	0.7785	0.135±0.005	−0.177±0.526	0.188±0.051

[a]The generation interval for line HR8 was 25 to 31 (see Table 1 and Fig. 2D).

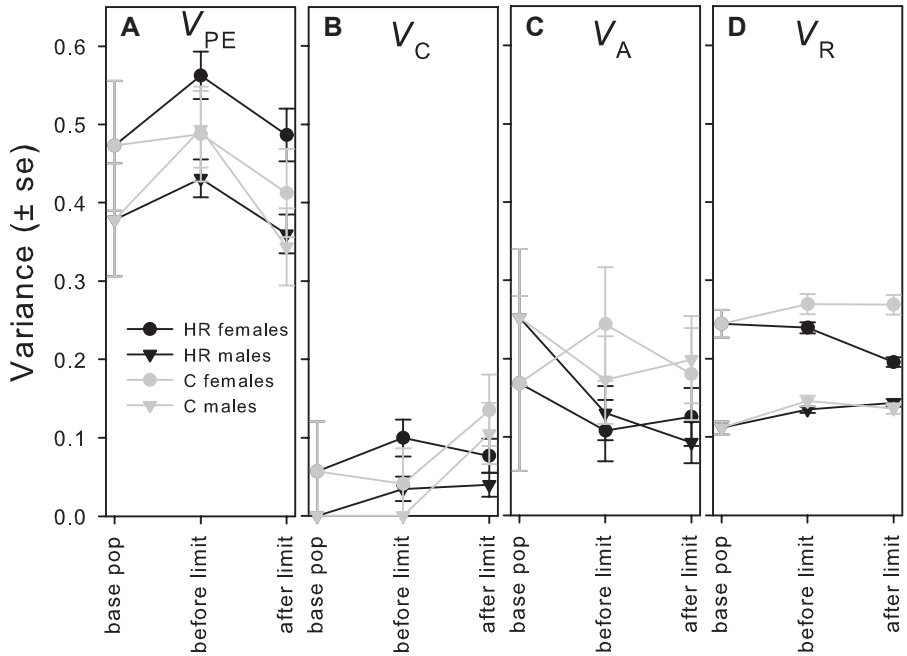

Figure 4. **Variance (± SE) in voluntary wheel running attributed to (A) permanent environmental (V_{PE}), (B) common environment (V_C), (C) additive genetic (V_A,), (D) and environmental (V_R) variances in univariate models allowing heterogeneous variance between sexes (females: circles; males: triangles) and between selection treatments (control lines C: gray; selected lines HR: black) in the base population, first 10 generations (before the selection limit), and from generation 21–31 (after the selection limit, except for line HR8 that reached the limit at generation 25: see Table 1 and Fig. 2D). Note that, as explained in the methods, wheel running within each generation was standardized to mean of 0 and variance of 1. See Figure S6 for a graphical representation of these components expressed as proportion of total phenotypic variance. These graphs show that V_A did not decline sufficiently to provide a general explanation for the selection limits (see Fig. 2).**

reproductive performance (smaller mice have smaller litters; Hayes et al. 1992). Another nonmutually exclusive mechanism would be via glucocorticoids (often referred to as "stress hormones"). Plasma corticosterone concentrations rise acutely in proportion to the amount of wheel-running activity in both C and HR mice (Girard and Garland 2002; Malisch et al. 2007; Malisch et al. 2009), which could have deleterious consequences on subsequent reproduction.

Indirect genetic effects are being increasingly recognized for their role in determining phenotypic responses to selection (Wolf et al. 1998). Räsänen and Kruuk (2007) suggested that selection experiments can be used to explicitly test how evolutionary responses may be affected by the presence of maternal effects. Although the negative COV_{AM} does not explain the selection limit per se, our study provides an example of effects of selection on maternal effects. Specifically, directional selection for wheel running seemed to increase V_M and the negative COV_{AM}, as both components became larger in HR and not C lines through the first 10 generations (Table 5A). After the limit, however, V_M was reduced and the COV_{AM} was no longer significantly different in HR versus C lines, perhaps because of the elimination of alleles with antagonistic effects in the HR lines.

Overall, our results suggest that a lack of V_A does not provide a general explanation for the observed selection limits, as most (six of eight) of the sex- and line-specific estimates of V_A in the HR lines after the limit were significantly greater than 0. These results add to the mixed empirical evidence from long-term experiments regarding the effect of consistent directional selection on V_A estimated with the animal model (see Introduction), with the distinction that we analyzed standing V_A in a behavioral trait before *and* after it had reached a selection limit. In *Drosophila serrata*, Hine et al. (2011) showed the existence of an evolutionary limit to selection for an increase in males attractiveness (a sexually selected trait), along with an increase in V_A that was partly associated with the increase in frequency of a partially recessive gene of major effect. By contrast, the two lines (i.e., HR3 and HR6) in which the recessive minimuscle allele (a gene of major effect) increased in frequency (from 0.067 in the base population; Garland et al. 2002) did not show such an increase in V_A nor any obvious difference with the two other HR lines. Aside from differences in modeling (e.g., use of pedigree back to base population, standardization of data), the variable results obtained from these (and other) studies indicate, unsurprisingly, that the change in V_A under selection depends on the population and trait considered.

Our result that V_A was significantly lower in HR than C lines, both before and after the selection limit—but still significantly greater than 0 after the limit—appear to be consistent with a "Bulmer effect" (Bulmer 1971). In R. A. Fisher's "infinitesimal model," formalized by Bulmer (1980), phenotypic traits are determined by an infinite number of unlinked loci, each with an infinitesimally small effect that is expressed additively at the phenotypic level. Bulmer (1971) divided V_A into two components: the "equilibrium genetic variance" and the "disequilibrium contribution" to show that directional selection induces a temporary negative disequilibrium contribution, and thus V_A decreases over the first few generations of selection (the "Bulmer effect"). As the effects of selection and recombination come to counterbalance each other, an equilibrium is reached where no further change in V_A occurs, at a level dependent on the selection intensity and h^2 of the trait (Bulmer 1971; Falconer and Mackay 1996). Considering the estimates of V_A and V_P for wheel running in the base population (i.e., 0.171 and 0.845, respectively) and the average proportion of the population selected over the first 10 generations (i.e., 0.189), the Bulmer effect should account for a reduction of 19.9% in V_A (calculated using eq. 9.46 in Bulmer 1980, assuming a harmonic-mean recombination rate of 0.4 and that all of the relevant loci were in linkage equilibrium in the base population). This contrasts with the 31.5% reduction in V_A over the first 10 generations. When the population size is finite, however, there is an additional reduction in V_A as inbreeding increases (Wei et al. 1996). Here, the average F increased from 0 to 0.134 $\pm$ 0.030 through the first 10 generations, which may explain why the reduction in V_A was greater than predicted from the Bulmer effect.

Several possibilities could explain the existence of selection limits when additive genetic variance remains (Falconer and Mackay 1996, p. 224), two of which may apply here. First, artificial selection should increase the frequency of dominant alleles tending to increase wheel running (Garland and Kelly 2006). However, any unfavorable recessive alleles will tend to remain at low frequencies for some number of generations. Consequently, at a selection limit dominance variance will contribute most to genetic variance (although some additive variance persists; see fig. 8.1 in Falconer and Mackay 1996) and unfavorable recessive alleles will lower the population trait mean when inbreeding occurs. Second, if overdominant loci contribute to wheel running, then at the selection limit overdominant genes would be in equilibrium at more-or-less intermediate frequencies, giving rise to nonadditive variance only (see fig. 8.1 in Falconer and Mackay 1996). Genetic dominance for high wheel running has been observed in crosses of house-mouse populations (Dohm et al. 1994; Nehrenberg et al. 2009; Kelly et al. 2010), but we did not detect dominance variance in wheel running either before or after the limits (Appendix S6). However, it must be noted that our breeding design makes this component hard to separate from V_C, V_{PE}, and V_A when inbreeding occurs. With overdominance, one expects increased within-family variance, which was indeed observed at generation 43 (Garland et al. 2011). Noteworthy in this regard is that a cross between lines HR7 and HR8 at generation 53 revealed heterosis for wheel running, but in males only (Hannon et al. 2011). Unfortunately, it is impossible to determine the exact nature of the observed selection limits, as this would require going back in time to conduct several key experiments (e.g., relaxing selection, reversing selection, analyzing line crosses) around the time the limits were reached.

Another way to explain the existence of selection limits is by consideration of physiological constraints that may prevent HR mice from running more (speed and/or duration) on a daily basis (Kolb et al. 2010). For example, the ability to dissipate heat generated by exercise might constrain wheel running (cf. Speakman and Król 2010). Availability of certain dietary lipids may also be constraining further evolutionary increases in wheel running, as suggested by the finding that a change in the diet (from normal chow to high-fat diet with added sucrose) increased distance run by as much as 75% in HR mice (Meek et al. 2010). Given that selection is applied on wheel running over the entire 24-h cycle, it is remarkable that the mice do not run during the day, even after the selection limit (Girard and Garland 2002; Malisch et al. 2009), perhaps due to constraints associated with the circadian rhythm and the need for sleep (Koteja et al. 2003; Rezende et al. 2009). Instead of being related to exercise abilities per se, the selection limits may be related to motivation. Perhaps the brain areas and/or signaling pathways (e.g., dopaminergic) involved in the reward received from wheel running reached a maximum level of possible activation. However, physiological and genetic constraints are not mutually exclusive, as they are two sides of the same coin. In other words, if a physiological constraint is reached, then it should be observable as a lack of V_A underlying this mechanism. For example, if the risk of hyperthermia limits wheel running, then it would imply that there is no V_A in the population for capacity to dissipate heat (e.g., body surface/volume ratio, fur density), store heat (which would be related to body composition), or tolerate a rise in body temperature.

Artificial selection represents a powerful, complimentary approach to the analysis of variation and evolution in natural populations (Blows and Hoffmann 2005; Räsänen and Kruuk 2007; Garland and Rose 2009). A useful feature of a selection experiment is to allow an empirical evaluation of the effect of sustained directional selection on key quantitative–genetic parameters, which are required to make accurate long-term predictions. Because multiple adaptive solutions should be more common for selected traits at relatively high levels of biological organization (Garland 2003; Careau and Garland 2012), more quantitative–genetic analyses from other selection studies that have targeted behavioral traits, such as thermoregulatory nesting (Lynch 1980)

and open-field behavior (DeFries et al. 1970), are needed to determine whether the type of trait selected (e.g., morphology, physiology, life-history, or behavior) impacts changes in key quantitative–genetic parameters and/or changes in the frequency of genes of major effect (Hine et al. 2011). Because populations in nature may exist at selection limits more frequently than is currently believed (Blows and Hoffmann 2005), this may prove crucial for our understanding of evolutionary processes.

ACKNOWLEDGMENTS

This project was supported by grants from National Science Foundation to TG, most recently (IOS-1121273), and to PC (EF-0328594), and from a Natural Sciences and Engineering Research Council postdoctoral fellowship to VC. The authors thank all the people who collected the data over the generations (e.g., J. G. Swallow), and colleagues with whom we discussed selection limits and analyses thereof (e.g., D. A. Roff, J. A. Endler). The authors also wish to thank S. Kane for her work on selection differentials and T. H. Meek for assistance with data checking. The authors thank two anonymous reviewers and M. Blows for comments on the initial version of the submitted manuscript.

LITERATURE CITED

Barton, N., and L. Partridge. 2000. Limits to natural selection. Bioessays 22:1075–1084.

Barton, N. H., and M. Turelli. 1987. Adaptive landscapes, genetic distance and the evolution of quantitative characters. Genet. Res. 49:157–173.

———. 1989. Evolutionary quantitative genetics: how little do we know. Annu. Rev. Genet. 23:337–370.

Beniwal, B. K., I. M. Hastings, R. Thompson, and W. G. Hill. 1992. Estimation of changes in genetic parameters in selected lines of mice using REML with an animal model. 1. Lean mass. Heredity 69:352–360.

Blows, M. W., and M. Higgie. 2003. Genetic constraints on the evolution of mate recognition under natural selection. Am. Nat. 161:240–253.

Blows, M. W., and A. A. Hoffmann. 2005. A reassessment of genetic limits to evolutionary change. Ecology 86:1371–1384.

Bulmer, M. G. 1971. Effect of selection on genetic variability. Am. Nat. 105:201–211.

———. 1980. The mathematical theory of quantitative genetics. Oxford Univ. Press, Oxford, U.K.

Bünger, L., U. Renne, G. Dietl, and S. Kuhla. 1998. Long-term selection for protein amount over 70 generations in mice. Genet. Res. 72:93–109.

Butler, D., B. R. Cullis, A. R. Gilmour, and D. J. Gogel. 2007. ASReml-R reference manual release 2.0. VSN International Ltd., Hemel Hempstead, U.K.

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

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

Coyne, J. A. 1987. Lack of response to selection for directional asymmetry in *Drosophila melanogaster*. J. Hered. 78:119–119.

DeFries, J. C., J. R. Wilson, and G. E. McClearn. 1970. Open-field behavior in mice: selection response and situational generality. Behav. Genet. 1:195–211.

Dewsbury, D. A. 1980. Wheel-running behavior in 12 species of muroid rodents. Behav. Process. 5:271–280.

Dohm, M. R., C. S. Richardson, and T. Garland, Jr. 1994. Exercise physiology of wild and random-bred laboratory house mice and their reciprocal hybrids. Am. J. Physiol.-Reg. I 36:R1098–R1108.

Dominicus, A., A. Skrondal, H. K. Gjessing, N. L. Pedersen, and J. Palmgren. 2006. Likelihood ratio tests in behavioral genetics: problems and solutions. Behav. Genet. 36:331–340.

Falconer, D. S. 1955. Patterns of response in selection experiments with mice. Cold Spring Harb. Sym. 20:178–196.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics, 4th edn. Longman: Harlow, U.K.

Flux, J. E. C., and M. M. Flux. 1982. Artificial selection and gene flow in wild starlings, *Sturnus vulgaris*. Naturwissenschaften 69:96–97.

Garland, T., Jr. 2003. Selection experiments: an underutilized tool in biomechanics and organismal biology. Pp. 23–56 *in* V. L. Bels, J.-P. Gasc, and A. Casinos, eds. Vertebrate biomechanics and evolution. BIOS Scientific Publishers, Oxford, U.K.

Garland, T., Jr., and S. A. Kelly. 2006. Phenotypic plasticity and experimental evolution. J. Exp. Biol. 209:2344–2361.

Garland, T., Jr., and M. R. Rose. 2009. Experimental evolution: concepts, methods, and applications of selection experiments. University of California Press, Berkeley, CA.

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

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

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

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

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

Hannon, R. M., T. H. Meek, W. Acosta, R. C. Maciel, H. Schutz, and T. Garland, Jr. 2011. Sex-specific heterosis in line crosses of mice selectively bred for high locomotor activity. Behav. Genet. 41:615–624.

Hayes, J. P., T. Garland, Jr., and M. R. Dohm. 1992. Individual variation in metabolism and reproduction of *Mus*: are energetics and life-history linked? Funct. Ecol. 6:5–14.

Heath, S. C., G. Bulfield, R. Thompson, and P. D. Keightley. 1995. Rates of change of genetic parameters of body weight in selected mouse lines. Genet. Res. 66:19–25.

Henderson, C. R. 1973. Sire evaluation and genetic trends. J. Anim. Sci. 76:10–41.

Hendry, A. P., and M. T. Kinnison. 1999. The pace of modern life: Measuring rates of contemporary microevolution. Evolution 53:1637–1653.

Hill, W. G., and A. Caballero. 1992. Artificial selection experiments. Annu. Rev. Ecol. Syst. 23:287–310.

Hill, W. G., and S. H. Mbaga. 1998. Mutation and conflicts between artificial and natural selection for quantitative traits. Genetica 102–3:171–181.

Hine, E., K. McGuigan, and M. W. Blows. 2011. Natural selection stops the evolution of male attractiveness. Proc. Natil. Acad. Sci. USA 108:3659–3664.

Keightley, P. D. 1998. Genetic basis of response to 50 generations of selection on body weight in inbred mice. Genetics 148:1931–1939.

Kelly, S. A., D. L. Nehrenberg, J. L. Peirce, K. Hua, B. M. Steffy, T. Wiltshire, F. Pardo Manuel de Villena, T. Garland, Jr., and D. Pomp. 2010. Genetic architecture of voluntary exercise in an advanced intercross line of mice. Physiol. Genom. 42:190–200.

Kelly, S. A., D. L. Nehrenberg, K. Hua, T. Garland, Jr., and D. Pomp. 2011. Exercise, weight loss, and changes in body composition in mice: phenotypic relationships and genetic architecture. Physiol. Genom. 43:199–212.

Kirkpatrick, M. 2009. Patterns of quantitative genetic variation in multiple dimensions. Genetica 136:271–284.

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

Koteja, P., J. G. Swallow, P. A. Carter, and T. Garland, Jr. 2003. Different effects of intensity and duration of locomotor activity on circadian period. J. Biol. Rhythms 18:491–501.

Kruuk, L. E. B., T. H. Clutton-Brock, J. Slate, J. M. Pemberton, S. Brotherstone, and F. E. Guinness. 2000. Heritability of fitness in a wild mammal population. Proc. Natil. Acad. Sci. USA 97:698–703.

Lerner, I. M., and E. R. Dempster. 1951. Attenuation of genetic progress under continued selection in poultry. Heredity 5:75–94.

Lynch, C. B. 1980. Response to divergent selection for nesting behavior in *Mus musculus*. Genetics 96:757–765.

———. 1994. Evolutionary inferences from genetic analyses of cold adaptation in laboratory and wild populations of the house mouse. Pp. 278–301 *in* C. R. B. Boake, ed. Quantitative genetic studies of behavioral evolution. University of Chicago Press, Chicago.

Lynch, M., and J. B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA.

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

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

Martinez, V., L. Bunger, and W. G. Hill. 2000. Analysis of response to 20 generations of selection for body composition in mice: fit to infinitesimal model assumptions. Genet. Sel. Evol. 32:3–21.

Mather, J. G. 1981. Wheel-running activity: a new interpretation. Mammal Rev. 11:41–51.

Mayr, E. 1961. Cause and effect in biology. Science 134:1501–1506.

Meek, T. H., J. C. Eisenmann, and T. Garland. 2010. Western diet increases wheel running in mice selectively bred for high voluntary wheel running. Int. J. Obesity 34:960–969.

Meyer, K., and W. G. Hill. 1991. Mixed model analysis of a selection experiment for food intake in mice. Genet. Res. 57:71–81.

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

Morrissey, M. B., D. J. Parker, P. Korsten, J. M. Pemberton, L. E. B. Kruuk, and A. J. Wilson. 2012. The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder's equation. Evolution 66:2399–2410.

Mousseau, T. A., and C. W. Fox. 1998. Maternal effects as adaptations. Oxford Univ. Press, Oxford, U.K.

Mousseau, T. A., and D. A. Roff. 1987. Natural selection and the heritability of fitness components. Heredity 59:181–197.

Muggeo, V. M. R. 2008. Segmented: an R package to fit regression models with broken-line relationships. R News 8:20–25.

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

Perrigo, G., and F. H. Bronson. 1985. Sex differences in the energy allocation strategies of house mice. Behav. Ecol. Sociobiol. 17:297–302.

Postma, E., J. Visser, and A. J. Van Noordwijk. 2007. Strong artificial selection in the wild results in predicted small evolutionary change. J. Evol. Biol. 20:1823–1832.

R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at http://www.R-project.org.

Räsänen, K., and L. E. B. Kruuk. 2007. Maternal effects and evolution at ecological time-scales. Funct. Ecol. 21:408–421.

Reeve, J. P., and D. J. Fairbairn. 1996. Sexual size dimorphism as a correlated response to selection on body size: an empirical test of the quantitative genetic model. Evolution 50:1927–1938.

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

Roberts, R. C. 1966. Limits to artificial selection for body weight in mouse. II. Genetic nature of limits. Genet. Res. Camb. 8:361–375.

Rundquist, E. A. 1933. Inheritance of spontaneous activity in rats. J. Comp. Psychol. 16:415–438.

Self, S. C., and K. Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. J. Am. Stat. Soc. 82:605–610.

Sorensen, D. A., and B. W. Kennedy. 1984. Estimation of genetic variances from unselected and selected populations. J. Anim. Sci. 59:1213–1223.

Speakman, J. R., and E. Król. 2010. Maximal heat dissipation capacity and hyperthermia risk: neglected key factors in the ecology of endotherms. J. Anim. Ecol. 79:726–746.

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

Thompson, R. 1973. Estimation of variance and covariance components with an application when records are subject to culling. Biometrics 29:527–550.

van der Werf, J. H. J., and I. J. M. de Boer. 1990. Estimation of additive genetic variance when base populations are selected. J. Anim. Sci. 68:3124–3132.

Walsh, B., and M. W. Blows. 2009. Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. Annu. Rev. Ecol. Evol. S 40:41–59.

Weber, K. E. 1996. Large genetic change at small fitness cost in large populations of *Drosophila melanogaster* selected for wind tunnel flight: Rethinking fitness surfaces. Genetics 144:205–213.

Wei, M., A. Caballero, and W. G. Hill. 1996. Selection response in finite populations. Genetics 144:1961–1974.

Willham, R. L. 1972. Role of maternal effects in animal breeding. 3. Biometrical aspects of maternal effects in animals. J. Anim. Sci. 35:1288–1293.

Wilson, A. J., and D. Réale. 2006. Ontogeny of additive and maternal genetic effects: lessons from domestic mammals. Am. Nat. 167:E23–E38.

Wolak, M. E. 2012. nadiv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. Methods Ecol. Evol. 3:792–796.

Wolf, J. B., E. D. Brodie, J. M. Cheverud, A. J. Moore, and M. J. Wade. 1998. Evolutionary consequences of indirect genetic effects. Trends Ecol. Evol. 13:64–69.

Associate Editor: M. Blows

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Animals, housing, and selection protocol.
Appendix S2. Data and pedigree.
Appendix S3. Parent–offspring regressions.
Appendix S4. Animal models with different pedigree.
Appendix S5. Control-adjusted means and seasonal variations.
Appendix S6. Dominance variance.
Appendix S7. Sex-specific ANCOVAs and ANOVAs.
Appendix S8. Litter size at weaning.
Appendix S9. Cross-sex genetic correlations.
Appendix S10. Proportion of variance.
Appendix S11. Line-specific estimates.
Appendix S12. Rate versus duration of response.

Supplementary Materials and Methods

ORIGINAL ARTICLE

LIMITS TO BEHAVIORAL EVOLUTION: THE QUANTITATIVE GENETICS OF A COMPLEX TRAIT UNDER DIRECTIONAL SELECTION

Vincent Careau[1,2], Matthew E. Wolak[1], Patrick A. Carter[3], and Theodore Garland, Jr.[1,*]

[1]Department of Biology, University of California, Riverside, CA 92521
[2]Present address: Centre for Integrative Ecology, Deakin University, 75 Pigdons Road, Waurn Ponds, Victoria 3216, Australia
[3]School of Biological Sciences, Washington State University, Pullman, WA 99164
*address for correspondence:
Theodore Garland, Jr., Department of Biology University of California, Riverside, Riverside CA 92521, Phone: (951) 827-3524; Fax: (951) 827-4286; Email: tgarland@ucr.edu

Table of Contents

APPENDIX S1

ANIMALS, HOUSING, AND SELECTION PROTOCOL

The base population for this selection experiment was the albino outbred Hsd:ICR strain. Further information on the history and maintenance of this strain can be found in previous publications (Dohm et al. 2001; Girard et al. 2002). A total of 224 individuals were purchased from Harlan Sprague Dawley (designated as generation=-2), subjected to two generations of random matings at the University of Wisconsin-Madison (generations -1 and 0; Swallow et al. 1998), and then separated into four non-selected control (C) lines (lab designations #C1, C2, C4, and C5) and four selected "high-runner" (HR) lines (#HR3, HR6, HR7, and HR8). Each line was 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 were maintained on a 12-h light/12-h dark cycle (lights on 0700 h), which is also maintained during the wheel-running trials.

Each generation, approximately 600 mice of 6-8 weeks of age were monitored for wheel revolutions for 6 consecutive days (testing occurred at 5-7 weeks of age in generation -1 and 0). Voluntary wheel running was 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). Three batches of approximately 200 mice each were measured during 3 successive weeks. Mice from a given batch were weighed and placed randomly on wheels during the morning of the first day; data collection was started at ~1300 h. Data were downloaded every 23.5 h, at which time wheels were checked to remove any food pellets or wood shavings and to ensure freedom of rotation. On the sixth day, mice were removed from the wheels and weighed. Although

wheel running was monitored around the clock, it occurs primarily during the dark phase in both C and HR lines (Malisch et al. 2009).

The selected trait was the residual average number of wheel revolutions run on days 5 and 6 (transformed as necessary to improve normality of residuals) from a multiple regression model used to control for several background variables, including measurement block (batches 1-3 and rooms 1-2), sex (in interaction with selection history), and family (see Swallow et al. 1998). The first four days of wheel access were not included as part of the selection criterion to avoid selecting on neophilia (attraction to novel objects). Within-family selection was used both to reduce the rate of inbreeding and to avoid possible complications of common environmental effects, such as maternal effects (e.g., see DeFries et al. 1970; Falconer 1973; Lynch 1980; Falconer and Mackay 1996). To produce offspring in the next generation for each of the four HR lines, the highest-running male and the highest-running female from each family were chosen to breed and paired randomly, with the provision of no sibling mating. Typically 2-3 extra pairs were made to ensure that 10 families were produced for each line. The additional males and females were generally the second-highest runners from the highest-running families, with the stipulation that no two of the additional mice came from the same family. In C lines, a male and female from within each family are chosen randomly to obtain breeders (some additional individuals were also retained to make “backup” pairings).

APPENDIX S2

DATA AND PEDIGREE

Wheel-running data from the base population through the first 31 generations of the experiment were compiled from raw data files using SPSS (v8.0). We stopped compiling data after generation 31 because at this point the colony was moved from the University of Wisconsin-Madison to the University of California, Riverside. This necessarily involved environmental changes that may have affected wheel running to an unknown extent, and in addition it was not possible to record wheel running for four generations.

Outliers were removed by a visual inspection of a graph comparing body mass before vs. after wheel-running, as a major change ($\geq$7g) in body mass over 6 days may indicate a typographical or biological problem (e.g., the individual was sick, a water bottle became air-locked). We also excluded all individuals who ran <0.4x less or >2.5x more on day 6 relative to day 5. Such an asymmetrical threshold is justified by the fact that wheel running increases constantly from day 1 to 6 in both C and HR lines (Garland and Kelly 2006). A total of 222 individuals were excluded by this procedure, consisting of 1.2% of the total number of initial samples (total n=17,988). This dataset is available in Dryad (doi:xx).

The pedigree of the entire population was re-constructed using dam and sire identity as recorded at weaning (available in Dryad; doi:xx). Potential numbering errors were checked in SPSS when merging the generational data files and then using Pedigree Viewer (Kinghorn and Kinghorn 2002). In the process, we found two events of accidental line mixing at generation 25 (an individual from C4 was bred as a HR3, and an individual

98 from C5 was bred as a C4). This line mixing should be taken into account by our use of a
99 complete pedigree in the animal model analyses (Pollak and Quaas 1983).
100 Also available in Dryad (doi:xx) are descriptive statistics (N=number of
101 individuals, mean, variance, sd=standard deviation, range, CV=coefficient of variation)
102 of voluntary wheel running (in revolutions per day) for four control lines (linetype=C)
103 and four selected lines (linetype=HR) as a whole and for the subset of individuals that
104 were selected for reproduction in each generation. These statistics were used to calculate
105 the selection differential (s), the standardized selection differential (s'), the cumulative
106 standardized selection differential (acc s'), the average of the four control lines at each
107 generation (Mean controls), and the selective gain relative to the average of the four
108 control lines at each generation (gain vs. control).

APPENDIX S3

DATA ANALYSIS – OFFSPRING-ON-SIRE REGRESSIONS

To establish a baseline estimate of h^2 before selection was applied, we first estimated h^2 using either the sire (h^2_{os}), dam (h^2_{od}), or midparent (h^2_{omp}) values in the base population. In the present experiment, males contributed no parental care because they were removed from the dams prior to parturition. In this case, twice the slope of the offspring-on-sire (OS) regression should give an estimate of narrow-sense h^2 that is not biased by maternal common environment effects (Lynch and Walsh 1998).

We first calculated the residual values of wheel running (average of days 5 and 6) line-by-line in a model that included several fixed effects fitted within generation [sex, age, inbreeding coefficient (F), and measurement block (batches 1-3 and rooms 1-2)]. The residuals of this model were used in the OS regression. We created a function "osw.R" that performs a regression weighted for litter size using an iterative process as described in Lynch and Walsh (1998, p541). The weight (w_i) of the ith family is calculated as $w_i=n_i/[n_i(t–B)+(1–t)]$, where n_i is the number of mice that were wheel-tested in that family, t the intraclass correlation coefficient (for that line at that generation), and B is the slope squared in the case of single parent regression, or simply the slope squared divided by two in the case of mid-parent regression. Some combinations of t and slopes yielded negative w_i, in which case uncorrected weighting was applied (i.e., using only w_i=simple family size, the number of offspring measured in that family).

The offspring-on-sire (OS) regression is not biased by selection of parents when that selection is based on the character whose h^2_{os} is being estimated (Falconer and Mackay 1996). However, the OS regression can be affected by assortative mating and

inbreeding (Falconer and Mackay 1996; Gibson 1996). Assortative mating should not occur in our population because individuals selected for breeding were mated randomly, other than disallowing mating of sibs. Indeed, the average correlation between the phenotypic values of mated sires and dams (r_{sd}) fluctuated randomly around a mean±se of 0.082±0.066, 0.019±0.066, 0.021±0.071, 0.003±0.060, 0.042±0.060, -0.029±0.056, -0.056±0.052, and 0.010±0.061 in line 1 to 8, respectively, and did not change across the 32 generations (linear regression of r_{sd} on generation: all eight lines, $t_{30}<0.851$; $P>0.402$). As F increased by 1.18% each generation (see main text), we corrected the weighted slopes of the OS regression (b_{os}) for the average F in the offspring generation following Nyquist (1991) ($b_F=b_{os}/[1+F(1-b_{os})]$). This correction assumes pure additive gene action and no gene-by-environment interaction (Gibson 1996). We tested whether b_F changed across generations using simple linear regression on generation number, weighting each b_F by the inverse of the standard error of b_{os}. We also tested whether the change in b_F was different among lines using ANCOVA with b_F as the dependent variable, generation (as a continuous variable), and line as the independent variables in addition to an interaction between the two. To give an overall h^2_{os} estimate in each of the C lines, we averaged all h^2_{os} estimates covering generations 0 to 31 and calculated the standard error of these 32 numbers. For each HR line, this procedure was implemented for generations 0 to 10 as well as through the selection limit and after the selection limit as determined by segmented regressions (see main text). The individuals that were accidentally transferred from one line to another at generation 25 were left out of the OS regressions.

OFFSPRING-ON-PARENT REGRESSIONS - BASE POPULATION

Before selection was applied, wheel running was measured in 188 and 557 individuals of generations -1 and 0, respectively. The slopes of the offspring-on-sire, offspring-on-dam, and offspring-on-midparent regressions were 0.143±0.058 (n=80), 0.105±0.044 (n=79), and 0.173±0.064 (n=78), respectively (sample sizes are slightly different because sometimes wheel running was only measured for one parent). Hence, the h^2_{os}, h^2_{od}, and h^2_{omp} estimates were 0.286, 0.209, and 0.173, respectively. There was a small positive correlation between the sire and dam phenotypes (r_{sd}=0.194, $d.f.$=76, P=0.089), which explains why h^2_{omp} is lower than h^2_{os} and h^2_{od} (the separate regression slopes for two correlated independent variables are higher than the slope for their mean; see also Falconer and Mackay 1996). The fact that h^2_{os} is higher than h^2_{od} may indicate the presence of a negative covariance between additive and maternal genetic effects (COV_{AM}; see main text).

OFFSPRING-ON-SIRE REGRESSIONS - BEFORE AND AFTER THE LIMIT

In the four replicate C lines, the weighted slopes of the OS regression corrected for inbreeding (b_F) ranged from -0.445 to 0.734 and did not systematically change across the 31 generations (linear regressions of b_F on generation: all four t_{30}<1.09; P>0.28). There was no systematic difference in the change of b_F estimates across generations among the replicate C lines, as indicated by a non-significant interaction term (generation by line) in an ANCOVA ($F_{3,120}$=0.36, P=0.780).

In the four replicate HR lines, b_F ranged from -0.611 to 0.611 and did not systematically changed across the 31 generations (linear regression of b_F vs. generation: all four t_{30}<1.734; P>0.093). However, there was some heterogeneity among the four

178 replicate HR lines in how b_F changed through the experiment, as suggested by a
179 marginally non-significant interaction term (generation by line) in an ANCOVA
180 ($F_{3,120}$=2.15, P=0.098; Fig. S1). In three of the HR lines, b_F decreased with generation
181 (estimates±se are -0.00295±0.00248, -0.00395±0.00247, and -0.00169±0.00282; Fig.
182 S1a-c). In line HR8, however, there was a tendency toward an increasing b_F with
183 generation (0.00433±0.00250; t_{30}=1.734; P=0.093; Fig. S1d). A close look at the
184 ANCOVA (for completeness) revealed that the change in b_F across generation was
185 significantly different in line HR8 vs. line HR3 (t_{120}=1.988; P=0.049), line HR6
186 (t_{120}=2.321; P=0.022), but not line HR7 (t_{120}=1.665; P=0.099). There were no other
187 significant differences in b_F among HR lines (for all 3 other contrasts: t_{36}<0.626; P>0.53
188 [not adjusted for multiple comparisons]).
189 The mean (±se) h^2_{os} (i.e., twice the b_F) estimated for HR lines up to generation 10,
190 before, and after the limit are shown in Table S1. The correspondence between h^2_{os} and
191 realized heritability (h^2_w) was relatively good, and the Pearson product moment
192 correlation (r) between the line-specific estimates was 0.96 (n=4; P=0.036) up to
193 generation 10 and 0.95 (n=4; P=0.050) after the limit (Table S1). The mean (±se)
194 estimated through generation 31 for lines C1, C2, C4, and C5, was 0.123±0.065,
195 0.087±0.052, 0.012±0.071, and 0.153±0.061, respectively. When averaged over all C
196 lines, the h^2_{os} is 0.092±0.031. The reason why h^2_{os} was slightly lower in C lines, other
197 than a sampling variance issue, is unclear.
198 There was an obvious discrepancy between the OS regression and the animal
199 model after the limit when looking at each HR line specifically: the rank order is reversed
200 (see Table S1 and Table S5). Again, the reason for this discrepancy is unclear, but

estimates from the animal model are likely more reliable as they incorporate all the information among relatives (Kruuk 2004; Åkesson et al. 2008).

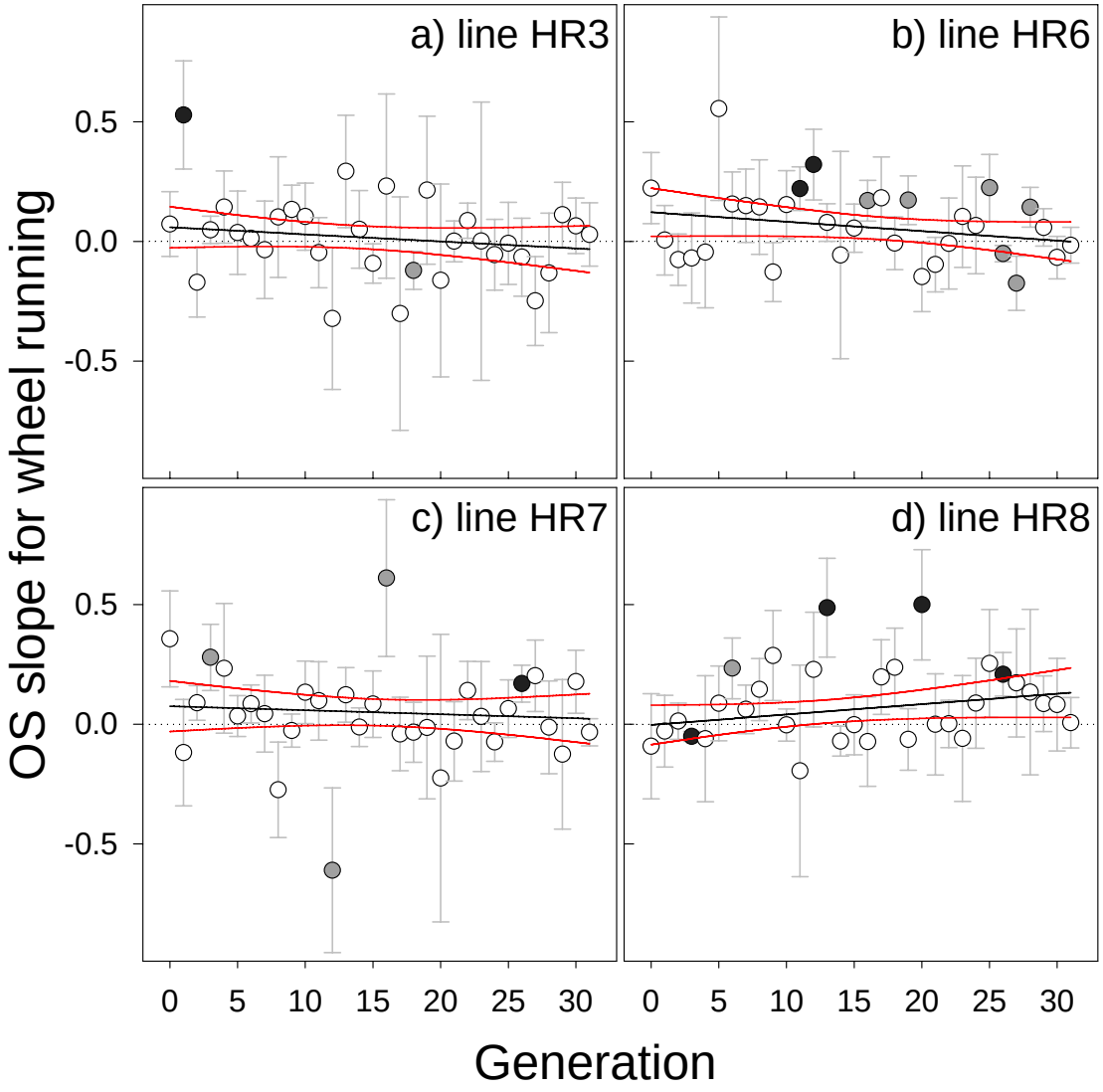

Figure S1. Slopes of the offspring-on-sire regressions (corrected for inbreeding) as a function of generation in each of the four replicate lines artificially selected for high voluntary wheel running (HR: high runner). Fitted lines were weighted according to the standard error of the slope. Red lines show 95% confidence intervals. The residual degrees of freedom in regression models to estimate the slope at each generation ranged from 4 to 10 (mean±sd = 7.91±0.81). Circle color denotes the *P*-value of the slope (black: $P<0.05$; grey: $P<0.10$; white: $P>0.10$).

Table S1. Heritability estimates (h^2_{os}: offspring-on-sire; h^2_{od}: offspring-on-dam, and h^2_{omp}: offspring-on-midparent) for wheel running (mean revolutions/day on days 5 & 6 of a 6-day period of wheel access) estimated (a) through generation 10, (b) through the selection limit, and (c) after the selection limit in each of the four replicate lines of mice artificially selected for high levels of voluntary wheel running. Also shown are the generations over which h^2_w was estimated (gens), the realized heritability (h^2_w; for within-family selection), and the realized heritability expected in a mass selection experiment (h^2_r).

Line	Gens	h^2_{os} ± se	h^2_{od} ± se	h^2_{omp} ± se	h^2_w ± se	h^2_r ± se
a. Through generation 10						
HR3	0-10	0.177 ± 0.103	-0.042 ± 0.094	0.035 ± 0.076	0.191 ± 0.027	0.288 ± 0.032
HR6	0-10	0.194 ± 0.117	0.258 ± 0.150	0.243 ± 0.138	0.253 ± 0.023	0.381 ± 0.027
HR7	0-10	0.152 ± 0.108	0.013 ± 0.093	0.010 ± 0.066	0.178 ± 0.029	0.269 ± 0.034
HR8	0-10	0.108 ± 0.075	-0.052 ± 0.086	0.033 ± 0.077	0.112 ± 0.022	0.168 ± 0.026
Mean HR		0.158 ± 0.019	0.044 ± 0.073	0.080 ± 0.054	0.183 ± 0.029	0.277 ± 0.044
b. Through limit						
HR3	0-18	0.070 ± 0.079	0.055 ± 0.083	0.085 ± 0.055	0.184 ± 0.018	0.277 ± 0.022
HR6	0-19	0.220 ± 0.069	0.240 ± 0.079	0.245 ± 0.081	0.253 ± 0.019	0.382 ± 0.022
HR7	0-21	0.067 ± 0.088	0.089 ± 0.072	0.064 ± 0.048	0.200 ± 0.013	0.302 ± 0.015
HR8	0-25	0.163 ± 0.063	0.016 ± 0.050	0.084 ± 0.042	0.142 ± 0.011	0.214 ± 0.013
Mean		0.130 ± 0.075	0.100 ± 0.049	0.119 ± 0.042	0.195 ± 0.023	0.294 ± 0.027
c. After limit						
HR3	19-31	-0.025 ± 0.068	0.287 ± 0.069	0.140 ± 0.044	-0.035 ± 0.077	-0.053 ± 0.091
HR6	20-31	0.006 ± 0.069	0.082 ± 0.075	0.128 ± 0.107	-0.018 ± 0.064	-0.027 ± 0.076
HR7	22-31	0.108 ± 0.073	0.050 ± 0.144	0.083 ± 0.042	0.074 ± 0.087	0.112 ± 0.103
HR8	26-31	0.230 ± 0.059	-0.005 ± 0.091	0.079 ± 0.072	0.095 ± 0.224	0.143 ± 0.264
Mean		0.080 ± 0.058	0.103 ± 0.064	0.107 ± 0.015	0.029 ± 0.033	0.044 ± 0.038

APPENDIX S4

ANIMAL MODELS WITH DIFFERENT PEDIGREE INFORMATION

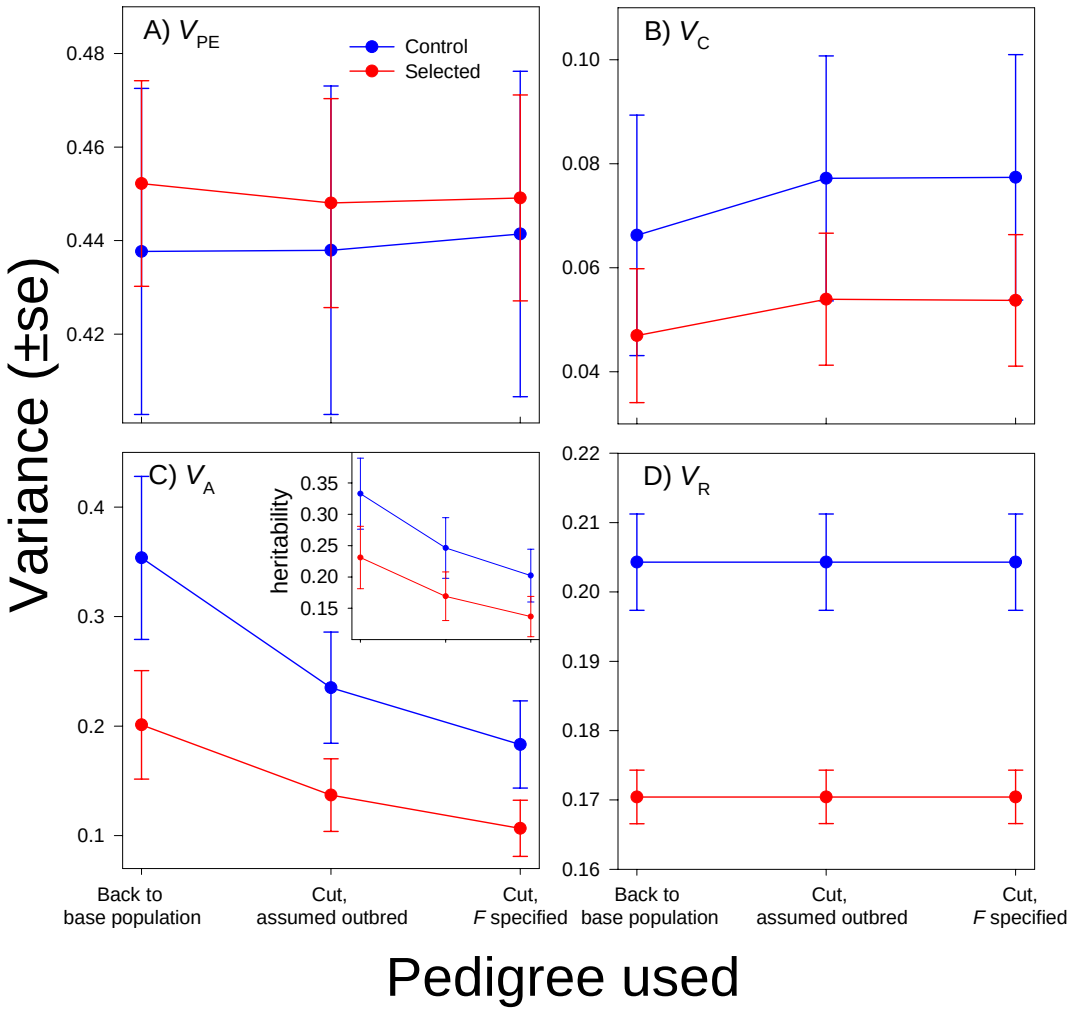

Figure S2. Variance components (±se; V_{PE}: permanent environmental variance [individual identity]; V_C: common environmental variance [dam identity]; V_A: additive genetic variance; V_R: residual variance) in voluntary wheel running using data after the selection limit (i.e., from generations 25 to 31 for line HR8 and 21 to 31 for all other lines) and A-inverse relationship matrices produced using different pedigree information. First, the A-inverse matrix was made from the complete pedigree (i.e., going "back to base population", beginning with generation -2). Second, the A-inverse matrix was made assuming that individuals of generations 21 (except for line HR8, generation 25) were unrelated and had an inbreeding coefficient (F) of zero ("cut, assumed outbred"). Third, the A-inverse was made assuming that animals of generation 21 were unrelated, but their exact F (calculated using the complete pedigree) was specified ("cut, F specified"; the variance components of this model are presented in Table 4d).

APPENDIX S5

CONTROL-ADJUSTED MEANS AND SEASONAL VARIATIONS

For visualization, we plotted wheel running as function of generation after subtracting the average of the control lines from each HR line (Fig. S3). The total phenotypic variance increased with the gradual increase of trait means across generations in the HR lines (the correlations between the mean and variance from generation 0 to 31 in HR3, HR6, HR7, and HR8 were 0.77, 0.76, 0.68, and 0.81 in females and 0.81, 0.77, 0.72, and 0.82 in males).

The presence of seasonal variations in wheel running was assessed in two steps for each line separately. First, we corrected wheel running for several variables (sex, age, measurement block [batches 1-3, rooms 1-2], date, and z-transformed date^2) and saved the residuals from this model (Fig. S4a). Second, we averaged the residuals of wheel running for each month, which revealed that wheel running is highest during winter and lowest in summer in both C and HR lines (Fig. S4b).

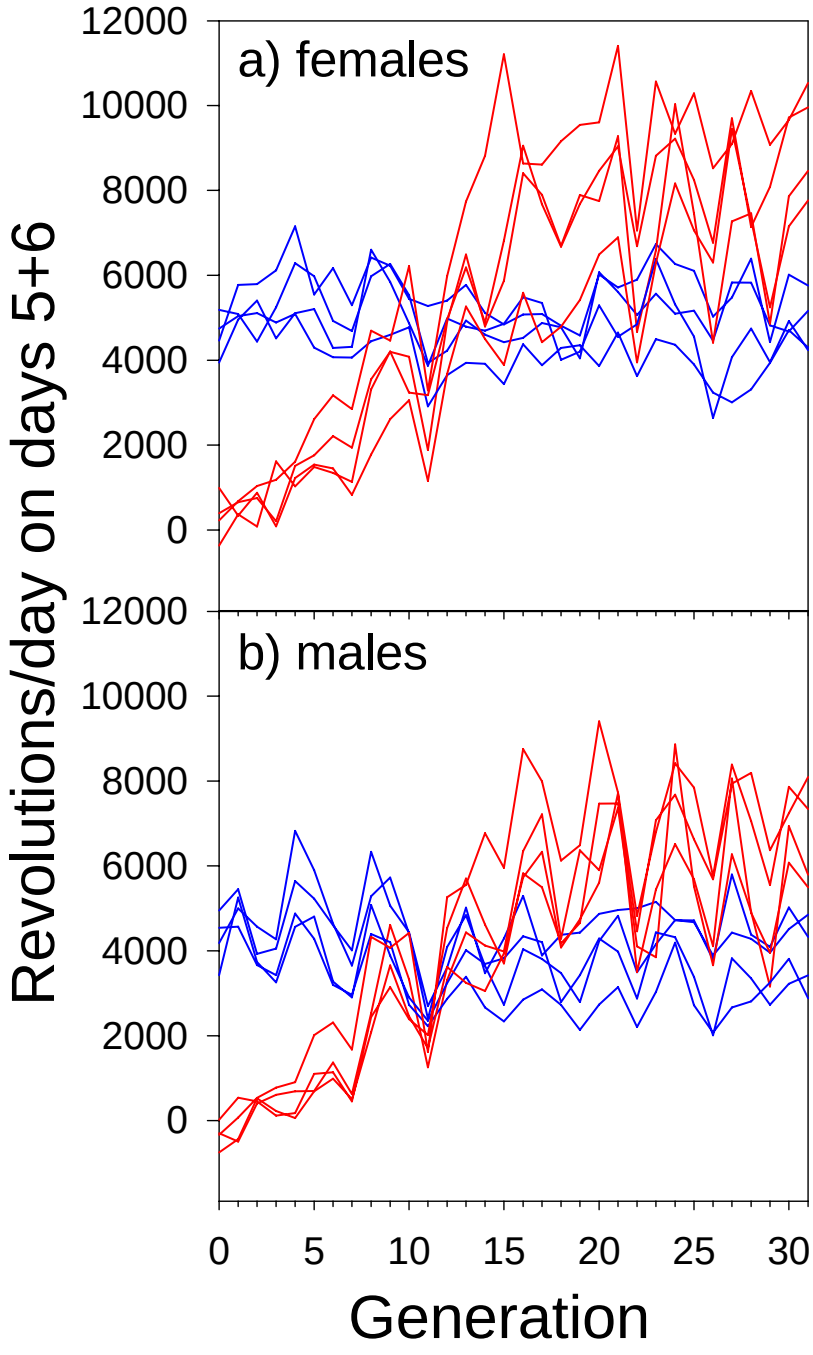

Figure S3. Average running distance (number of revolutions run on days 5 and 6) in (a) female and (b) male mice from each of the four control lines (blue lines) and the four selected lines (red lines; adjusted for the mean running distance in control lines) plotted against generation. Note: the total phenotypic variance increased with the gradual increase of trait means across generations in the HR lines (the correlations between the mean and variance from generation 0 to 31 in HR3, HR6, HR7, and HR8 were 0.77, 0.76, 0.68, and 0.81 in females and 0.81, 0.77, 0.72, and 0.82 in males).

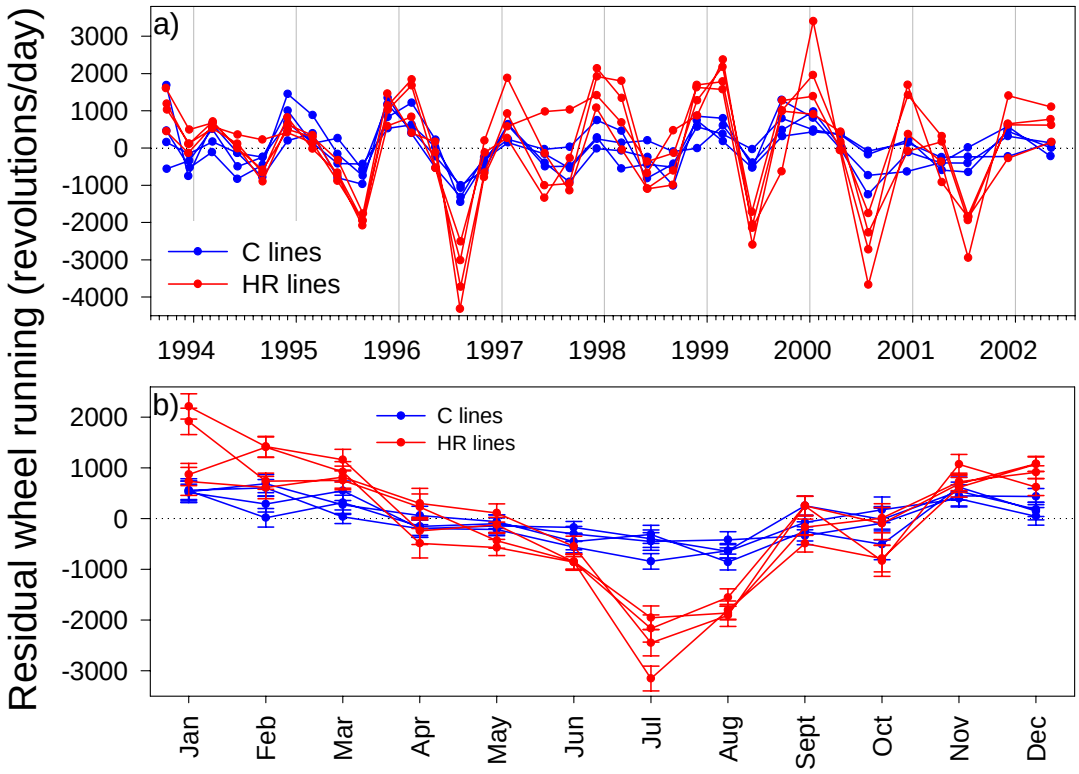

Figure S4. Residual wheel running in the four control (C; blue lines) and four selected lines (High Runner; HR, red lines) from the base population (i.e., generation -2) up to generation 31 plotted as function of (a) date and (b) month.

APPENDIX S6

DOMINANCE VARIANCE

We estimated dominance genetic variance (V_D) in our models using the inverse of the dominance genetic relationship matrix ($\mathbf{D^{-1}}$) computed with the "nadiv" package in R (Wolak 2012). As the computing requirements to fit V_D into the models were very large, we had to run one model before and one after the limit for each of the eight lines (16 models). However, models with V_D never fit the data better (assessed via log-likelihood ratio tests, see main text) than reduced models excluding this random effect. Furthermore, V_D was estimated at the lower boundary of the parameter space in 13 of 16 models (Table S2). Subsequent analyses were thus conducted without V_D in the model. This allowed us to run a single model for all lines.

It must be noted, however, that our breeding design makes V_D hard to estimate. Aside from full sibs, which also share a common environment at least until weaning, the only other type of relationship present in our pedigree for which dominance effects are expressed are double first cousins, which were relatively uncommon. More specifically, the breeding design in HR lines yielded a total of 34 and 67 pairs of families producing double first cousins before and after the limit, respectively, out of a total of 532 and 482 families.

Table S2. Components of variance (±se; V_{PE}, permanent environmental; V_C, common environmental; V_A, additive genetic; V_D, dominance; V_R, environmental variance) in univariate quantitative genetic analyses of voluntary wheel running in four control lines of randomly bred mice (C lines) and four lines artificially selected for high levels of voluntary wheel running (High-runner, HR lines). Statistical significance of each variance component was tested using a log-likelihood ratio statistics of the full model vs. a reduced that excluded the variance component of interest. Wheel-running data were standardized (mean=0, variance=1) within each generation for each line, such that they are relatively comparable (depending on the proportion of variance explained by the fixed effects, which were fitted within generation). Pooled estimates are also provided with line fitted within generation as an additional fixed effect (computational requirement made it impossible to include V_D in these models). Also shown are the ratios (proportion of variance explained) of V_{PE}, V_C, V_A, and V_D over the sum of V_{PE}, V_C, V_A, V_D and V_R (respectively, p^2, c^2, h^2, and d^2).

Line	Permanent environment				Common environment				Additive genetic				Dominance				Residual	Ratios			
	V_{PE} ± se	χ^2	P		V_C ± se	χ^2	P		V_A ± se	χ^2	P		V_D ± se	χ^2	P		V_R ± se	p^2	c^2	h^2	d^2
C1	0.380 ± 0.079	22.52	0.0000		0.089 ± 0.069	2.40	0.0608		0.320 ± 0.135	9.15	0.0012		0.000 ± 0.000	0.00	0.5000		0.225 ± 0.015	0.375	0.088	0.316	0.000
C2	0.513 ± 0.078	33.50	0.0000		0.000 ± 0.000	0.00	0.5000		0.125 ± 0.097	2.36	0.0621		0.000 ± 0.000	0.00	0.5000		0.237 ± 0.016	0.587	0.000	0.143	0.000
C4	0.537 ± 0.078	33.03	0.0000		0.000 ± 0.000	0.00	0.5000		0.122 ± 0.096	2.10	0.0734		0.000 ± 0.000	0.00	0.5000		0.209 ± 0.014	0.619	0.000	0.140	0.000
C5	0.312 ± 0.089	9.90	0.0008		0.000 ± 0.000	0.00	0.5000		0.540 ± 0.166	31.59	0.0000		0.000 ± 0.000	0.00	0.5000		0.148 ± 0.010	0.312	0.000	0.540	0.000
C pooled	*0.47 ± 0.035*	*155.79*	*0.0000*		*0.012 ± 0.020*	*0.36*	*0.2749*		*0.250 ± 0.050*	*50.46*	*0.0000*						*0.207 ± 0.007*	*0.503*	*0.012*	*0.265*	
HR3	0.487 ± 0.050	4.65	0.0155		0.076 ± 0.076	0.27	0.3010		0.081 ± 0.087	0.83	0.1809		0.000 ± 0.000	0.00	0.4999		0.218 ± 0.009	0.565	0.088	0.094	0.000
HR6	0.481 ± 0.052	33.47	0.0000		0.048 ± 0.035	3.09	0.0394		0.193 ± 0.092	9.55	0.0010		0.000 ± 0.000	0.00	0.5000		0.160 ± 0.007	0.545	0.055	0.219	0.000
HR7	0.383 ± 0.431	3.47	0.0312		0.043 ± 0.152	0.07	0.3926		0.094 ± 0.099	0.92	0.1684		0.228 ± 0.631	0.13	0.3592		0.144 ± 0.006	0.430	0.048	0.105	0.269
HR8	0.434 ± 0.089	12.50	0.0002		0.000 ± 0.000	0.00	0.5000		0.063 ± 0.081	0.55	0.2285		0.180 ± 0.142	2.09	0.0742		0.222 ± 0.009	0.483	0.000	0.070	0.201
HR poole	*0.51 ± 0.021*	*259.33*	*0.0000*		*0.055 ± 0.013*	*30.38*	*0.0000*		*0.124 ± 0.032*	*29.55*	*0.0000*						*0.186 ± 0.004*	*0.584*	*0.063*	*0.141*	

After the selection limit (generations 20-31, except for line 8 which starts at generation 25)

Line	Permanent environment				Common environment				Additive genetic				Dominance				Residual	Ratios			
	V_{PE} ± se	χ^2	P		V_C ± se	χ^2	P		V_A ± se	χ^2	P		V_D ± se	χ^2	P		V_R ± se	p^2	c^2	h^2	d^2
C1	0.370 ± 0.105	8.16	0.0021		0.000 ± 0.000	0.00	0.5000		0.059 ± 0.119	0.24	0.3107		0.251 ± 0.268	1.04	0.1541		0.266 ± 0.017	0.391	0.000	0.063	0.265
C2	0.307 ± 0.082	13.16	0.0001		0.083 ± 0.064	2.32	0.0637		0.280 ± 0.125	6.67	0.0049		0.000 ± 0.000	0.00	0.5000		0.216 ± 0.014	0.346	0.094	0.316	0.000
C4	0.403 ± 0.087	19.68	0.0000		0.000 ± 0.000	0.00	0.5000		0.365 ± 0.130	23.82	0.0000		0.000 ± 0.000	0.00	0.5000		0.142 ± 0.009	0.443	0.000	0.401	0.000
C5	0.405 ± 0.076	21.83	0.0000		0.000 ± 0.000	0.00	0.5000		0.211 ± 0.099	9.18	0.0012		0.000 ± 0.000	0.00	0.5000		0.186 ± 0.012	0.505	0.000	0.264	0.000
C pooled	*0.42 ± 0.033*	*143.76*	*0.0000*		*0.037 ± 0.021*	*3.59*	*0.0290*		*0.231 ± 0.043*	*50.08*	*0.0000*						*0.204 ± 0.007*	*0.472*	*0.041*	*0.259*	
HR3	0.342 ± 0.063	10.63	0.0006		0.020 ± 0.032	0.42	0.2595		0.320 ± 0.105	7.65	0.0028		0.000 ± 0.000	0.00	0.5000		0.164 ± 0.007	0.404	0.024	0.379	0.000
HR6	0.475 ± 0.041	52.86	0.0000		0.025 ± 0.028	1.02	0.1562		0.082 ± 0.053	3.41	0.0323		0.000 ± 0.000	0.00	0.5000		0.148 ± 0.006	0.651	0.034	0.113	0.000
HR7	0.453 ± 0.045	34.46	0.0000		0.106 ± 0.046	7.40	0.0033		0.012 ± 0.065	0.02	0.4449		0.000 ± 0.000	0.00	0.5000		0.127 ± 0.005	0.648	0.152	0.017	0.000
HR8	0.430 ± 0.075	13.47	0.0001		0.093 ± 0.060	3.90	0.0242		0.104 ± 0.104	0.50	0.5000		0.000 ± 0.000	0.00	0.5000		0.286 ± 0.015	0.470	0.102	0.114	0.000
HR poole	*0.44 ± 0.022*	*165.00*	*0.0000*		*0.053 ± 0.013*	*27.32*	*0.0000*		*0.120 ± 0.030*	*27.32*	*0.0000*						*0.171 ± 0.004*	*0.563*	*0.067*	*0.152*	

APPENDIX S7

Table S3. Results from ANCOVAs with cumulative selection response as dependent variable and cumulative selection differential and line, in addition to their interaction, as independent variables, testing for heterogeneity in realized heritability (h^2_w) among replicate lines in females, males, and the average of the two (mean, as presented in the main text). Also shown are results of ANOVAs with cumulative selection response as the dependent variable and line as the independent variable, testing whether the height of the selection plateau differed among lines by restricting the data to generations after the limit. In both ANCOVAs and ANOVAs the *F*-statistic is presented for the main effect (a and c) along with the *t*-statistic for the individual line contrasts (b and d). Parameter estimates of the segmented regression and their standard errors are given in Table 1 of the main text. Note: resetting the cumulative selection differential to zero for the generation at which the limit was reached yielded identical results.

	ANCOVA: interaction of cumulative selection X line (realized heritability)								
	Females			Males			Mean		
a. Main effect	*F*	df	*P*	*F*	df	*P*	*F*	df	*P*
up to gen 10	**8.288**	**3,36**	**<0.001**	2.159	3,36	0.110	**5.367**	**3,36**	**0.004**
up to limit	**18.164**	**3,80**	**<0.001**	**3.005**	**3,75**	**0.036**	**10.551**	**3,77**	**0.000**
after limit	0.644	3,33	0.592	0.307	3,37	0.820	0.254	3,35	0.858
b. Line contrasts	*t*	df	*P*	*t*	df	*P*	*t*	df	*P*
HR lines 3 vs. 6	**3.157**	**80**	**0.002**	1.581	75	0.118	**2.463**	**77**	**0.016**
HR lines 3 vs. 7	0.943	80	0.349	0.629	75	0.531	0.306	77	0.761
HR lines 3 vs. 8	**3.025**	**80**	**0.003**	1.009	75	0.316	**2.351**	**77**	**0.021**
HR lines 6 vs. 7	**2.504**	**80**	**0.014**	**2.327**	**75**	**0.023**	**2.418**	**77**	**0.018**
HR lines 6 vs. 8	**6.607**	**80**	**<0.001**	**2.916**	**75**	**0.005**	**5.344**	**77**	**0.000**
HR lines 7 vs. 8	**4.668**	**80**	**<0.001**	0.338	75	0.736	**3.151**	**77**	**0.002**
	ANOVA: main line effect (height of the plateau)								
c. Main effect	*F*	df	*P*	*F*	df	*P*	*F*	df	*P*
after limit	**7.109**	**3,37**	**<0.001**	**2.558**	**3,41**	**0.014**	**5.901**	**3,39**	**0.002**
d. Line contrasts	*t*	df	*P*	*t*	df	*P*	*t*	df	*P*
HR lines 3 vs. 6	**4.345**	**37**	**<0.001**	**3.120**	**41**	**0.003**	**4.110**	**39**	**<0.001**
HR lines 3 vs. 7	**2.042**	**37**	**0.048**	0.153	41	0.879	1.037	39	0.306
HR lines 3 vs. 8	0.206	37	0.838	**2.558**	**41**	**0.014**	1.391	39	0.172
HR lines 6 vs. 7	2.010	37	0.052	**2.782**	**41**	**0.008**	**2.743**	**39**	**0.009**
HR lines 6 vs. 8	**3.033**	**37**	**0.004**	0.037	41	0.971	1.833	39	0.075
HR lines 7 vs. 8	1.370	37	0.179	**2.321**	**41**	**0.025**	0.483	39	0.632

APPENDIX S8
LITTER SIZE AT WEANING

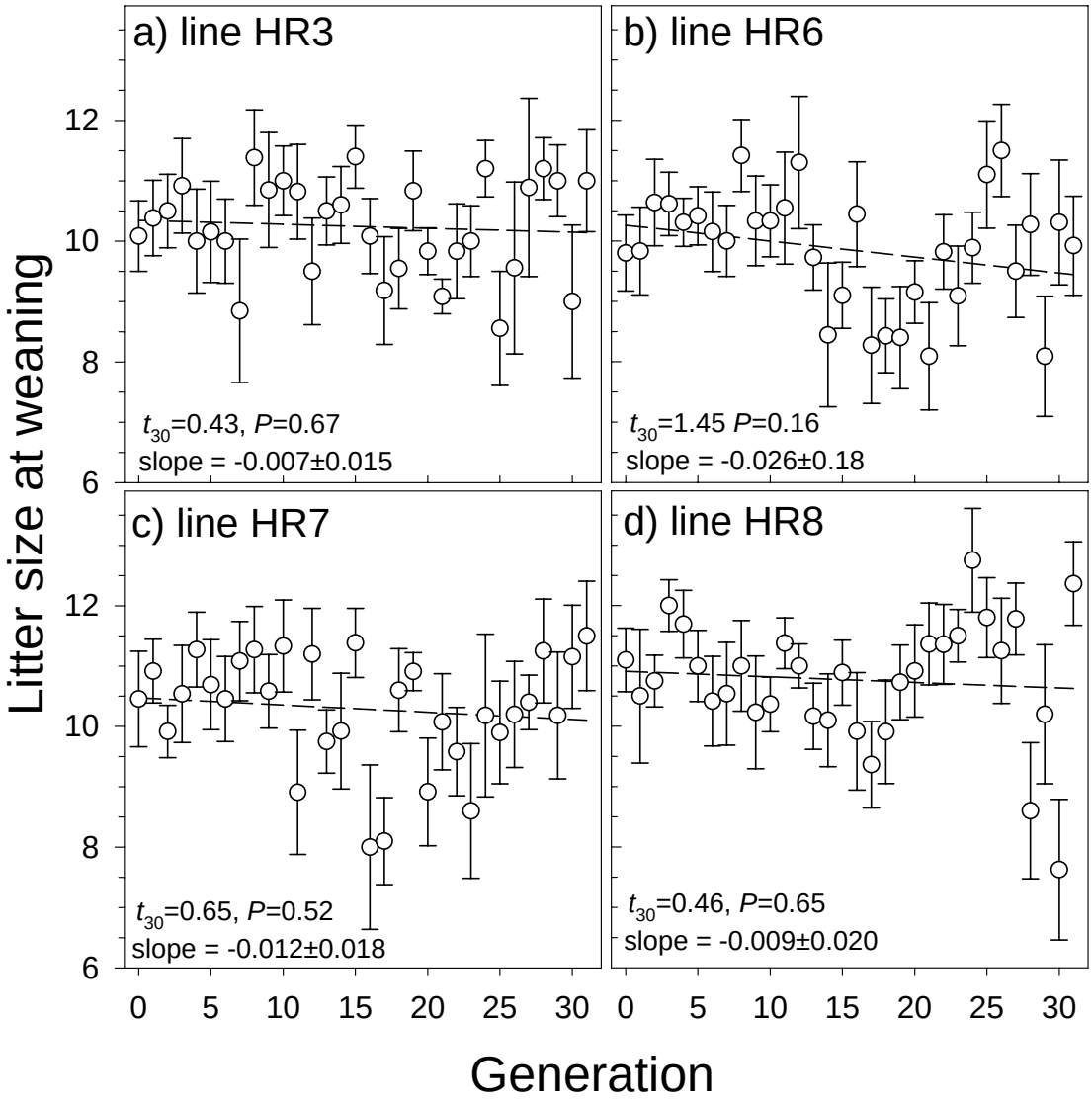

Figure S5. Litter size at weaning (±se) in four lines (a-d) of mice artificially selected for high voluntary wheel running over 31 generations. Note: litter size was impossible to estimate in control lines since only four pups were weaned per family.

APPENDIX S9

CROSS-SEX GENETIC CORRELATIONS

Table S4. Sex-specific additive genetic variances (females: $V_{A\text{-}F}$; males: $V_{A\text{-}M}$) and cross-sex genetic correlations (r_{MF}±se) in quantitative-genetic analyses of voluntary wheel running (treating revolutions run on days 5 & 6 as repeated measures of the same trait) in four non-selected control (C) lines of house mice and four lines artificially selected for high levels of voluntary wheel running (High Runner or HR lines) and their founders (base population). Models were run for (a) the base population (generations -2 to 0), (b) before and (c) after the selection limit with heterogeneous variance between selection groups (C vs. HR lines), and (d) before and (e) after the limit with heterogeneous variance among lines. Log-likelihood ratio tests were used to assess whether r_{MF} was significantly smaller than unity (r_{MF}<0.999) and different from zero (r_{MF}≠0). Note: as the former test evaluated a component against a boundary of its parameter space (i.e., r_{MF}<0.999), we used an equally weighted mixture of χ^2-distributions with one and zero *d.f.* ($\chi^2_{0:1}$) (Self and Liang 1987).

												$r_{MF} < 0.999$		$r_{MF} \neq 0$	
Gens	Group	$V_{A\text{-}F}$	±	se	$V_{A\text{-}M}$	±	se	r_{MF}	±	se		$\chi^2_{0:1}$	*P*	χ^2_1	*P*
a. Base population															
-2-0		0.172	±	0.107	0.250	±	0.085	0.949	±	0.317		0.023	0.4401	10.590	0.0011
b. Before the limit, heterogeneous variance for selection group															
0-10	C	0.277	±	0.069	0.199	±	0.055	0.952	±	0.080		0.490	0.2421	35.648	0.0000
0-10	HR	0.170	±	0.037	0.156	±	0.033	0.954	±	0.074		0.462	0.2484	40.713	0.0000
c. After the limit, heterogeneous variance for selection group															
21-31	C	0.212	±	0.056	0.218	±	0.053	0.916	±	0.101		0.993	0.1595	29.457	0.0000
21-31[a]	HR	0.195	±	0.036	0.129	±	0.028	0.840	±	0.080		7.459	0.0032	37.549	0.0000
d. Before the limit, heterogeneous variance for line															
0-10	C1	0.317	±	0.138	0.145	±	0.087	0.999	±	NA		0.002	0.4834	11.253	0.0008
0-10	C2	0.269	±	0.132	0.175	±	0.113	0.516	±	0.321		6.461	0.0055	2.106	0.1468
0-10	C4	0.090	±	0.102	0.165	±	0.126	0.999	±	NA		0.000	0.5000	3.538	0.0600
0-10	C5	0.438	±	0.168	0.298	±	0.117	0.999	±	NA		0.000	0.5000	23.764	0.0000
0-10	HR3	0.236	±	0.078	0.221	±	0.075	0.999	±	NA		0.000	0.4984	16.881	0.0000
0-10	HR6	0.163	±	0.069	0.147	±	0.065	0.999	±	NA		0.000	0.4993	12.281	0.0005
0-10	HR7	0.184	±	0.067	0.159	±	0.063	0.999	±	NA		0.000	0.4988	13.451	0.0002
0-10	HR8	0.112	±	0.061	0.162	±	0.062	0.688	±	0.274		1.261	0.1308	4.827	0.0280
e. After the limit, heterogeneous variance for line															
21-31	C1	0.183	±	0.103	0.005	±	0.017	0.999	±	NA		0.000	0.5000	0.442	0.5063
21-31	C2	0.215	±	0.111	0.219	±	0.121	0.966	±	0.220		0.446	0.2521	5.961	0.0146
21-31	C4	0.238	±	0.113	0.320	±	0.105	0.999	±	NA		0.433	0.2553	19.792	0.0000
21-31	C5	0.196	±	0.097	0.419	±	0.125	0.999	±	NA		0.000	0.4987	10.198	0.0014
21-31	HR3	0.223	±	0.066	0.178	±	0.057	0.999	±	NA		0.000	0.4986	21.570	0.0000
21-31	HR6	0.182	±	0.057	0.087	±	0.041	0.999	±	NA		0.433	0.2554	13.077	0.0003
21-31	HR7	0.243	±	0.061	0.117	±	0.049	0.999	±	NA		0.000	0.4987	16.449	0.0000
25-31	HR8	0.060	±	0.048	0.221	±	0.078	0.999	±	NA		0.433	0.2552	6.507	0.0107

APPENDIX S10

PROPORTION OF VARIANCE

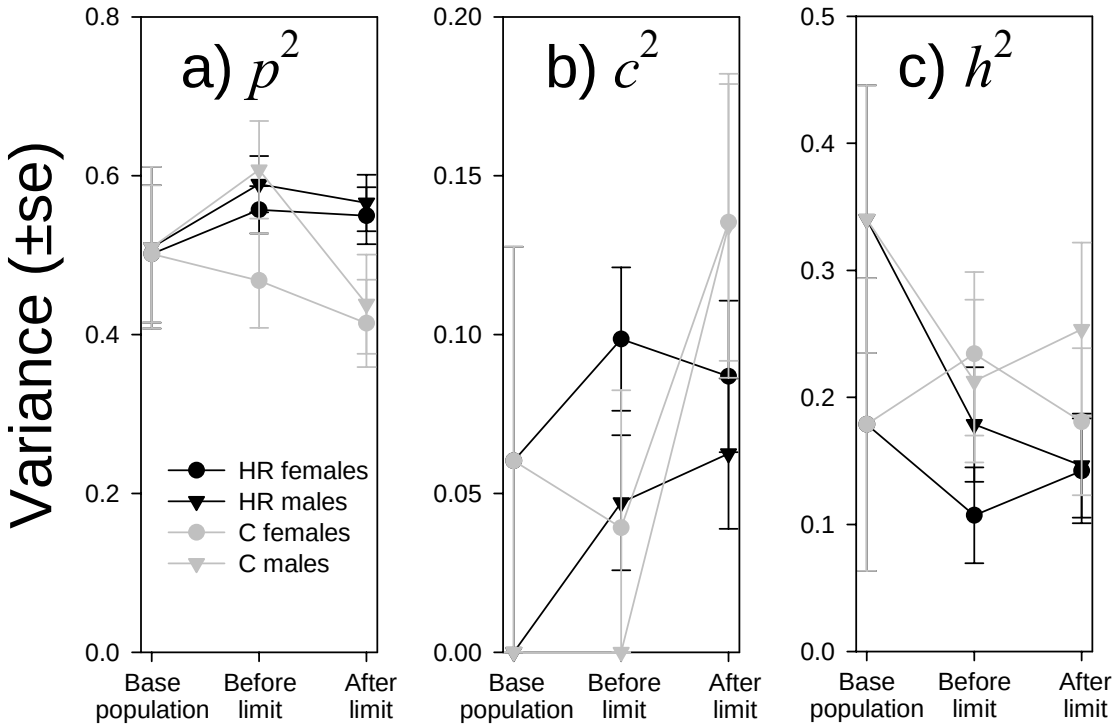

Figure S6. Proportion of variance (ratio±se) in voluntary wheel running explained by (a) permanent environment (p^2), (b) common environment (c^2), and (c) additive genetic (h^2) effects in models allowing heterogeneous variance between sexes (females: circles; males: triangles) and selection history (control lines C: grey; selected lines HR: black) from the base population up to generation 10 (before the selection limit) and from generation 20 to 31 (after the selection limit, except for the line that reached the limit at generation 25, see Table 1 and Fig. 2d).

APPENDIX S11

ANIMAL MODELS WITH HETEROGENEOUS VARIANCE BETWEEN SEXES AND AMONG LINES

Table S5. Sex- and line-specific components of variance (±se; V_{PE}, permanent environmental; V_C, common environmental; V_A, additive genetic; V_R, environmental variance) in univariate quantitative genetic analyses of voluntary wheel running (treating revolutions run on days 5 & 6 as repeated measures of the same trait) in four non-selected control (C) lines of house mice and four lines artificially selected for high levels of voluntary wheel running (High Runner or HR lines). Models were run separately on different intervals of generations (gen), from either (a) 0 to 10 (before the limit) or (b) generations 21 to 31 (after the limit, except for line HR8 for which the interval was 25 to 31[a]). See Table 4 for more information.

			Permanent environment			Common environment			Additive genetic			Residual	Ratios		
Gens	Sex	Line	V_{PE} ± se	χ^2	P	V_C ± se	χ^2	P	V_A ± se	χ^2	P	V_R ± se	p^2 ± se	c^2 ± se	h^2 ± se
a. Heterogeneous variance between sexes and lines: before the limit															
0-10	f	C1	0.361 ± 0.111	11.56	0.0003	0.120 ± 0.093	1.86	0.0860	0.325 ± 0.155	8.93	0.0014	0.289 ± 0.028	0.33 ± 0.11	0.11 ± 0.08	0.30 ± 0.13
0-10	f	C2	0.444 ± 0.107	16.42	0.0000	0.000 ± NA	0.00	0.5000	0.239 ± 0.127	6.31	0.0060	0.279 ± 0.027	0.46 ± 0.11	0.00 ± 0.00	0.25 ± 0.12
0-10	f	C4	0.703 ± 0.086	32.50	0.0000	0.000 ± NA	0.00	0.5000	0.000 ± NA	0.00	0.4996	0.295 ± 0.028	0.70 ± 0.03	0.00 ± 0.00	0.00 ± 0.00
0-10	f	C5	0.454 ± 0.128	10.43	0.0006	0.065 ± 0.090	0.60	0.2189	0.355 ± 0.174	8.69	0.0016	0.216 ± 0.021	0.42 ± 0.13	0.06 ± 0.08	0.33 ± 0.14
0-10	f	HR3	0.493 ± 0.060	39.38	0.0000	0.131 ± 0.050	12.02	0.0003	0.131 ± 0.084	4.69	0.0152	0.295 ± 0.017	0.47 ± 0.06	0.12 ± 0.05	0.13 ± 0.08
0-10	f	HR6	0.598 ± 0.059	58.48	0.0000	0.129 ± 0.051	10.29	0.0007	0.081 ± 0.066	3.14	0.0383	0.197 ± 0.012	0.60 ± 0.06	0.13 ± 0.05	0.08 ± 0.06
0-10	f	HR7	0.638 ± 0.063	69.91	0.0000	0.105 ± 0.049	7.49	0.0031	0.136 ± 0.082	5.25	0.0110	0.175 ± 0.010	0.61 ± 0.06	0.10 ± 0.05	0.13 ± 0.07
0-10	f	HR8	0.512 ± 0.055	65.58	0.0000	0.025 ± 0.034	0.59	0.2218	0.120 ± 0.067	5.74	0.0083	0.286 ± 0.017	0.54 ± 0.06	0.03 ± 0.04	0.13 ± 0.07
0-10	m	C1	0.457 ± 0.096	17.40	0.0000	0.000 ± NA	0.00	0.5000	0.152 ± 0.110	2.52	0.0562	0.158 ± 0.015	0.60 ± 0.13	0.00 ± 0.00	0.20 ± 0.13
0-10	m	C2	0.564 ± 0.104	23.53	0.0000	0.000 ± NA	0.00	0.5000	0.146 ± 0.108	3.49	0.0309	0.201 ± 0.019	0.62 ± 0.11	0.00 ± 0.00	0.16 ± 0.11
0-10	m	C4	0.487 ± 0.091	17.57	0.0000	0.133 ± 0.082	2.79	0.0474	0.032 ± 0.095	0.09	0.3815	0.136 ± 0.013	0.62 ± 0.11	0.17 ± 0.10	0.04 ± 0.12
0-10	m	C5	0.438 ± 0.094	19.85	0.0000	0.024 ± 0.070	0.12	0.3644	0.211 ± 0.115	7.26	0.0035	0.089 ± 0.009	0.58 ± 0.13	0.03 ± 0.09	0.28 ± 0.14
0-10	m	HR3	0.385 ± 0.050	32.43	0.0000	0.023 ± 0.030	0.64	0.2110	0.156 ± 0.079	4.83	0.0140	0.142 ± 0.008	0.55 ± 0.08	0.03 ± 0.04	0.22 ± 0.10
0-10	m	HR6	0.461 ± 0.047	50.33	0.0000	0.037 ± 0.031	1.54	0.1071	0.084 ± 0.061	2.65	0.0519	0.127 ± 0.007	0.65 ± 0.07	0.05 ± 0.04	0.12 ± 0.08
0-10	m	HR7	0.392 ± 0.041	51.76	0.0000	0.068 ± 0.033	6.73	0.0047	0.092 ± 0.057	5.22	0.0111	0.111 ± 0.007	0.59 ± 0.07	0.10 ± 0.05	0.14 ± 0.08
0-10	m	HR8	0.500 ± 0.052	67.86	0.0000	0.016 ± 0.032	0.30	0.2923	0.166 ± 0.067	15.53	0.0000	0.161 ± 0.009	0.59 ± 0.06	0.02 ± 0.04	0.20 ± 0.07
b. Heterogeneous variance between sexes and lines: after the limit															
21-31	f	C1	0.444 ± 0.114	14.34	0.0001	0.037 ± 0.081	0.21	0.3241	0.181 ± 0.103	6.25	0.0062	0.340 ± 0.032	0.44 ± 0.11	0.04 ± 0.08	0.18 ± 0.10
21-31	f	C2	0.147 ± 0.093	2.17	0.0702	0.153 ± 0.073	5.82	0.0079	0.224 ± 0.118	4.09	0.0216	0.290 ± 0.028	0.18 ± 0.12	0.19 ± 0.08	0.28 ± 0.14
21-31	f	C4	0.612 ± 0.124	25.12	0.0000	0.111 ± 0.101	1.35	0.1225	0.169 ± 0.119	3.58	0.0292	0.215 ± 0.020	0.55 ± 0.11	0.10 ± 0.09	0.15 ± 0.11
21-31	f	C5	0.403 ± 0.122	8.94	0.0014	0.245 ± 0.104	7.04	0.0040	0.175 ± 0.135	2.13	0.0724	0.228 ± 0.022	0.38 ± 0.12	0.23 ± 0.09	0.17 ± 0.13
21-31	f	HR3	0.508 ± 0.073	30.62	0.0000	0.040 ± 0.040	1.15	0.1419	0.206 ± 0.091	9.22	0.0012	0.193 ± 0.012	0.54 ± 0.08	0.04 ± 0.04	0.22 ± 0.09
21-31	f	HR6	0.487 ± 0.061	41.23	0.0000	0.097 ± 0.044	8.09	0.0022	0.126 ± 0.066	8.18	0.0021	0.159 ± 0.010	0.56 ± 0.07	0.11 ± 0.05	0.14 ± 0.08
21-31	f	HR7	0.495 ± 0.057	43.33	0.0000	0.060 ± 0.037	3.64	0.0282	0.119 ± 0.060	11.30	0.0004	0.145 ± 0.009	0.60 ± 0.06	0.07 ± 0.04	0.14 ± 0.07
25-31	f	HR8	0.453 ± 0.054	27.74	0.0000	0.151 ± 0.054	9.87	0.0008	0.000 ± NA	0.00	0.4991	0.338 ± 0.025	0.48 ± 0.05	0.16 ± 0.05	0.00 ± 0.00
21-31	m	C1	0.403 ± 0.071	20.80	0.0000	0.087 ± 0.063	2.08	0.0744	0.000 ± NA	0.00	0.5000	0.199 ± 0.019	0.58 ± 0.09	0.13 ± 0.09	0.00 ± 0.00
21-31	m	C2	0.477 ± 0.121	15.65	0.0000	0.211 ± 0.106	4.81	0.0142	0.200 ± 0.132	3.83	0.0252	0.135 ± 0.013	0.47 ± 0.12	0.21 ± 0.10	0.20 ± 0.13
21-31	m	C4	0.315 ± 0.088	13.98	0.0001	0.035 ± 0.068	0.27	0.3023	0.294 ± 0.111	14.72	0.0001	0.075 ± 0.007	0.44 ± 0.13	0.05 ± 0.09	0.41 ± 0.14
21-31	m	C5	0.169 ± 0.097	2.66	0.0515	0.031 ± 0.060	0.29	0.2944	0.357 ± 0.130	8.58	0.0017	0.137 ± 0.013	0.24 ± 0.15	0.05 ± 0.09	0.51 ± 0.16
21-31	m	HR3	0.339 ± 0.046	36.80	0.0000	0.021 ± 0.026	0.92	0.1682	0.119 ± 0.054	9.49	0.0010	0.137 ± 0.008	0.55 ± 0.07	0.03 ± 0.04	0.19 ± 0.08
21-31	m	HR6	0.353 ± 0.041	53.32	0.0000	0.066 ± 0.034	5.10	0.0120	0.065 ± 0.041	4.92	0.0132	0.126 ± 0.008	0.58 ± 0.06	0.11 ± 0.05	0.11 ± 0.07
21-31	m	HR7	0.365 ± 0.040	41.84	0.0000	0.070 ± 0.030	8.30	0.0020	0.007 ± 0.036	0.03	0.4323	0.097 ± 0.006	0.68 ± 0.06	0.13 ± 0.05	0.01 ± 0.07
25-31	m	HR8	0.428 ± 0.072	25.58	0.0000	0.019 ± 0.039	0.29	0.2963	0.147 ± 0.072	9.78	0.0009	0.236 ± 0.017	0.52 ± 0.08	0.02 ± 0.05	0.18 ± 0.09

APPENDIX S12

RATE VS. DURATION OF RESPONSE TO SELECTION

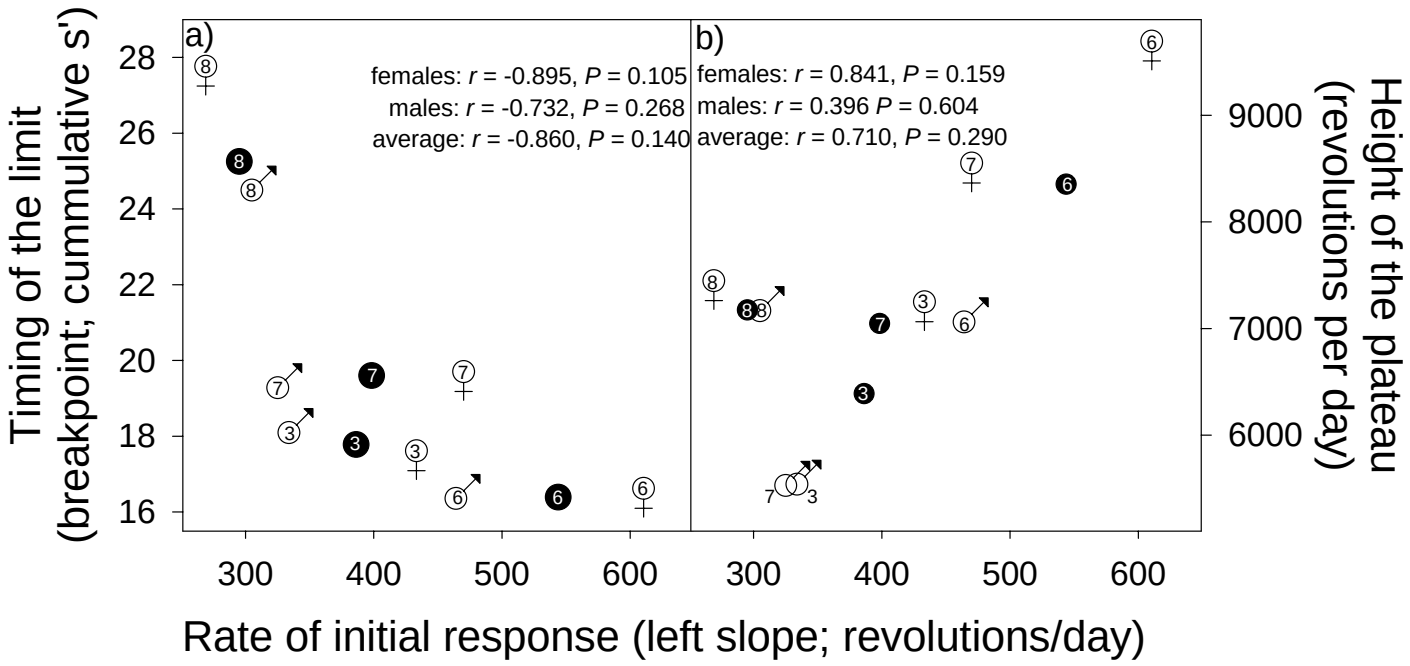

Figure S7. Rate of initial response to selection (left slope of the segmented regression, increase in revolutions per day per generation) as a function of (a) timing at which the selection limit was reached (breakpoint of the segmented regression, in cumulative standardized selection differential s') and (b) height of the selection plateau (intercept of the null right slope in the segmented regression), for males, females, and their average (solid dots). Line numbers are indicated within symbols.

LITERATURE CITED

Åkesson, M., S. Bensch, D. Hasselquist, M. Tarka, and B. Hansson. 2008. Estimating heritabilities and genetic correlations: comparing the "animal model" with parent-offspring regression using data from a natural population. Public Library of Science - Biology 3:e1739.

DeFries, J. C., J. R. Wilson, and G. E. McClearn. 1970. Open-field behavior in mice: Selection response and situational generality. Behav. Genet. 1:195-211.

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

Falconer, D. S. 1973. Replicated selection for body weight in mice. Genet. Res. Camb. 22:291-321.

Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to Quantitative Genetics, 4th edn. Longman: Harlow.

Garland, T., Jr. and S. A. Kelly. 2006. Phenotypic plasticity and experimental evolution. The Journal of Experimental Biology 209:2344-2361.

Gibson, P. T. 1996. Correcting for inbreeding in parent-offspring regression estimates of heritability with non-additive and genotype X environment effects present. Crop Sci 36:594-600.

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

Kinghorn, B. P. and A. J. Kinghorn. 2002. Pedigree Viewer 5.0 - a program for drawing and manipulating pedigree diagrams,, University of New England Armidale, Australia.

Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the 'animal model'. Phil. Trans. R. Soc. B. 359:873-890.

Lynch, C. B. 1980. Response to divergent selection for nesting behavior in *Mus musculus*. Genetics 96:757-765.

Lynch, M. and J. B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, Massachusetts.

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

Nyquist, W. E. 1991. Estimation of heritability and prediction of selection response in plant populations. Crit Rev Plant Sci 10:235-322.

Pollak, E. J. and R. L. Quaas. 1983. Definition of group effects in sire evaluation models. J Dairy Sci 66:1503-1509.

Self, S. C. and K. Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. J. Am. Stat. Soc. 82:605-610.

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

Wolak, M. E. 2012. nadiv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. Methods Ecol. Evol. 3:792-796.