EVOLUTION OF A SMALL-MUSCLE POLYMORPHISM IN LINES OF HOUSE MICE SELECTED FOR HIGH ACTIVITY LEVELS

Theodore Garland, Jr.,[1] Martin T. Morgan,[2] John G. Swallow,[3] Justin S. Rhodes,[4] Isabelle Girard,[4] Jason G. Belter,[5] and Patrick A. Carter[2]

[1]*Department of Biology, University of California, Riverside, California 92521*
E-mail: tgarland@citrus.ucr.edu
[2]*School of Biological Sciences, Washington State University, Pullman, Washington 99164*
E-mail: mmorgan@wsu.edu; pacarter@wsu.edu
[3]*Department of Biology, University of South Dakota, Vermillion, South Dakota 57069*
E-mail: jswallow@usd.edu
[4]*Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706*
E-mail: justinrhodes@students.wisc.edu; igirard@facstaff.wisc.edu
[5]*Department of Biochemistry, University of Minnesota, 1479 Gortner Avenue, Saint Paul, Minnesota 55108*
E-mail: belt0042@tc.umn.edu

Abstract.—To study the correlated evolution of locomotor behavior and exercise physiology, we conducted an artificial selection experiment. From the outbred Hsd:ICR strain of *Mus domesticus*, we began eight separate lines, each consisting of 10 breeding pairs. In four of the lines, we used within-family selection to increase voluntary wheel running. The remaining four lines were random-bred (within lines) to serve as controls. Various traits have been monitored to test for correlated responses. Here, we report on organ masses, with emphasis on the triceps surae muscle complex, an important extensor of the ankle. Mice from the selected lines exhibit reduced total body mass, increased relative (mass-corrected) kidney mass, and reduced relative triceps surae mass. In addition, a discrete muscle-mass polymorphism was observed: some individuals had triceps surae that were almost 50% lighter than normal for their body mass. This small-muscle phenotype was observed in only three of the eight lines: in one control line, it has fluctuated in frequency between zero and 10%, whereas in two of the selected lines it has increased in frequency to approximately 50% by generation 22. Data from a set of parents and offspring (generations 23 and 24) are consistent with inheritance as a single autosomal recessive allele. Evidence for the adaptive significance of the small-muscle allele was obtained by fitting multiple-generation data to hierarchical models that include effects of genetic drift and/or selection. The small-muscle allele is estimated to have been present at low frequency (approximately 7%) in the base population, and analysis indicates that strong selection favors the allele in the selected but not control lines. We hypothesize that the small muscles possess functional characteristics and/or that the underlying allele causes pleiotropic effects (e.g., reduced total body mass; increased relative heart, liver, and kidney mass) that facilitate high levels of wheel running. Nevertheless, at generation 22, wheel running of affected individuals did not differ significantly from those with normal-sized muscles, and the magnitude of response to selection has been similar in all four selected lines, indicating that multiple genetic ''solutions'' are possible in response to selection for high activity levels.

Key words.—Behavior, correlated response, exercise physiology, locomotion, organ masses, polymorphism, recessive allele.

Natural and sexual selection are thought to act relatively directly on behavior and whole-organism performance traits, as compared with lower-level morphological, physiological, and biochemical phenotypes (Arnold 1983; Irschick and Garland 2001). Because much animal behavior involves locomotion of some sort, the ability to perform various types of locomotion has been of keen interest to ecological and evolutionary physiologists and morphologists (Bennett and Huey 1990; Garland and Losos 1994; Autumn et al. 1999; Boggs and Frappell 2000; Miles et al. 2000; Kramer and McLaughlin 2001; Van Damme et al. 2002). At the same time, movement patterns of animals in nature have received considerable attention from behavioral ecologists (Bunnell and Harestad 1990; Perry 1999; Nunn and Barton 2000; Kelt and Van Vuren 2001; Kramer and McLaughlin 2001; Perry and Garland 2002). Less often studied is the relationship between performance abilities and behavioral propensities, that is, whether and how often animals actually use their maximal abilities (Irschick and Garland 2001). For example, few comparative studies have asked whether locomotor abilities have evolved in concert with ranging patterns, largely because it

is difficult to obtain quantitative information on field movements (Garland 1999).

Selection experiments are a powerful way to elucidate correlated evolution (Rose et al. 1996; Gibbs 1999; Feder et al. 2000; Endler et al. 2001; Garland 2001, 2002) and may be of particular use for traits that are difficult to study in wild organisms. To study the correlated evolution of exercise physiology and locomotor behavior (propensity to be active), we have selected for increased wheel-running behavior in house mice. Beginning with a base population of outbred Hsd:ICR mice (see Carter et al. 1999; Dohm et al. 2001; refs. therein), we produced four replicate lines of mice selected for high voluntary wheel-running activity, while also maintaining four randomly bred lines as controls (Swallow et al. 1998a). After 10 generations of selection, the high-selected lines were running, on average, 75% more (total revolutions per day) than the control lines (Swallow et al. 1998a; Koteja et al. 1999a; Carter et al. 2000). This level of activity exceeds that of wild house mice born and raised under the same conditions (Dohm et al. 1994). In subsequent generations, the differential between selected and control lines eventually in-

creased to approximately 160% (see fig. 1 in Rhodes et al. [2000] for summary data through 17 generations), which seems to represent a selection plateau through 30 generations (Garland 2001, 2002 and unpubl. data). The selection-control differential in wheel running almost spans the range of variation in wheel running observed among 13 species of wild murid rodents (Dewsbury 1980; fig. 4 in Garland 2002).

Various behavioral, physiological, and morphological traits have been shown to differ between the replicate selected and control lines. For example, body mass (see also Swallow et al. 1999) and food consumption at 76 days of age showed a correlated response to selection by generation 10, with mice from selected lines being smaller and eating more food (on a mass-adjusted basis) than those from control lines (Koteja et al. 1999a). Mice from selected lines also build smaller thermoregulatory nests (at generation 10, Carter et al. 2000) and have higher body temperatures at night when active on wheels (generation 17, Rhodes et al. 2000). Mice from selected lines have more symmetrical hindlimb bone lengths (Garland et al. 2000) and, under some housing conditions, mature males from the selected lines have shown a maximal aerobic capacity (oxygen consumption elicited during forced treadmill exercise) approximately 6% higher than in control lines (Swallow et al. 1998b). Males from selected lines also exhibit elevated insulin-stimulated glucose uptake in some hindlimb muscles (Dumke et al. 2001), which should facilitate the relatively high-speed, but still probably aerobically supported, wheel running of selected-line mice (Girard et al. 2001). When housed with access to functional running wheels for several weeks, mice from selected lines often show greater training responses than do controls (e.g., greater increases in muscle citrate synthase activity), which constitutes a genotype-by-environment interaction, presumably because of their higher activity levels (Houle-Leroy et al. 2000). With respect to the motivational basis of elevated wheel running in the selected lines, pharmacological studies suggest differences from control lines in the dopaminergic neuromodulatory system (Rhodes et al. 2001).

Here, we investigate the genetic basis and selective significance of a discrete polymorphism in hindlimb muscle mass that has been observed mainly in the selected lines. Routine dissections of the triceps surae (a complex of four muscles in the calf), beginning in the first generation after selection was imposed, revealed occasional individuals with muscles that were reduced by almost half in mass, as compared with ''normal'' mice of similar body mass. Such dramatic variation in muscle mass could have important effects on locomotor abilities, thus we wished to determine if the frequency of small-muscles individuals was evolving as a correlated response to selection for high wheel running.

METHODS

Selection Experiment and Muscle Dissection

Mice were from a replicated artificial selection experiment for increased wheel running behavior, as described in Swallow et al. (1998a). Lines with assigned lab numbers of 1, 2, 4, and 5 are bred randomly as controls, whereas lines 3, 6, 7, and 8 are selected for high voluntary wheel running. In each generation, within-family selection of breeders was based on amount of wheel running (total revolutions) on days 5 and 6 of a six-day test on wheels of 1.12 m circumference. Within the control lines, breeders were chosen randomly. In all lines, sibling-mating is disallowed.

In generations 1 through 7, 10, 14, 18, and 22, one male and one female were randomly chosen for dissection from each of the 10 families in each of the eight lines (total $n =$ 160/generation). As some of these individuals were also required as breeders to produce the next generation, dissections did not occur until their litters had been weaned. Triceps surae muscles (which include the lateral and medial heads of the gastrocnemius, the soleus, and the plantaris) were dissected and weighed as described in Carter et al. (1999). To reduce processing time, only the right triceps surae was dissected in generations 1–7. Inspection of data for these early generations made it clear that something unusual was occurring, so we then began dissecting both right and left muscles, and mean muscle mass was analyzed for subsequent generations. The small-muscle phenotype was identified by inspection of log-log plots of muscle mass versus body mass. In all generations, individuals with the small-muscle phenotype could be identified visually, irrespective of their sex, linetype (selected vs. control), or body mass, based on having a much-reduced muscle mass for their body mass (e.g., see Fig. 2).

For mice from the generation 22 sample, we also dissected and weighed the liver, spleen, kidneys, and ventricle (blotted free of blood) (age at dissection averaged 129.5 days, range = 113–140). SAS PROC MIXED (SAS Institute 1996) was used to implement nested analysis of covariance models to compare $\log_{10}$ organ masses and square-root transformed wheel running of the generation-22 mice. Main effects were linetype (selected vs. control), sex, and small muscle. Random factors were replicate line, nested within linetype, and family, nested within line. Covariates included (depending on the trait) $\log_{10}$ body mass, age, time of day at sacrifice, and (z-transformed time of day)^2 (the last term allows for a nonlinear relationship with time of day). We did not test for interactions because (1) they were not of primary interest, and (2) the representation of individuals with small muscles was highly unequal across lines. Statistical significance was judged at $P < 0.05$.

Genetic Basis of the Small-Muscle Phenotype

For a sample of nine available families from each of lines 3 and 6 (the two selected lines that showed the phenotype), we weighed triceps surae muscles of parents (generation 23) and their offspring (generation 24). Parents with the small-muscle phenotype were assumed to be homozygous recessive, parents with the normal muscle phenotype that produced both normal- and small-muscle offspring were assumed to be heterozygous, and parents that produced only normal-muscle offspring were assumed to be homozygous for the normal allele (for an analogous analysis, see Carter and Watt 1988). Observed muscle phenotypes in the offspring were compared to the phenotype frequencies expected to be produced from these putative genotype crosses by a G-test (critical value at $\alpha = 0.05$, 1 df $= 3.84$).

Selective Significance of the Small-Muscle Phenotype

The strength of selection on the putative allele can be estimated from fitting multiple-generation data to hierarchical models that include effects of genetic drift and/or selection. Under genetic drift alone, the expected frequency of a recessive mutant phenotype is $p^2 + F_t p(1 - p)$ (Crow and Kimura 1970). The inbreeding coefficient in generation t, F_t, is determined by population size, and given by $(1 - (1 - 1/2N_e))^t$. The effective population size, N_e, for each line is approximately 35 (Swallow et al. 1998a); t is the number of generations during which drift operates. Under genetic drift alone, the expected frequency of individuals exhibiting the mutant phenotype is thus a function of generation number t and the estimable parameter p.

Selection is a deterministic force changing the frequency of the recessive allele. Suppose that phenotypically normal individuals have fitness 1, while individuals with the mutant phenotype have fitness $1 + s$. Standard population genetics of infinite populations shows that a single generation of selection changes the frequency of the allele from p to $p' = (p(1 - p) + p^2(1 + s))/(1 + p^2s)$. The expected allele frequency after t generations of selection can be found by iterating this equation t times.

An approximate method combining the effects of drift and selection uses equations for selection to determine allele frequency and equations for drift to translate allele frequency into expected genotype frequency. Expected genotype frequency under models of selection and drift are then a function of t, with estimable parameters p and s. The possibility of selection in only one treatment (e.g., amongst mice selected for voluntary wheel running) can be entertained by estimating the selection coefficient s from only those lines in the treatment. Differential selection in the two treatments (control, selected) can be accommodated by estimating two selection coefficients, s_c and s_+ (selection in the control and selected treatment, respectively).

Data for parameter estimation consisted of 11 estimates of phenotype frequency in each of the eight experimental lines (generations 1 through 7, 10, 14, 18, and 22). Each frequency estimate was weighted by the inverse of the standard error of the estimate. The standard error was estimated as the square root of $f(1 - f)/n$, where f is the observed frequency of the phenotype in a sample of size n. This is the standard error expected from a binomial model. Samples with no mutant phenotypes were weighted as though $f = 1/n$. The nonlinear models were fit to observed data by least squares using the Levenberg-Marquardt method (Press et al. 1992) implemented in Igor Pro (WaveMetrics 1996). This method provides parameter estimates and a measure of the goodness of fit, χ^2, of the model to the data.

The models of drift and selection form a nested hierarchy. The simplest model, of genetic drift alone, estimates a single parameter, p. Explicitly evaluating the significance of this model is difficult because the number of degrees of freedom available from the complex experimental design is not apparent. However, models with more parameters can be compared to models with a subset of parameters. Each additional parameter uses a degree of freedom. The difference in χ^2 values, $\Delta\chi^2$, between a model with n parameters and a model
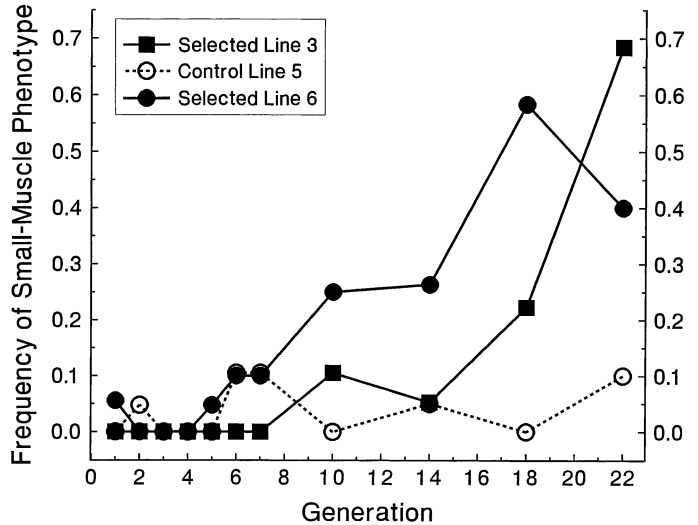

FIG. 1. Frequency of small-muscle phenotype in selected and control treatments over 22 generations (sample size is 10 males and 10 females per line for each generation, with the exception of occasional missing data). Phenotype was never observed in two of the selected and in three of the control lines.

with a subset $k < n$ parameters is asymptotically distributed as χ^2 with $n - k$ degrees of freedom; the large size of the overall experiment lends confidence to the use of asymptotic results. We therefore start with the simplest model, including only drift. By analogy with stepwise regression, we fit two more complex models, with selection acting only in the control treatment or only in the selected treatment. If neither of these models produces a significant increase in goodness of fit, then we conclude that drift alone explains the data and perform no further statistical tests. Otherwise, we provisionally accept the model with highest statistical significance (e.g., including drift and selection in the selected treatment, but without selection in the control treatment). We compare the goodness of fit of this provisional model to the model with drift and selection in both treatments. If the improvement in goodness of fit is not statistically significant, then we accept our provisional model as the best explanation for the data. Otherwise, we accept the full model as the best explanation.

RESULTS

Phenotype Observation and Characterization

The small-muscle phenotype was observed in three of the eight lines, two of which had experienced selection for high voluntary wheel running (lines numbered 6 and 3) and one of which had been bred randomly as a control (line 5). In line 6 (selected), the phenotype was observed in generations 1, 5, 6, 7, 10, 14, 18, and 22, with frequency eventually rising to 0.4 at generation 22 (Fig. 1). In line 3 (selected), the phenotype was not observed until the generation-10 sample, and then rose in frequency to 0.7 by generation 22 (Fig. 1). In the control line (line 5), the phenotype appeared in generations 2, 6, 7, 14, and 22, but the frequency of the phenotype never exceeded 0.1.

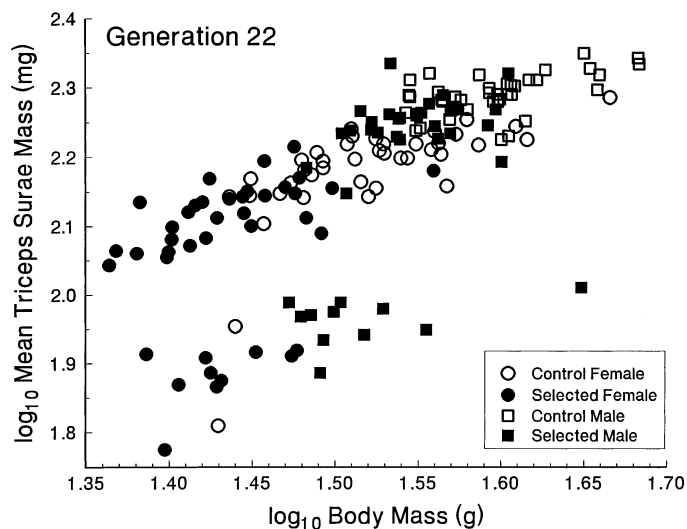Figure 2 shows the relationship of triceps surae muscle

FIG. 2. Mass of triceps surae muscle in relation to body mass for mice from generation 22.

mass to total body mass in mice sampled from generation 22. In this generation, as in all others, muscles from individuals with the small-muscle phenotype could be identified visually in such plots, irrespective of their sex, linetype or body mass. Table 1 presents results of a nested analysis of covariance comparing muscle mass by sex, linetype, and ''small muscle,'' with body mass as a covariate. The triceps surae of individuals with the small-muscle phenotype average about 44% smaller in mass as compared with unaffected individuals (comparison of back-transformed least squares means reported in Table 1). In addition, mice from selected lines have significantly smaller triceps surae ($P = 0.0139$), even after accounting for the effect of the small-muscle phenotype. Finally, males have larger muscles than do females.

Individuals with the small-muscle phenotype have ventricles that average about 8% larger on the arithmetic scale as compared with unaffected individuals (Table 2). Males have significantly larger mass-adjusted ventricles than do females, but selected and control lines do not differ.

Spleen mass ($n = 158$) is larger ($F = 71.49$, df = 1,72,

TABLE 2. Nested analysis of covariance of $\log_{10}$ ventricle mass (mg) in relation to various factors and covariates for mice from generation 22 ($n = 159$).

Source	Degree of freedom		F	P
	Numerator	Denominator		
Selected vs. control[1]	1	6	1.10	0.3346
Sex	1	73	9.98	0.0023
Small muscle	1	73	14.07	0.0004
$\log_{10}$ body mass[2]	1	73	56.73	0.0001
Age	1	73	2.76	0.1012
Time of day	1	73	3.12	0.0813
Squared time of day	1	73	0.14	0.7138

[1] SAS PROC MIXED with line nested within linetype (selected vs. control) and family nested within line as random factors.
[2] Partial regression coefficient = 0.472 ± 0.0626 (± SE).
Least squares (adjusted) means ± SE are: 2.147 ± 0.0075 (control lines) and 2.157 ± 0.0070 (selected lines), 2.141 ± 0.0069 (females) and 2.163 ± 0.0049 (males), 2.135 ± 0.0046 (normal) and 2.169 ± 0.0089 (small-muscled mice).

$P = 0.0001$) in females ($\log_{10}$ L.S. Mean ± SE = 2.089 ± 0.0129) than in males (1.975 ± 0.0113), but does not differ between selected and control lines ($P = 0.8339$) or between individuals with small versus normal-sized triceps surae ($P = 0.1215$). The scaling exponent for spleen mass was 0.906 ± 0.1204.

Liver mass ($n = 157$) does not vary by sex ($P = 0.4244$) or linetype ($P = 0.1206$), but averages significantly larger ($F = 27.47$, df = 1,71, $P = 0.001$) in small-muscled individuals ($\log_{10}$ L.S.Mean ± SE = 3.345 ± 0.0077) as compared with normal animals (3.301 ± 0.0032). The scaling exponent for liver mass was 0.963 ± 0.0636.

Mean $\log_{10}$ kidney mass ($n = 156$) averages larger ($F = 166.60$, df = 1,70, $P = 0.0001$) in males (2.427 ± 0.0066) than in females (2.321 ± 0.0074), larger ($F = 8.07$, df = 1,6, $P = 0.0295$) in selected (2.388 ± 0.0073) than in control lines (2.360 ± 0.0078), and larger ($F = 11.86$, df = 1,70, $P = 0.0010$) in individuals with small (2.392 ± 0.0099) versus normal-sized (2.356 ± 0.0046) triceps surae. The scaling exponent for kidney mass was 0.619 ± 0.0726.

Analysis of $\log_{10}$ body mass (Table 3) indicates that individuals with the small-muscle phenotype average about 13% smaller as compared with unaffected individuals. In addition, after accounting for effects of the small-muscle phenotype, females are smaller than males, and mice from se-

TABLE 1. Nested analysis of covariance of $\log_{10}$ triceps surae muscle mass (mg) in relation to various factors and covariates for mice from generation 22 ($n = 158$). One individual removed as statistical outlier.

Source	Degrees of freedom		F	P
	Numerator	Denominator		
Selected vs. control[1]	1	6	11.78	0.0139
Sex	1	72	70.20	0.0001
Small muscle	1	72	921.96	0.0001
$\log_{10}$ body mass[2]	1	72	87.54	0.0001
Age	1	72	0.75	0.3897
Time of day	1	72	2.66	0.1074
Squared time of day	1	72	0.03	0.8706

[1] SAS PROC MIXED with line nested within linetype (selected vs. control) and family nested within line as random factors.
[2] Partial regression coefficient = 0.557 ± 0.0613 (± SE).
Least squares (adjusted) means ± SE are: 2.101 ± 0.0057 (control lines) and 2.077 ± 0.0050 (selected lines), 2.061 ± 0.0055 (females) and 2.116 ± 0.0049 (males), 2.213 ± 0.0030 (normal) and 1.964 ± 0.0076 (small-muscled mice).

TABLE 3. Nested analysis of covariance of $\log_{10}$ body mass (g) in relation to various factors and covariates for mice from generation 22 ($n = 158$). One individual removed as statistical outlier.

Source	Degrees of freedom		F	P
	Numerator	Denominator		
Selected vs. control[1]	1	6	12.09	0.0132
Sex	1	73	259.66	0.0001
Small muscle	1	73	20.33	0.0001
Age	1	73	12.86	0.0006
Time of day	1	73	1.32	0.2538
Squared time of day	1	73	0.14	0.7135

[1] SAS PROC MIXED with line nested within linetype (selected vs. control) and family nested within line as random factors.
Least squares (adjusted) means ± SE are: 1.533 ± 0.0129 (control lines) and 1.474 ± 0.0121 (selected lines), 1.459 ± 0.0096 (females) and 1.548 ± 0.0096 (males), 1.528 ± 0.0085 (normal) and 1.479 ± 0.0125 (small-muscled mice).

lected lines are smaller than those from controls, as has been reported previously for most generations (e.g., Koteja et al. 1999a,b; Swallow et al. 1999, 2001; Dumke et al. 2001; Koteja and Garland 2001; Girard et al. 2001 [but see Carter et al. 2000; Swallow et al. 1998a,b]).

Wheel running (revolutions/day, square-root transformed) of these mice, as measured during days 5 and 6 of the regular six-day test used for choosing breeders (see Swallow et al. 1998a), was also analyzed by nested ANCOVA (line within linetype, family within line; covariates were age at start of six-day test [mean = 57.6 days, range = 48–66] and the square root of a measure of wheel freeness [mean = 8.78, range = 4.9–16.1]; $n = 154$). As expected, mice from selected lines ran more than did controls ($P = 0.0014$) and females ran more than males ($P = 0.0001$), but individuals with the small-muscle phenotype did not differ significantly from other mice ($P = 0.3334$; normal 75.4 ± 2.51, small muscled 70.9 ± 4.64) (age had a significant negative effect [$P = 0.0233$] and wheel freeness a marginally positive effect [$P = 0.0782$]).

Additional analyses were performed after excluding mice from the control lines in order to test for replicate-line heterogeneity in the response to selection ($n = 76$). Again, females ran more than males ($P = 0.0001$), but individuals with the small-muscle phenotype did not differ significantly from other mice ($P = 0.6365$) and the four selected lines showed no differences ($P = 0.5680$) (age had a negative effect [$P = 0.0613$] and wheel freeness a nonsignificant positive effect [$P = 0.2255$]). When small muscle was excluded as a factor, sex remained significant ($P = 0.0001$) and line remained nonsignificant ($P = 0.5526$) (age $P = 0.0528$, freeness $P = 0.1733$). Thus, for the sample of generation-22 mice from which organs were dissected, we find no evidence for differences in wheel running among the selected lines.

Genetic Basis

The discrete nature of the phenotype provides some evidence for simple Mendelian inheritance. Initial observations of the phenotype only after several generations in some lines rule out the possibility that the phenotype is caused by a dominant allele. Both sexes express the phenotype in approximately equal frequency, indicating that the trait is neither sex linked nor sex limited. The simplest remaining explanation is that the trait is caused by a single recessive Mendelian factor. This hypothesis cannot easily be ruled out by examination of family-level pedigrees, especially because most observations of the trait are separated by several generations for which data on presence of the phenotype are unavailable.

Dissections of parents (generation 23) and offspring (generation 24) yielded distributions of the small-muscle phenotype that were generally consistent with a model of inheritance as a simple Mendelian recessive. As would be expected under such a model, all four of the putative homozygous recessive X homozygous recessive crosses (three in line 3, one in line 6) yielded only offspring with small muscles. For the nine line-6 families, none of the offspring phenotype frequencies deviated significantly from expected. For the nine line-3 families, two deviated significantly from ex-

pectations; both of these were putative heterozygote X homozygous recessive crosses, and in both cases all offspring, instead of the expected 50%, had small muscles. The pooled G-statistics for each type of cross indicated no deviation from the expected for matings of presumed heterozygote X heterozygote or homozygous recessives X homozygous recessive, but a significant excess of small-muscles offspring for matings of presumed heterozygotes with homozygous recessives ($G = 6.36$, $P < 0.02$). (Two families were left out of the pooled G calculations because it was difficult to determine the genotype of both parents and because they were one-of-a-kind crosses (one was probably homozygous dominant [wild type] X homozygous dominant, the other homozygous dominant X homozygous recessive.)

Selective Significance

A simple calculation shows that the occurrence of the small-muscle phenotype in two selected but only one control line does not, in itself, provide evidence for a selective advantage to the small-muscle phenotype. The total number of ways to assign four of the eight lines to each treatment group is

$$\binom{8}{4}.$$

Five of the eight lines do not have the small-muscle phenotype. The number of ways of assigning two of the five lines without the small-muscle phenotype *and* two of the three lines with the small-muscle phenotype to the selected treatment is

$$\binom{5}{2}\binom{3}{2}.$$

Thus, the probability of observing, by chance alone, two of the four selected lines with the small-muscle phenotype *and* two of the four selected lines without the small-muscle phenotype is

$$\binom{5}{2}\binom{3}{2} \bigg/ \binom{8}{4} = 0.43.$$

Therefore, the appearance of the small-muscle phenotype in two selected and one control line does not provide evidence for selective significance.

Table 4 summarizes parameter estimates for models of neutral and selective consequences of the small-muscle phenotype. The model assuming neutrality in both treatments (Model 1) estimates an initial allele frequency $p = 0.088$. The model of selection only in the control treatment (Model 2) estimates initial allele frequency as $p = 0.128$, with strong selection against the mutant allele ($s_c = -0.99$). The model of selection only in the selected treatment (Model 3) estimates initial allele frequency at $p = 0.067$, with strong selection in favor of the mutation ($s_+ = 0.60$). Allowing selection in both treatments (Model 4) reinforces results of the simpler models, with strong selection against the allele in the control treatment ($s_c = -0.99$) and strong selection in favor of the mutation in the selected treatment ($s_+ = 0.41$). Model 4 estimates an initial allele frequency of $p = 0.087$.

Table 4 also compares goodness of fit under the four mod-

TABLE 4. Parameter estimates, goodness of fit, and hypothesis tests for four models describing change in frequency of the recessive allele underlying the small-muscle phenotype (neutrality of the phenotype in both selected and control treatments; neutrality in the selected but not control treatment; neutrality in the control but not selected treatment; non-neutral evolution in both control and selected treatments). p is estimated initial frequency of the allele, s_c is estimated strength of selection in the control lines, s_+ is selection in the lines selected for high voluntary wheel running. Model 3 (bold) provides the most parsimonious fit and is illustrated in Figure 3. See text for discussion.

Model	p	s_c	s_+	χ^2	Test	P
1. Neutral	0.088			114.4		
2. Selection in control treatment	0.128	−0.99		109.2	2 vs. 1	<0.023
3. Selection in selected treatment	**0.067**		**0.60**	**104.7**	**3 vs. 1**	**<0.005**
4. Selection in both treatments	0.087	−0.99	0.41	103.1	4 vs. 1	<0.005
					4 vs. 2	<0.014
					4 vs. 3	≈0.201

els. The improvement in fit of Model 2 compared with the neutral model ($\Delta\chi^2 = 114.4 - 109.2 = 5.2$, df = 1) is statistically significant ($P < 0.023$), suggesting that the phenotype is strongly deleterious in the control treatment. However, the improvement in fit of Model 3 compared with the neutral model ($\Delta\chi^2 = 9.7$, 1 df) is even more highly significant ($P < 0.005$), suggesting that the phenotype is strongly advantageous under selection for voluntary wheel running. The improvement in fit of Model 4 (selection in both treatments) is significant compared to the model of selection only in the control treatment ($\Delta\chi^2 = 6.1$, 1 df, $P < 0.014$), but not when compared with the model including selection only in the selected treatment ($\Delta\chi^2 = 1.6$, 1 df, $P \approx 0.2$).

Therefore, the model with maximum statistically meaningful explanatory power involves selection favoring the small-muscle phenotype in the selected lines, but neutral evolution in the control lines. Adding an additional parameter (i.e., selection in the control treatment) does not provide sta-

tistically significant improvement of fit. Consequently, no additional parameters are added to the model. Although there is evidence for strong selection against the small-muscle phenotype in the control lines (i.e., improved fit of Model 2 compared with Model 1), this selection is not required for a parsimonious explanation of the data. It is important to note, however, that even strong selection against a recessive phenotype is difficult to detect when the phenotype is at low frequency.

Figure 3 shows the fit between Model 3 and treatment average phenotype frequency (with line contributions weighted by the inverse of the standard error, as outlined above). Model 3 fits the observed data well for the control lines, and for the early generations of the selected lines. Several statistical features contribute to the relatively poor fit to the data for the selected lines in later generations. First, sampling is more frequent during the early, rather than the late, generations. Therefore, processes acting during early generations contribute more to the fitted model than do processes acting during later generations. Phenotype and thus small-muscle allele frequencies are low during early generations in the selected lines, therefore the early samples involve a substantial random drift component. Higher phenotype frequencies in later generations of the selected line allow greater opportunity for selection to increase allele frequency. The less frequent sampling in later generations tends to decrease the importance of selection (i.e., cause underestimation of s_+) and result in a fitted curve below the actual data. Second, standard errors of intermediate frequencies are intrinsically large because of the binomial nature of the sampling variance. Because line means are weighted by the inverse of their standard errors, the later data points are less important for determining the fitted curve than are early data points. In the selected lines, this again emphasizes the importance of drift during early generations rather than selection during later generations. Finally, genetic drift over the course of the experiment causes variation within lines to be converted to variation among lines. This means, in effect, that there are more degrees of freedom (i.e., independent observations) early in the experiment compared to later in the experiment. Again, early generation observations are more important in determining fit than are later generations.
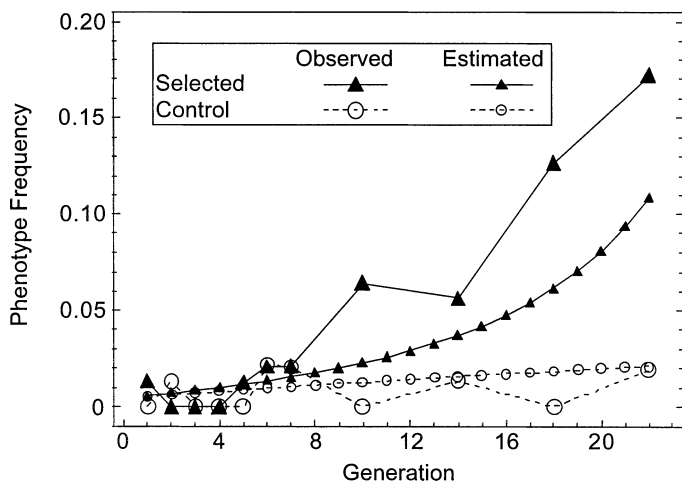

FIG. 3. Summary of most-parsimonious model (no. 3 in Table 4) of selection and/or random genetic drift, as fitted to cross-generation data on frequency of the small-muscle phenotype in control and selected lines of house mice. This model posits positive selection on the underlying recessive allele in the selected lines, but neutrality in the control lines. Large, solid triangles represent weighted (see text) average frequency (across all four replicate lines) of the small-muscle phenotype in the selected treatment. Large, open circles are weighted average frequency of the four replicate control lines. Smaller, solid triangles and open circles represent expected frequencies under the model of selection in the selected, but not control, lines.

DISCUSSION

Expression of the ''small-muscle'' phenotype during this selection experiment is consistent with inheritance of a single

autosomal recessive allele. The phenotype is not expressed in early generations of some lines, ruling out dominant inheritance. Both sexes express the phenotype, in approximately equal frequency, so the phenotype is neither sex linked nor sex limited. The simplest remaining genetic model is of an autosomal recessive allele, which is largely supported by the dissections of generation 23 parents and generation 24 offspring: none of the crosses produced ''impossible'' offspring (e.g., a normal offspring from two affected parents), and only two of 18 families deviated from expected phenotypic ratios in the offspring.

Many recessive alleles are associated with harmful phenotypic effects. Such alleles are introduced into a population through mutations that disrupt normal developmental or physiological pathways (Charlesworth and Charlesworth 1987). Deleterious recessive alleles can nonetheless attain relatively high frequency (e.g., 5–10%). This is because, under random mating, low-frequency alleles usually occur in heterozygous genotypes, where their phenotypic effects are not apparent, thus shielding them from selection. Allele frequency is then largely determined by the rate of mutation (and potentially migration, at least in natural populations), balanced by the influence of stochastic forces (i.e., genetic drift) and selection against the rare homozygotes (Crow and Kimura 1970). A reasonable hypothesis is that the allele responsible for the small-muscle phenotype was present at low frequency (approximately 7%, Table 4) in the base population. The improved fit of Model 2 compared with Model 1 provides evidence that the allele is deleterious under control conditions (Table 4), although no obvious life history or other behavioral differences have been detected in informal study. The frequency in the base population likely reflects a balance between mutation, selection, and genetic drift.

The most parsimonious model (Model 3) shows that strong selection favors the small-muscle phenotype in the selected lines. Two qualitative features of the data support this interpretation. First, the phenotype occurs in two of the selected but only one of the control lines. Second, the frequency of the phenotype in the selected lines rises to high values, but remains comparatively low in the control line (Fig. 1). That the underlying allele is beneficial in selected lines receives quantitative support from hierarchical models incorporating genetic drift and selection (Table 4, Fig. 3).

A potentially important caveat is that selection may act differently from inheritance. For instance, the recessive allele encoding the small-muscle phenotype may have dominant or overdominant effects on wheel running. It is possible to fit models like those in Table 4 to the data, but with dominant or overdominant fitness effects. Measures of goodness-of-fit of these alternative fitness effect models are very similar to those in Table 4 (M. T. Morgan, unpubl. data). In particular, alternative fitness effect models lead to qualitative conclusions about initial allele frequency and pattern of selection (i.e., favored in the selected but not control treatment) that are identical to those from the recessive model presented in Table 4. The estimated strength of selection is less under dominant or overdominant models (e.g., $s_c = 0.10$–0.12), consistent with heterozygote expression of fitness effects. Unfortunately, recessive, dominant, and overdominant models do not form a nested hierarchy, thus it is not possible to distinguish between alternative fitness effect models using goodness-of-fit criteria.

Why might selection be favoring small hindlimb muscles in highly active mice? The simplest possibility is that they function in a way that facilitates higher levels of wheel running. For example, reduced muscle mass in the limbs could reduce the energetic cost of cycling the limbs (e.g., see Myers and Steudel 1985; Steudel 1990). In this context, it is important to note that, in addition to effects of the small-muscle allele, mice from the selected lines show a further, general reduction in triceps surae mass, as compared with mice from the control lines (Table 1). Interestingly, Pollock et al. (1977) found reduced thigh girths in a sample of elite class human distance runners, as compared to average young men. This study did not find that calf (where the triceps surae is located) girths were reduced, but the parallel with our mice cannot be too closely drawn because mice are quadrupedal whereas humans are bipedal. Although other muscles were not weighed in the present study, it was apparent on dissection that the hindlimb of affected individuals had generally reduced muscle mass. Indeed, analyses of data from Houle-Leroy et al. (2000) indicate that total hindlimb muscle mass is also reduced by almost 50% (Houle-Leroy et al., unpubl. data). Further studies will be required to determine to what extent the small-muscle allele affects mass of individual muscles.

The small-muscle phenotype might also have other attributes that are conducive to running for long periods of time at relatively high, but still probably aerobic speeds (Girard et al. 2001). Preliminary observations (P. Houle-Leroy, T. Garland, J. G. Swallow, and H. Guderley, unpubl. data) gathered during a study of the effects of prolonged access to running wheels (Houle-Leroy et al. 2000) indicate that the small muscles have a very high per gram aerobic capacity. We are now planning further biochemical and histological studies to characterize the small-muscle phenotype with respect to functional properties.

Aside from possible direct positive effects on locomotor ability, the small-muscle allele might have beneficial pleiotropic effects for mice that engage in high levels of wheel running. For example, total body mass is reduced by approximately 13% in affected individuals (Table 3), presumably largely as a result of their reduced muscle mass, and this could be conducive to lowering energetic costs of locomotion. Ventricle (heart) mass is higher in small-muscled individuals (Table 2), which could enhance cardiac output. The liver, which is involved in exercise metabolism, is also larger in small-muscled animals, as is the kidney. Whether the small-muscle allele might have beneficial pleiotropic effects in heterozygotes is unknown because we are presently unable to detect its presence based on hindlimb muscle mass. In any case, the increase in frequency of the small-muscle allele constitutes an example of adaptation partly through selection on a gene of major effect. The importance of such ''major genes'' in natural populations has recently been reemphasized (Orr and Coyne 1992; Smith and Girman 2000).

Despite apparent strong selection favoring it, the small-muscle phenotype has been observed in only two of the four selection lines. Two reasons can explain why more selection lines do not exhibit the phenotype, both related to the ap-

parent low frequency of the allele in the base population. First, by chance, the sample of individuals chosen to establish each selection line may not have contained the allele (i.e., a founder effect). However, as lines were established with 10 male and 10 female parents, the probability that an allele at frequency 0.067 in the base population is not represented in a newly founded line is only $(1–0.067)^{40}$, or about 0.06. Therefore, most lines probably received one or more copies of the allele during founding. Second, and more likely, small-muscle alleles amongst founders might have been lost by random genetic drift during early generations, when most individuals remain heterozygous.

All four selected lines have exhibited a similar magnitude of response to selection for increased voluntary wheel running, at least based on analysis of the sample of mice dissected at generation 22. Our results suggest that the small-muscle allele represents one genetic mechanism underlying this response, but apparently only in two of the four lines. Thus, the genetic architecture of selection response differs among lines. Our observations of small-muscle phenotype frequency suggest that stochastic events during line establishment, coupled with genetic drift in early generations, have resulted in genetic differentiation among lines. Such among-line differences define the pool of genetic variation available for response to selection, setting the stage for unique genetic architecture of selection response in each line, even in the absence of unique mutations that might occur in one or more lines. Thus, genetic drift has been a potent force in shaping the genetic architecture of selection response in this experiment, as has been reported in other selection experiments with house mice (e.g., Rutledge et al. 1974; Falconer et al. 1978; Eisen and Pomp 1990; Hastings and Hill 1990; Bult and Lynch 1996) and with many other organisms (e.g., Endler et al. 2001), as may be expected on theoretical grounds (Cohan 1984; Gromko 1995).

LITERATURE CITED

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

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

Bennett, A. F., and R. B. Huey. 1990. Studying the evolution of physiological performance. Pp. 251–284 in D. J. Futuyma and J. Antonovics, eds. Oxford surveys in evolutionary biology. Vol. 7. Oxford Univ. Press, Oxford, U.K.

Boggs, D. F., and P. B. Frappell. 2000. Unifying principles of locomotion: foreword. Physiol. Biochem. Zool. 73:647–650.

Bult, A., and C. B. Lynch. 1996. Multiple selection responses in house mice bidirectionally selected for thermoregulatory nest-building behavior: crosses of replicate lines. Behav. Genet. 26: 439–446.

Bunnell, F. L., and A. S. Harestad. 1990. Activity budgets and body weight in mammals: How sloppy can mammals be? Pp. 245–305 in H. H. Genoways, ed. Current mammalogy. Vol. 2. Plenum, New York.

Carter, P. A., and W. B. Watt. 1988. Adaptation at specific loci. V. Metabolically adjacent enzyme loci may have very distinct experiences of selective pressures. Genetics 119:913–924.

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

Carter, P. A., J. G. Swallow, S. J. Davis, and T. Garland, Jr. 2000. Nesting behavior of house mice (Mus domesticus) selected for increased wheel-running activity. Behav. Genet. 30:85–94.

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. 18:237–268.

Cohan, F. M. 1984. Can uniform selection retard random genetic divergence between isolated conspecific populations? Evolution 38:495–504.

Crow, J. F., and M. Kimura. 1970. An introduction to population genetics theory. Harper and Row, New York.

Dewsbury, D. A. 1980. Wheel-running behavior in 12 species of muroid rodents. Behav. Processes 6:271–280.

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. 267(Regulatory Integrative Comp. Physiol. 36):R1098–R1108.

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.

Dumke, C. L., J. S. Rhodes, T. Garland, Jr., E. Maslowski, J. G. Swallow, A..C. Wetter, and G. D. Cartee. 2001. Genetic selection of mice for high voluntary wheel running: effect on skeletal muscle glucose uptake. J. Appl. Physiol. 91:1289–1297.

Eisen, E. J., and D. Pomp. 1990. Replicate differences in lines of mice selected for body composition. Genome 33:294–301.

Endler, J. A., A. Basolo, S. Glowacki, and J. Zerr. 2001. Variation in response to artificial selection for light sensitivity in guppies (Poecilia reticulata). Am. Nat. 158:36–48.

Falconer, D. S., I. K. Gauld, and R. C. Roberts. 1978. Cell numbers and cell sizes in organs of mice selected for large and small body size. Genet. Res., Camb. 31:287–301.

Feder, M. E., A. F. Bennett, and R. B. Huey. 2000. Evolutionary physiology. Annu. Rev. Ecol. Syst. 31:315–341.

Garland, T., Jr. 1999. Laboratory endurance capacity predicts variation in field locomotor behaviour among lizard species. Anim. Behav. 57:77–83.

———. 2001. Phylogenetic comparison and artificial selection: two approaches in evolutionary physiology. Pp. 107–132 in R. C. Roach, P. D. Wagner, and P. H. Hackett, eds. Hypoxia: from genes to the bedside. Advances in Experimental Biology and Medicine, vol. 502. Kluwer Academic/Plenum Publishers, New York.

———. 2002. Selection experiments: an underutilized tool in biomechanics and organismal biology. In V. L. Bels, J.-P. Gasc, and A. Casinos, eds. Biomechanics and evolution. Bios Scientific Publishers, Oxford, U.K. In press.

Garland, T., Jr.,and J. B. Losos. 1994. Ecological morphology of locomotor performance in squamate reptiles. Pp. 240–302 in P. C. Wainwright and S. M. Reilly, eds. Ecological morphology: integrative organismal biology. Univ. of Chicago Press, Chicago, IL.

Garland, T., Jr., J. G. Swallow, I. Girard, J. S. Rhodes, P. Houle-Leroy, H. Guderley, P. W. Freeman, C. L. Dumke, G. D. Cartee, P. Koteja, M. W. McAleer, G. R. Hosack, J. G. Belter, and P. A. Carter. 2000. Exercise adaptations in lines of house mice genetically selected for high voluntary wheel-running behavior. Physiologist 43:328.

Gibbs, A. G. 1999. Laboratory selection for the comparative physiologist. J. Exp. Biol. 202:2709–2718.

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

Gromko, M. H. 1995. Unpredictability of correlated responses to

selection: pleiotropy and sampling interact. Evolution 49:685–693.

Hastings, I. M., and W. G. Hill. 1990. Analysis of lines of mice selected for fat content. 2. Correlated responses in the activities of enzymes involved in lipogenesis. Genet. Res., Camb. 55:55–61.

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

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

Kelt, D. A., and D. H. Van Vuren. 2001. The ecology and macroecology of mammalian home range area. Am. Nat. 157:637–645.

Koteja, P., and T. Garland, Jr. 2001. Forum: response to R. Eikelboom. Anim. Behav. 61:F25–F26.

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

Koteja, P., T. Garland, Jr., J. K. Sax, J. G. Swallow, and P. A. Carter. 1999b. Behaviour of house mice artificially selected for high levels of voluntary wheel running. Anim. Behav. 58:1307–1318.

Kramer, D. L., and R. L. McLaughlin. 2001. The behavioral ecology of intermittent locomotion. Am. Zool. 41:137–153.

Miles, D. B., B. Sinervo, and W. A. Frankino. 2000. Reproductive burden, locomotor performance, and the cost of reproduction in free ranging lizards. Evolution 54:1386–1395.

Myers, M. J., and K. Steudel. 1985. Effect of limb mass and its distribution on the energetic cost of running. J. Exp. Biol. 116:363–373.

Nunn, C. L., and R. A. Barton. 2000. Allometric slopes and independent contrasts: a comparative test of Kleiber's Law in primate ranging patterns. Am. Nat. 156:519–533.

Orr, H. A., and J. A. Coyne. 1992. The genetics of adaptation: a reassessment. Am. Nat. 140:725–742.

Perry, G. 1999. The evolution of search modes: ecological versus phylogenetic perspectives. Am. Nat. 153:98–109.

Perry, G., and T. Garland, Jr. 2002. Lizard home ranges revisited: effects of sex, body size, diet, habitat, and phylogeny. Ecology *In press.*

Pollock, M. L., L. R. Gettman, A. Jackson, J. Ayres, A. Ward, and A. C. Linnerud. 1977. Body composition of elite class distance runners. Ann. N. Y. Acad. Sci. 301:361–370.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery.

1992. Numerical recipes in C. 2d ed. Cambridge Univ. Press, Cambridge, U.K.

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

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

Rose, M. R., T. J. Nusbaum, and A. K. Chippindale. 1996. Laboratory evolution: the experimental wonderland and the Cheshire Cat syndrome. Pp. 221–241 *in* M. R. Rose and G. V. Lauder, eds. Adaptation. Academic Press, San Diego, CA.

Rutledge, J. J., E. J. Eisen, and J. E. Legates. 1974. Correlated responses in skeletal traits and replicate variation in selected lines of mice. Theor. Appl. Genet. 45:26–31.

SAS Institute. 1996. SAS PROC MIXED. Vers. 6.12. SAS Institute, Inc., Cary, NC.

Smith, T. B., and D. J. Girman. 2000. Reaching new adaptive peaks: evolution of alternative bill forms in an African finch. Pp. 139–156 *in* T. A. Mousseau, B. Sinervo, and J. A. Endler, eds. Adaptive genetic variation in the wild. Oxford Univ. Press, New York.

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

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

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

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

———. 2001. Food consumption and body composition in mice selected for high wheel-running activity. J. Comp. Physiol. B 171:651–659.

Van Damme, R., B. Vanhooydonck, P. Aerts, and F. De Vree. 2002. Evolution of lizard locomotion: context and constraint. *In* V. L. Bels, J.-P. Gasc, and A. Casinos, eds. Biomechanics and evolution. Bios Scientific Publishers, Oxford, U.K. *In press.*

WaveMetrics, Inc. 1996. Igor Pro User's Guide. Vers. 3. WaveMetrics Inc,, Lake Oswego, OR.