OXFORD GENETICS

GENETICS, 2023, 223(1), iyac165
https://doi.org/10.1093/genetics/iyac165
Advance Access Publication Date: 28 October 2022
Investigation

Multiple solutions at the genomic level in response to selective breeding for high locomotor activity

David A. Hillis [1], Theodore Garland Jr [2,*]

[1]Genetics, Genomics, and Bioinformatics Graduate Program, University of California, Riverside, CA 92521, USA,
[2]Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, CA 92521, USA

*Corresponding author: Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, CA 92521, USA. Email: tgarland@ucr.edu

Abstract

Replicate lines under uniform selection often evolve in different ways. Previously, analyses using whole-genome sequence data for individual mice (*Mus musculus*) from 4 replicate High Runner lines and 4 nonselected control lines demonstrated genomic regions that have responded consistently to selection for voluntary wheel-running behavior. Here, we ask whether the High Runner lines have evolved differently from each other, even though they reached selection limits at similar levels. We focus on 1 High Runner line (HR3) that became fixed for a mutation at a gene of major effect (*Myh4*^{Minimsc}) that, in the homozygous condition, causes a 50% reduction in hindlimb muscle mass and many pleiotropic effects. We excluded HR3 from SNP analyses and identified 19 regions not consistently identified in analyses with all 4 lines. Repeating analyses while dropping each of the other High Runner lines identified 12, 8, and 6 such regions. (Of these 45 regions, 37 were unique.) These results suggest that each High Runner line indeed responded to selection somewhat uniquely, but also that HR3 is the most distinct. We then applied 2 additional analytical approaches when dropping HR3 only (based on haplotypes and nonstatistical tests involving fixation patterns). All 3 approaches identified 7 new regions (as compared with analyses using all 4 High Runner lines) that include genes associated with activity levels, dopamine signaling, hippocampus morphology, heart size, and body size, all of which differ between High Runner and control lines. Our results illustrate how multiple solutions and "private" alleles can obscure general signatures of selection involving "public" alleles.

Keywords: complex traits; gene of major effect; experimental evolution; multiple solutions; polygenic adaptation; mammalian genetics

Introduction

By their very nature, complex traits can evolve in multiple ways. Thus, when a given form of directional selection is applied to replicate lines, adaptive responses are likely to be somewhat different (Mayr 1961; Cohan 1984a, 1984b; Tenaillon *et al.* 2012; Wone *et al.* 2019), a phenomenon often termed multiple solutions (e.g. see Bock 1959; Bennett 2003; Garland *et al.* 2011). These variable evolutionary pathways underscore the versatility of the genome and also provide opportunities for insight concerning the developmental and physiological mechanisms that underlie variation in complex traits.

The particular genomic and/or genetic features and processes that underlie a complex trait may affect the likelihood of multiple adaptive responses to a given type of selection. For example, duplications can create redundancy in genes, thus enabling altered function in 1 or both copies without detrimental effect on the organism. This has been seen in myosin *MLC2* genes (Gerrits *et al.* 2012) and in hemoglobin (Natarajan *et al.* 2015; Storz 2016). Multiple solutions can also be modulated by highly impactful single-nucleotide polymorphisms (SNPs). A well-known example of this is malaria. Here, infection by a parasitic *Plasmodium* invokes typical immunological responses (Malaguarnera and Musumeci 2002), with a lethality rate of up to 30% in severe cases (i.e. multisyndromic and often manifesting as cerebral malaria, severe malaria anemia, and respiratory distress) (Karlsson *et al.* 2014). However, the sickle-cell mutation, which is an A-to-T substitution causing glutamate to be substituted with valine in the β-globin gene, is associated with substantial resistance to the disease in both heterozygotes and homozygotes, but with notable health detriments in homozygotes (Aluoch 1997; Griffiths *et al.* 2015). Despite the deleterious pleiotropic effects of the sickle-cell mutation, this allele has been favored by selection in populations where malaria is present (Karlsson *et al.* 2014), thus providing an alternative solution to the typical immunological responses.

One genomic feature that may promote multiple adaptive solutions is the presence of so-called genes of major effect (GOMEs), also referred to as major QTL, which are defined as genes whose allelic variants explain a large proportion of quantitative variation (Tanksley 1993). GOMEs may enhance the probability of divergent genomic pathways among replicate lines by affecting genetic variances and covariances (Agrawal *et al.* 2001; Garland 2003; Hannon *et al.* 2008; Stinchcombe *et al.* 2009). For example, Stinchcombe *et al.* (2009) demonstrated that the ERECTA allele in *Arabidopsis thaliana* had a small but clear impact on the G-matrix structure, although without a discernable impact on the response to selection. Epistatic genetic variance is also likely enhanced by

Received: September 20, 2022. **Accepted:** October 14, 2022

the presence of GOMEs. Thus, if some populations have a given GOME and others do not, then they are likely to evolve genetically in somewhat different ways. (As noted in the *Discussion*, founder effects and random genetic drift can also increase the likelihood of multiple responses to selection.)

Replicated selection experiments offer excellent opportunities for discovering multiple adaptive responses to a defined and reproducible selective regime (Garland 2003; Garland and Rose 2009). Here, we test for multiple genomic responses to selection in the context of a replicated selection experiment that has a well-documented GOME that causes a phenotype termed mini-muscle (Garland *et al.* 2002: see below). Specifically, the High Runner (HR) mouse experiment includes 4 replicate lines of mice that have been bred (within family) for long daily distances of voluntary running on a wheel (HR3, HR6, HR7, and HR8) and 4 non-selected control lines (C1, C2, C4, and C5) (Swallow *et al.* 1998). A statistically significant response to selection could be detected by generation 6, and all lines reached selection limits around generations 17–27, running on average 2.5–3 times more than the control line (Careau *et al.* 2013). The 4 replicate HR lines vary in the extent to which daily running distance has evolved via increases in average speed vs duration of running, and a significant negative correlation between average running speed and duration of daily activity had evolved among the HR lines by generation 43 (Garland *et al.* 2011). For example, on average, mice from line HR3 (which became fixed for the allele underlying the mini-muscle phenotype) run faster but for fewer minutes per day than other HR lines, whereas the opposite is true for HR8 (see Fig. 3 in Garland *et al.* 2011).

Numerous differences among the HR lines have been identified at various points during the selection experiment, although these results have yet to be synthesized or approached from the perspective of a meta-analysis. These include pleiotropic effects attributable to the mini-muscle allele ($Myh4^{Minimsc}$) when in the homozygous state, in addition to differences that do not involve the mini-muscle phenotype. Those nonmini-muscle differences among the replicate HR lines have been documented for a variety of traits at the level of behavior, whole-animal performance, morphology, and physiology. For example, the HR lines have been shown to differ in both male-male (Klomberg *et al.* 2002) and maternal aggression (Gammie *et al.* 2003), as well as behavior in an open-field arena test (measuring aspects of exploration and risk-taking behavior) and in a plus maze (measuring aspects of anxiety) (Jónás *et al.* 2010). Among-line differences in performance and physiology have been documented for daily energy expenditure (Rezende *et al.* 2009), basal metabolic rate (Kane *et al.* 2008), endurance capacity during forced treadmill exercise (Meek *et al.* 2009), the ability to clear a parasitic nematode species (*Nippostrongylus brasiliensis*) from the small intestine (Malisch *et al.* 2009), and circulating corticosterone levels under both baseline conditions and after 40 min of restraint stress (Malisch *et al.* 2007), among other traits. Body mass differs among the HR lines (e.g. Klomberg *et al.* 2002; Hiramatsu *et al.* 2017), as do the masses of individual hindlimb muscles (controlling statistically for variation in body mass and even excluding those with the mini-muscle phenotype)(Houle-Leroy *et al.* 2003). Muscle fiber-type composition differs among lines and, at the level of muscle biochemistry, HR lines differ in the mass-specific activities of various metabolic enzymes (e.g. palmitoyl transferase, citrate synthase, cytochrome C oxidase) (Guderley *et al.* 2008). As these differences are in traits of functional relevance for endurance running, they suggest multiple solutions.

The mini-muscle phenotype noted above is caused by the recessive $Myh4^{Minimsc}$ allele (a single base pair replacement) at the *Minimsc* locus in the eleventh intron of the *Myosin heavy polypeptide 4* gene (chr11:67,244,850, GRCm38/mm10 assembly) (Kelly *et al.* 2013). The mini-muscle GOME was serendipitously discovered relatively early in the HR mouse selection experiment, based on systematic muscle dissections (Garland *et al.* 2002). The $Myh4^{Minimsc}$ allele was uncommon in the base population (frequency ~7%) and the phenotype has only been observed in 2 of the HR lines and in 1 control line (Garland *et al.* 2002; Syme *et al.* 2005). Of these lines, C5 apparently lost the allele to drift by generation 36 (Syme *et al.* 2005), HR3 becomes fixed for the $Myh4^{Minimsc}$ allele by generation 36 (Garland *et al.* 2002; Syme *et al.* 2005), and HR6 has remained polymorphic for the allele through generation 98 (unpublished data; Cadney *et al.* 2021). Population genetic modeling indicated positive selection on the allele in the HR lines and either neutrality or negative selection in the C lines (Garland *et al.* 2002).

When present in the homozygous condition, the $Myh4^{Minimsc}$ allele causes a 50% reduction of the mass of the triceps surae (calf) muscle, as well as total hindlimb muscle mass, earning it the name mini-muscle allele or phenotype (Garland *et al.* 2002; Hannon *et al.* 2008; Bilodeau *et al.* 2009; Kelly *et al.* 2013). The $Myh4^{Minimsc}$ allele has a variety of pleiotropic effects when in the homozygous condition, such as increasing the mass of several organs, including the heart, spleen, liver, kidney, lung, stomach, and soleus muscle (Garland *et al.* 2002; Swallow *et al.* 2005; Syme *et al.* 2005; Guderley *et al.* 2006; Hannon *et al.* 2008; Kelly *et al.* 2017; and references therein), and altering the size and/or shape of various skeletal elements (Castro, Karakostis, *et al.* 2021; Castro, Rabitoy, *et al.* 2021). Possible effects in heterozygotes have not yet been studied. Perhaps most relevant for the concept of multiple responses to selection, mice from line HR3 (fixed for mini-muscle) and mini-muscle individuals in general tend to run faster but for fewer minutes per day as compared with the other HR lines (Kelly *et al.* 2006; Hannon *et al.* 2008; Dlugosz *et al.* 2009; Garland *et al.* 2011).

Loci with such far-reaching pleiotropic effects as mini-muscle have great potential to result in nonadditive epistatic effects with other genes, which may enhance their benefits or compensate their detriments (Pavlicev and Wagner 2012). Thus, we expected that the genomic basis of high voluntary wheel running in HR3—beyond the change in frequency of this 1 underlying allele—would differ from that of the other 3 HR lines. Although previous analyses involving all HR lines detected signatures of selection at various genomic regions (Hillis *et al.* 2020), we hypothesized that fixation for the $Myh4^{Minimsc}$ allele in HR3 may mask additional signatures when this genetically divergent line is included in the analyses. To test this, we have repeated analyses using single-nucleotide polymorphism (SNP) data, dropping each of the HR lines. After confirming that dropping HR3 produced more novel selection signatures than when dropping any other HR line, we incorporated additional analyses used by Hillis *et al.* (2020) to highlight signatures of selection. Overall, our results illustrate how multiple solutions and "private" alleles (those unique to 1 or 2 lines) can obscure general signatures of selection involving "public" alleles (those present in all lines) (cf. Partridge and Gems 2002).

Materials and methods

HR mouse model

As described previously (Swallow *et al.* 1998; Careau *et al.* 2013; Hillis *et al.* 2020), 112 male and 112 female mice were obtained from Harlan Sprague Dawley (outbred Hsd: ICR strain) in 1993.

Following 2 generations of random mating, 10 breeding pairs were randomly chosen to be founders for each of 8 closed lines (generation 0). Four of these lines were randomly designated as HR lines (lab designated HR3, HR6, HR7, and HR8), which would undergo selection based on voluntary wheel running. The remaining 4 lines would serve as unselected control (C) lines (lab designated C1, C2, C4, and C5) (Fig. 1). Each generation, all mice were given access to wheels at 6–8 weeks of age for 6 days. The highest-running (total revolutions on days 5 plus 6) male and female of each HR family were used to propagate the line (within-family selection, no sib-mating). This selection criterion was continued even after reaching selection limits at around generation 17–27 (Careau *et al.* 2013). The male and female from each C family were chosen randomly with respect to wheel running.

Whole-genome sequencing

As described previously (Hillis *et al.* 2020), DNA was collected from 80 mice (10 from each line), from generation 61, via phenol-chloroform extraction and sequenced on an Illumina HiSeq 2500 1T platform. Libraries were constructed using Nextera kit and reads were trimmed and aligned to the GRCm38/mm10 mouse genome assembly as described in Didion *et al.* (2016). This generated an average read depth of 12× per mouse. SNPs were filtered to keep those with genotype quality ("GQ") >5, read depth >3, MAF >0.0126 for all samples (as done by Hillis *et al.* 2020 to preserve all variable loci in data set), and Mapping Quality ("MQ") >30. Of the 80 mice, 1 was excluded due to likely contamination, as in Xu and Garland (2017), leaving 79 for the following analyses. SNPs not found to be present in at least 2 of the 80 mice were also removed from analysis. This leaves 5,932,148 SNPs for analyses involving all 8 lines. The number decreased when dropping certain lines due to the remaining 7 lines being fixed for the same allele. Although Xu and Garland (2017) had identified these 80 mice from generation 61 as females, they were in fact all males with exception of 1 female from line C5.

Principal components analysis

Principal components analysis (PCA) was performed in R with the SNPRelate library (Zheng *et al.* 2012). Of the 5,932,148 variable SNPs across all lines (HR and C), we used 4,679,533 variable SNPs across the 9–10 mice within each of the HR lines.

SNP analyses excluding individual HR lines

To assess the hypothesis that fixation of the mini-muscle allele would cause HR3 to differentiate from the other HR lines in genomic regions relevant to wheel-running behavior, a mixed-model ANOVA was used to calculate differentiation between C and HR lines while dropping each of the other HR lines. The mixed-model ANOVA used minimum variance quadratic unbiased estimation (mivque) method of variance estimation (Rao 1971; Xu and Garland 2017). In addition, *P*-values and the Aikake Information Criterion corrected (AICc) for small sample sizes were calculated for 4 models with different variance structures (equal within-line variance for HR and C lines, equal among-line variance, both variances equal, and both variances different) and we then used AICc scores to choose the best model (following Hillis *et al.* 2020). The results of each of these analyses were then compared, with the expectation that more selection signatures would be present after dropping HR3 as compared with dropping any other HR line. "Differentiated regions" were defined by the following 3-step process. First, we identified all SNPs differentiated with *P*-value $\leq$0.001. Second, we considered that any 2 such SNPs within 1mbp of each other were part of the same region. Finally, we considered any gap between SNPs (with $P \leq 0.001$) larger than 1 mbp as delineating separate regions.

Power and type I error simulations

All else being equal, dropping one of the 8 lines from the analyses would be expected to reduce the power to detect differentiation between the HR and C lines, due to the loss of a denominator degree of freedom. To estimate this expected drop in power, we performed simulations. Data reflecting the alternative hypothesis were simulated by taking a region from chromosome 17 that had been shown to be differentiated by Hillis *et al.* (2020). Approximately 22,700 SNP loci in this region (chr17:17,846,983–23,586,163) were variable and were differentiated across the region for the 8-line analyses (mean *P*-value = 0.104, median = 0.137, lowest = 7.54E−05, highest = 0.952). To generate simulated data, a variable locus was randomly sampled from the region, then the alleles for each line were created by randomly sampling (with replacement) from the alleles at that locus for that line. This was done for each of the 8 lines and the whole process was repeated to produce 100,000 simulated loci. Membership of each line within the set of either HR or C lines was always retained. Simulated data were analyzed using the multimodel ANOVA method (Hillis *et al.* 2020), first with all 8 lines and then dropping each of the HR lines 1 at a time.

For calculating relative type I error rate, data reflecting the null hypothesis were generated with a method similar to that for the power analyses. Alleles were sampled (with replacement) from a single line in the previously indicated chromosome 17 region, but then assigned to any of the 8 lines at random. This process was repeated for all 8 lines in sequence. A total of 100,000 loci were thus created, and multimodel ANOVA was performed with all 8 lines as well as dropping each of the HR lines.

Haplotype and nonstatistical analyses excluding HR3 (mini-muscle)

Following Hillis *et al.* (2020), we performed 2 additional analyses to gauge differentiation between the 4 C and 3 HR lines (excluding HR3). First, we used the haplotype data that were used by Hillis *et al.* (2020) and applied the mixed-model ANOVA method used for the SNP analyses, dropping HR3. A critical threshold of $P \leq 0.00526$ was used for these haplotype analyses, following Hillis *et al.* (2020). Next, loci that were fixed for a given allele (either reference or alternative) for all HR lines (excluding HR3) and simultaneously polymorphic for all C lines, were identified as "FixedHR/PolyC" (as in Hillis *et al.* 2020). Any loci or genomic region identified as differentiated in all 3 tests (SNP ANOVA, haplotype ANOVA, and FixedHR/PolyC) are referred to as "consistent" regions and regarded as having the strongest evidence of differentiation. Selection signatures implicated by these analyses were compared to those implicated by analyses including all 8 lines (as reported in Hillis *et al.* 2020).

Gene annotations and knockout phenotyping

Gene annotations were determined using the University of California, Santa Cruz Genome Browser for GRCm38/mm10 (http://genome.ucsc.edu/, accessed Oct 2021) (Kent *et al.* 2002) and the Rat Genome Database for the GRCm38/mm10 mouse genome browser (https://rgd.mcw.edu/, accessed May 2022) (Smith *et al.* 2019). Mouse Genome Informatics' Batch Query database was used for the knockout phenotyping (http://www.informatics.jax.org/batch/, accessed Nov 2021) (Bult *et al.* 2019).

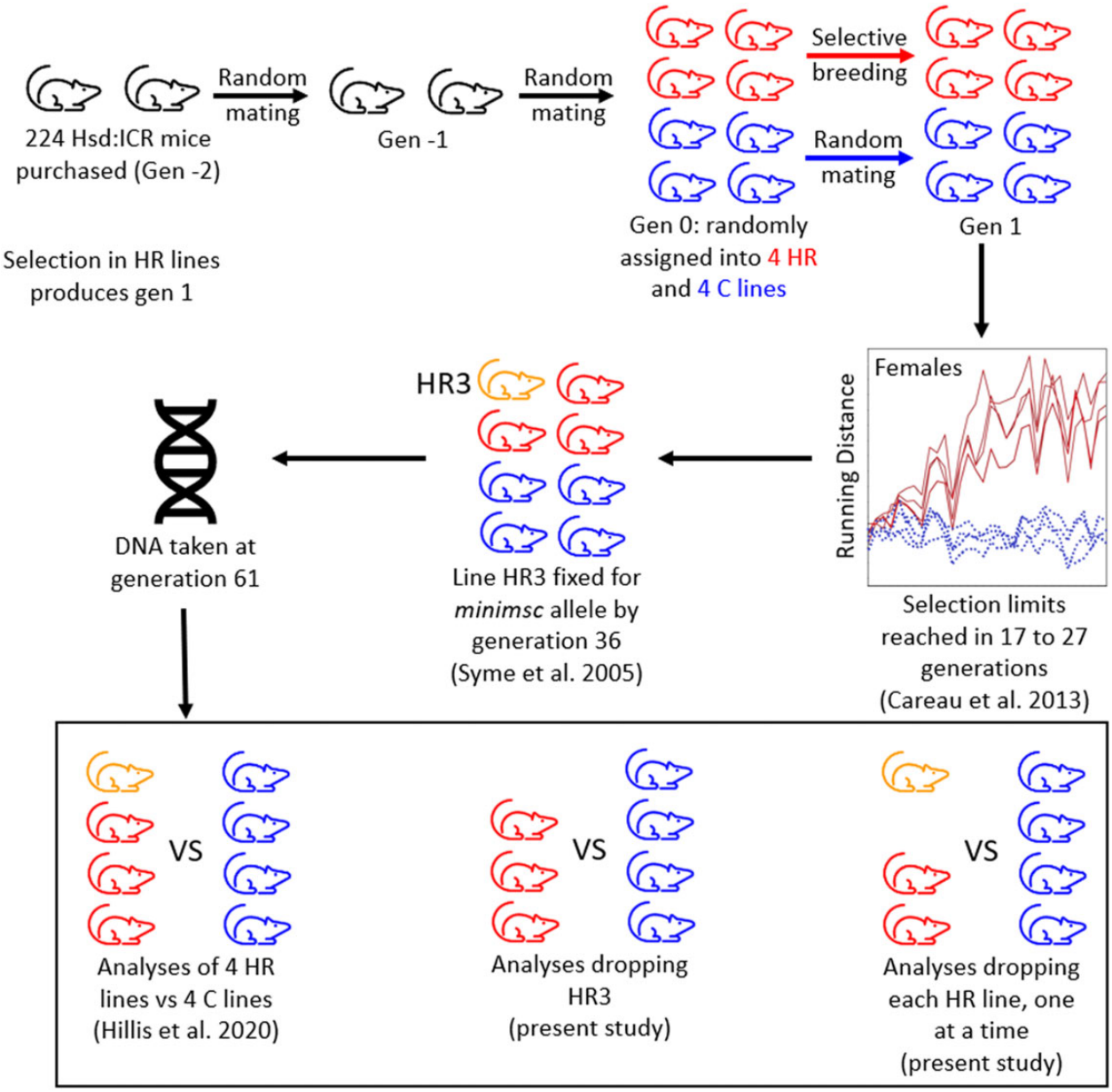

Fig. 1. Schematic illustration of the HR mouse artificial selection experiment, begun in 1993 with a base population of 224 outbred mice.

Results

Principal components analysis

PCA across all lines (79 individuals) produced 7 eigenvalues >1: PC1 = 10.6 (13.6% of variance), PC2 = 9.4 (12.0% of variance), PC3 = 8.8 (11.3% of variance), PC4 = 8.4 (10.7% of variance), PC5 = 7.9 (10.1% of variance), PC6 = 6.9 (8.9% of variance), and PC7 = 6.8 (8.7% of variance). The 3D scatterplot of eigenvectors for PC1, PC3, and PC5 demonstrates a clear differentiation between the HR and C lines and also that HR3 differs from other HR lines (Fig. 2a).

PCA of the 39 individuals in the HR lines included 4,679,533 variable SNP loci and produced 3 eigenvalues >1: PC1 = 10.0 (26.3% of variance), PC2 = 8.8 (23.1%), and PC3 = 7.7 (20.2%). Line HR3 was remarkably different from the other 3 HR lines for scores on PC2 (Fig. 2, b–d).

SNP analyses excluding individual HR lines

Each of the analyses dropping 1 of the HR lines produced some new peaks, as compared with the original analyses (Fig. 3). However, dropping HR3 produced generally lower *P*-values across the genome than dropping any of the other HR lines (paired *t*-test, $t = -149.91$, -126.2, and -163.56, when comparing results after HR3 to dropping lines HR6, HR7, and HR8, respectively). The overall reduction in *P*-values when dropping HR3 is due largely to the increase in SNPs with $P < 0.001$, which is 4 times greater than in the analyses including all 4 HR lines (Table 1). More specifically, this difference is attributable mainly to a large increase in loci with *P*-values in this range in 2 genomic regions (chr3:46,438,071–52,624,971 and chr10:101,652,005–106,038,129). Both regions contain some loci with *P*-values ≤1e−03 after dropping any of the other lines; however, dropping HR3 produces

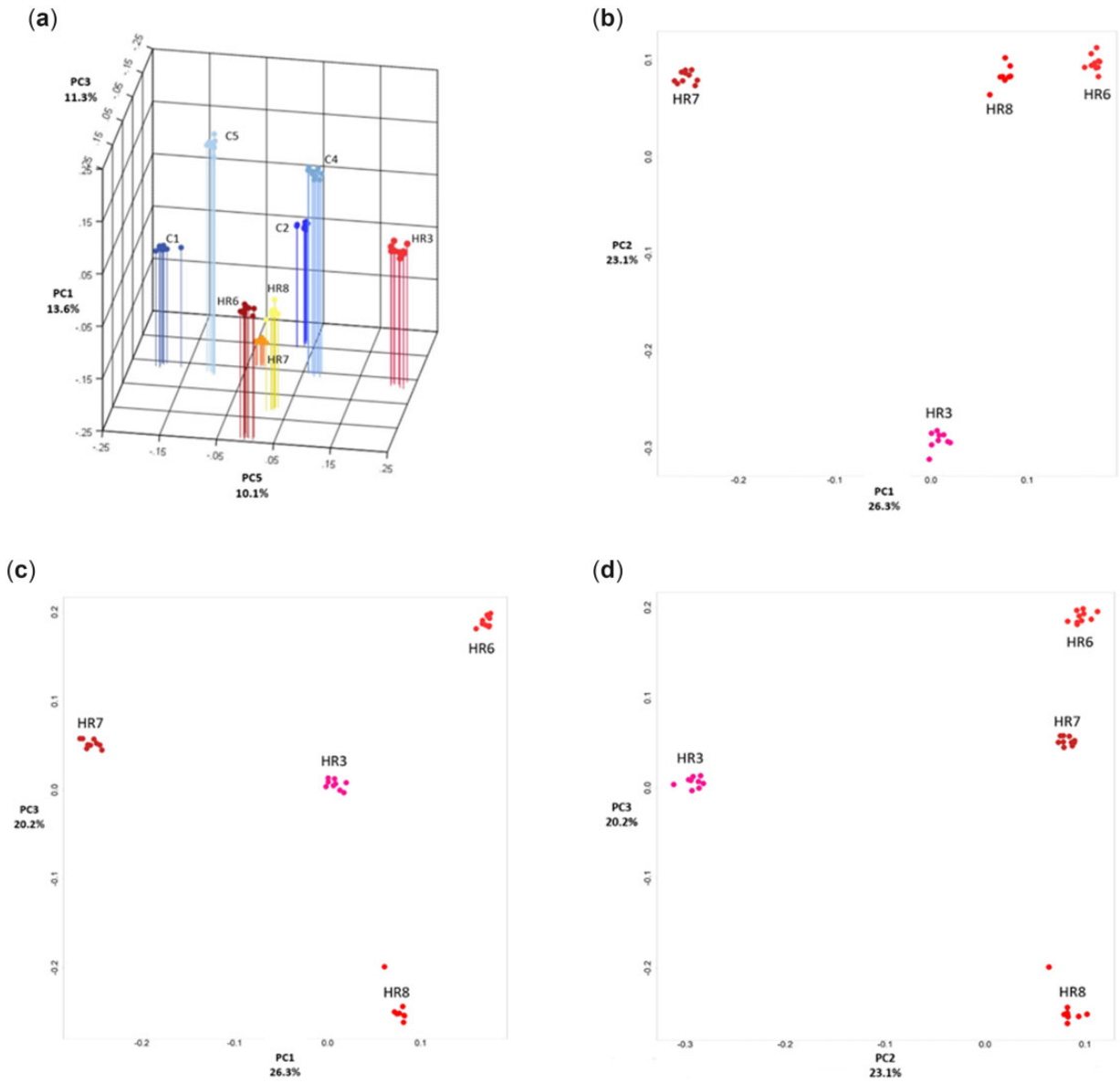

Fig. 2. Principal components analysis of variable SNP loci. a) Scatterplot of principle components (PCs) using all 8 lines: PC1, PC3, and PC5 account for a combined total of 35.0% of the variance. b) Bivariate scatterplot of scores on PC1 vs PC3 (49.4% of variance), from an analysis of variable SNP loci for the 4 HR lines only. c) PC1 vs PC3 (46.5% of variance). d) PC2 vs PC3 (43.3% of variance). As can be seen in (b) and (d), line HR3 is very different from the other 3 HR lines for scores on PC2.

about 40,000 additional loci with low *P*-values in these 2 regions (Table 2). Dropping HR8 resulted in a notable increase in loci with *P*-values in the 1e−06 to 1e−08 range (Table 1), largely due to a single region containing 1,414 loci with uniquely low *P*-values (chr7:115,169,726–116,129,821) (Table 2). This region contains *Sox6*, a gene whose knockout phenotypes include abnormal skeletal muscle fiber-type ratio (van Rooij *et al.* 2009), and was also identified in the original 8-line analyses (Hillis *et al.* 2020).

Although Table 1 seems to generally show that dropping HR3 produces more differentiated regions ($N = 75$) than dropping any other HR line ($N = 63$–70), some of these regions will contain only 1 or a few SNPs, which may be a result of sampling error and thus a Type I error (see section on Type I error, below). Therefore, Table 2 and Supplementary Table 1 concentrate on those regions with at least 10 SNPs with $P \leq 0.001$.

Supplementary Table 1 contains all regions with at least 10 SNPs with $P \leq 0.001$ for any analyses where an HR line was dropped. Dropping HR3 from the analyses resulted in 34 such regions, which is more than those identified after dropping any of the other HR lines (noHR6 = 23 regions, noHR7 = 27 regions, and noHR8 = 19) and also more than the 21 regions that were produced when analyzing all 8 lines.

The 45 regions listed in Table 2 are a subset of those shown in Supplementary Table 1, excluding regions where similar numbers of SNPs with $P \leq 0.001$ were produced when dropping any HR line. The regions in Table 2 are highlighted because they are where the HR lines responded differently from each other to the selection protocol. This leaves regions with (1) at least 10 SNPs with $P \leq 0.001$ after dropping only 1 specific HR line (e.g. chr1:155,052,375–157,767,127, see Fig. 4 for illustration) or (2) a substantial increase in significant loci when dropping a specific line (e.g. chr3:46,438,071–52,624,971). Of the 45 regions listed in Table 2, dropping the line fixed for $Myh4^{Minimsc}$ (HR3) produced more of these unique regions than dropping any other line (19

Table 1. Number of SNPs below different *P*-value thresholds in generation 61 individual mouse analyses ($N = 5{,}923{,}148$).

P threshold	8-Line	Regions	No HR3	Regions	No HR6	Regions	No HR7	Regions	No HR8	Regions
0.05	248,549	455	315,791	474	283,789	467	275,305	488	296,948	433
0.01	87,600	129	132,865	196	100,845	185	88,311	193	97,760	173
0.001	16,844	44	68,372	75	16,561	70	23,107	63	21,256	69
0.0001	5,667	19	15,852	35	6,237	23	6,694	19	8,308	18
0.00001	3,522	6	4,862	19	3,924	12	1,364	11	2,806	9
0.000001	984	3	482	8	272	6	167	7	2,176	5
1.00E-07	2	1	215	6	7	4	138	5	1,148	4
1.00E-08	2	1	207	5	2	1	126	5	608	4
1.00E-09	0	0	155[a]	4	0	0	63[a]	1	485[a]	3

Number of SNPs below different *P*-value thresholds in generation 61 individual mouse analyses ($N = 5{,}923{,}148$). Results for the 8-line comparison are from Hillis *et al.* (2020). This includes 155 from the analyses that excluded line HR3 that were fixed for opposite alleles in the C and HR lines producing low *P*-values. See Fig. 3 for the Manhattan plot of these regions (Fig. 3e for 8-line Manhattan plot).

[a] These SNPs have very low *P*-values (2E−39) due to fixation for opposite alleles in the C vs HR lines.

regions for HR3, 8 regions for HR6, 12 regions for HR7, and 6 regions for HR8).

Although none of the SNPs were fixed for opposite alleles between all 4 HR and 4 C lines (Hillis *et al.* 2020), dropping individual lines did produce some loci where the remaining C and HR lines were fixed for opposite alleles (Table 3). When dropping HR3, 155 SNPs are fixed for opposite alleles between the C and HR lines, clustered in 4 regions. Dropping any other HR line produces 0–3 such regions (Table 3).

Power and type I error simulations

Dropping any one of the 8 lines generally reduced the number of *P*-values lower than 0.05 and lower than other relevant significance thresholds, although the difference was sometimes negligible (Supplementary Table 2). Overall, these comparisons suggest that, as would be expected, the statistical power to detect differentiation between the HR and C lines is reduced when an HR line is excluded from the analyses.

The relative change in *P*-value appears to increase as the *P*-value decreases. For example, those loci whose 8-line analyses produced a *P*-value in the $0.05 < P < 0.01$ range, had an average increase in *P*-value by about 4.5% when a line was dropped, whereas loci whose 8-line analyses produce a *P*-value <1.00E−05 had an average increase of about 36.6% in *P*-values when a line was dropped.

Table 4 illustrates that type I error rates for $\alpha = 0.05$ are deflated in the 8-line analyses, as was noted previously (Hillis *et al.* 2020), and a similar deflation occurs for $\alpha = 0.01$. Dropping an HR line from the analyses increases the type I error rate for both $\alpha = 0.05$ and $\alpha = 0.01$ (Table 4). For $\alpha = 0.001$, the type I error rates were inflated for both the 8-line analyses (0.00319) and when dropping a line (range = 0.00276–0.00286), and even more so for $\alpha = 0.0001$ (range = 0.00060–0.00078). For some of the *P* thresholds (e.g. $P \leq 0.001$), the increase was quite large relative to the type I error rate for the 8-line analyses (Table 4).

To compare type I error rate to the *P*-values for the real data, total *P*-values below each of these thresholds (found in Table 1) were scaled to be out of 100,000 to match the simulation. When the estimated type I error rate (Table 4) is subtracted from the frequency of calculated total positives (scaled from Table 1), many signatures of selection remain, particularly when dropping HR3 (Table 4).

Haplotype and nonstatistical analyses excluding HR3 (mini-muscle)

Hillis *et al.* (2020) had identified 13 "consistent" regions (i.e. differentiated in SNP ANOVA, haplotype ANOVA, and FixedHR/PolyC) when performing analyses using all 8 lines. All 13 of those regions are listed in Table 5, including 1 region (chr16:40,742,298–41,357,426) that was inadvertently not identified as consistent by Hillis *et al.* (2020). When dropping HR3 from the analyses, 17 regions were identified as consistent, 7 of which were not identified in the 8-line analyses by any of the 3 analytical methods (Table 5). These 7 regions included genes associated with systems known to be different in the HR lines as compared with the C lines, including skeletal, heart, and neuronal development (see *Discussion*). For completeness, Table 5 also lists 15 additional regions identified by at least 2 of the 3 analytical approaches when dropping HR3.

Discussion

In the present study, we took advantage of the serendipitous discovery of a gene of major effect, named mini-muscle, which is part of the adaptive response to selection for high voluntary wheel running (see *Introduction*). Given its major effect on muscle mass and fiber-type composition, the observation that mini-muscle mice (and line HR3 in general) tend to run faster but for fewer minutes per day, as well as its pervasive pleiotropic effects on other behaviors, physiological traits, and organ sizes, we hypothesized that line HR3, which became fixed for the $Myh4^{Minimsc}$ allele, would show evidence of multiple solutions at the genomic level, as compared with the other 3 HR lines. Our results provide substantial support for this hypothesis, and encourage the application of similar analytical approaches to other replicated selection experiments.

SNP analyses excluding individual HR lines

Much of the increase in significant SNPs that we see when dropping HR3 can be attributed to 2 regions, chr3:46,438,071–52,624,971 and chr10:101,652,005–106,038,129 (Table 2), which had been identified by the 8-line analyses (Hillis *et al.* 2020). Because these regions were also detected with the 8-line analyses, it stands to reason that they responded to selection in all 4 HR lines. However, the wider areas implicated by the other 3 HR lines may correspond to stronger selection and hence a faster response to selection, as compared with HR3, thus not allowing sufficient time for recombination to break the haplotype in the other 3 HR lines (Smith and Haigh 1974; Kaplan *et al.* 1989; Kim and Stephan 2002).

Multiple solutions among the HR lines

A new genomic region that emerges as statistically significant only after dropping one of the HR lines (i.e. 4 C lines vs 3 HR lines)

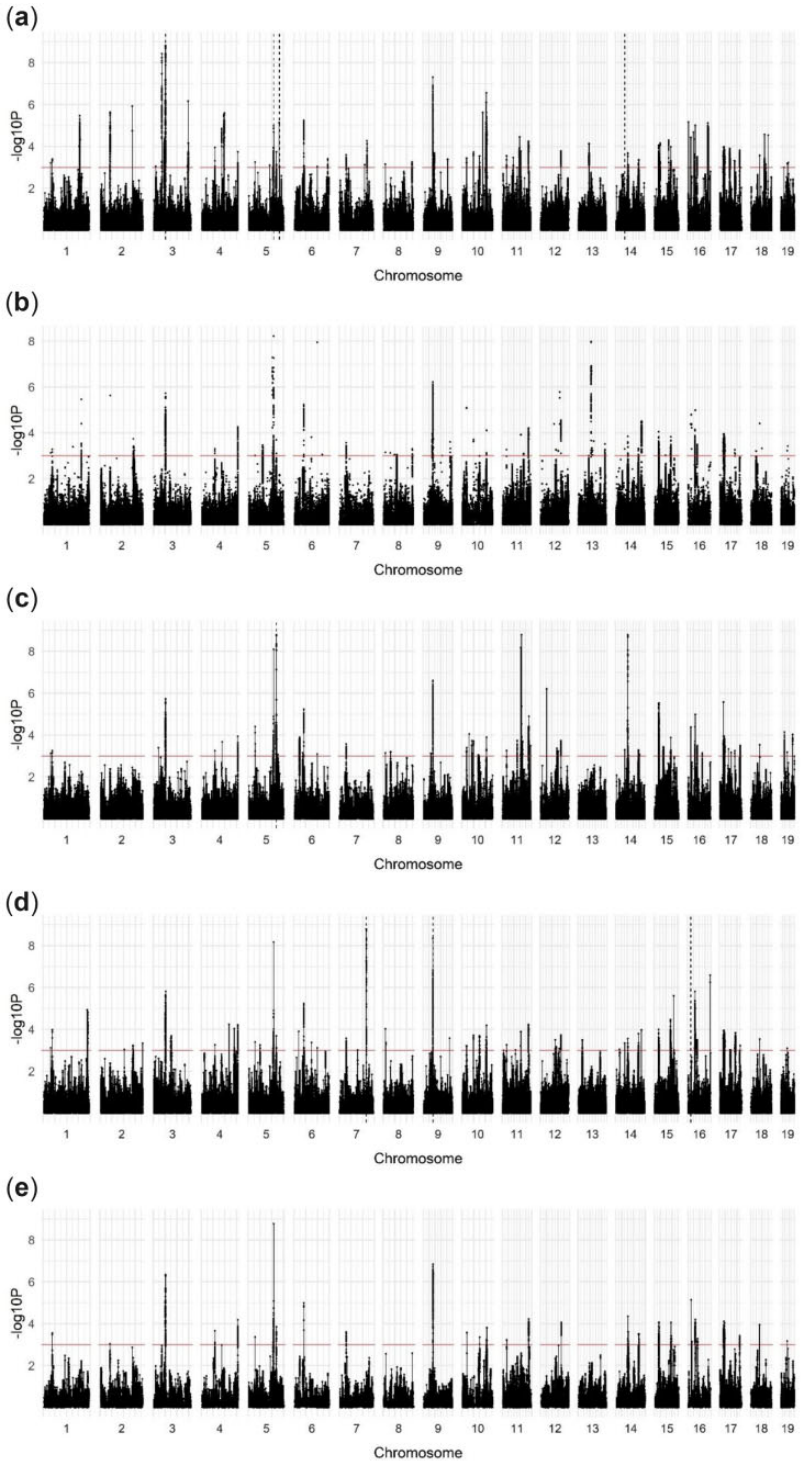

Fig. 3. a) Manhattan plot of the generation 61 individual mouse analyses, *excluding HR3* ($N = 5{,}931{,}993$). Vertical, dashed lines included represent loci fixed for opposite alleles (155 total loci) in the C and HR lines (e.g. $P = 2.43E{-}39$). b) Manhattan plot of the generation 61 individual mouse analyses, *excluding HR6* ($N = 5{,}932{,}148$). c) Manhattan plot of the generation 61 individual mouse analyses, *excluding HR7* ($N = 5{,}932{,}085$). Vertical, dashed lines included represent loci fixed for opposite alleles (63 total loci) in the C and HR lines (e.g. $P = 2.43E{-}39$). d) Manhattan plot of the generation 61 individual mouse analyses, *excluding HR8* ($N = 5{,}931{,}663$). Vertical, dashed lines included represent loci fixed for opposite alleles (485 total loci) in the C and HR lines (e.g. $P = 2.43E{-}39$). e) Manhattan plot of the generation 61 individual mouse analyses, *all 8 lines* ($N = 5{,}932{,}148$) (modified from Fig. 2 of Hillis *et al.* 2020).

implies: (1) the region is likely relevant to wheel-running behavior (though not as strongly supported as genomic regions identified with all 4 HR lines) and (2) the HR line that was dropped does not show the same response to selection as the other 3 HR lines (Fig. 5). Therefore, each of the 37 new regions listed in Table 2 may be thought of as relevant to voluntary wheel running in 3 of the 4 HR lines, thus providing evidence of "multiple solutions" at the genomic level.

Table 2. Genomic regions with different response to selection among the 4 replicate HR lines.

Chr	minPOS	maxPOS	Width (bp)	loci_noHR3	loci_noHR6	loci_noHR7	loci_noHR8	Significant SNPs in 8-line analysis
1	155,052,375	157,767,127	2,714,753	1,063	NA	NA	NA	
1	163,002,979	163,450,173	447,195	NA	18	NA	NA	
1	189,994,733	190,372,872	378,140	NA	NA	NA	277	
2	42,643,461	43,450,389	806,929	744	NA	NA	NA	
2	141,166,416	145,485,766	4,319,351	NA	30	NA	NA	
3	35,723,757	36,887,906	1,164,150	631	NA	NA	NA	
3	46,438,071	52,624,971	6,186,901	18,517	1,167	1,169	2,126	x
3	75,213,797	76,058,839	845,043	NA	NA	NA	324	
3	148,414,742	148,630,252	215,511	240	NA	NA	NA	
4	87,042,703	87,114,728	72,026	64	NA	NA	NA	
4	96,750,866	97,151,452	400,587	307	NA	NA	NA	
4	98,501,069	99,358,924	857,856	126	NA	NA	NA	
5	61,441,085	63,038,738	1,597,654	NA	58	NA	NA	
5	103,680,064	104,223,799	543,736	NA	38	NA	NA	
5	118,866,355	119,908,192	1,041,838	3	2	816	3	x
6	23,817,774	25,279,190	1,461,417	NA	NA	1,807	NA	
6	144,501,313	144,545,689	44,377	32	NA	NA	NA	
7	115,169,726	116,129,821	960,096	NA	NA	NA	1,414	
7	117,620,491	119,188,853	1,568,363	372	NA	NA	NA	
9	38,651,820	39,097,109	445,290	NA	260	NA	NA	
9	47,437,476	48,268,374	830,899	109	NA	NA	NA	
10	101,652,005	106,038,129	4,386,125	23,176	6	5	23	x
11	64,166,197	64,725,826	559,630	NA	NA	167	NA	
11	73,267,237	74,873,424	1,606,188	551	NA	NA	NA	
11	79,985,635	80,596,899	611,265	1	1	383	1	x
11	82,262,541	82,854,023	591,483	NA	NA	198	NA	
11	114,193,898	114,629,516	435,619	NA	NA	241	NA	x
12	71,753,929	72,115,917	361,989	NA	NA	24	NA	
13	48,002,204	48,725,034	722,831	222	NA	NA	NA	
13	57,093,620	57,687,257	593,638	NA	387	NA	NA	
13	115,026,145	115,159,660	133,516	NA	58	NA	NA	
14	51,204,847	54,600,493	3,395,647	11	10	195	6	x
14	111,028,017	111,103,065	75,049	NA	855	NA	NA	
15	38,434,855	40,095,009	1,660,155	NA	NA	1,115	NA	
15	61,388,625	61,598,676	210,052	175	NA	NA	NA	
15	69,108,096	71,591,605	2,483,510	40	496	36	2,414	x
16	31,440,034	33,322,583	1,882,550	144	772	103	1,999	x
16	87,078,841	88,667,433	1,588,593	5,078	NA	NA	NA	
17	44,043,299	44,712,180	668,882	3,480	NA	NA	NA	
17	67,989,641	68,777,695	788,055	NA	NA	NA	212	
17	87,221,754	87,282,285	60,532	168	NA	NA	NA	
17	88,660,339	88,885,305	224,967	NA	NA	632	1	
18	59,052,362	60,490,520	1,438,159	31	NA	NA	NA	
19	18,635,879	18,895,311	259,433	NA	NA	480	NA	
19	53,052,787	53,102,861	50,075	NA	NA	11	NA	

Genomic regions ($N = 45$; and counts of SNP loci within those regions) that became significant at the $P \leq 0.001$ level (1) for at least 10 SNPs after dropping the indicated line but no other HR line (e.g. chr1:155,052,375–157,767,127) and (2) for substantially more SNPs than any other dropped line (e.g. chr3:46,438,071–52,624,971).

Possible explanations for different responses to selection among the HR lines include:

- *Founder effects*. Different starting allele frequencies (i.e. founder effects, sensu Mayr 1942) could alter the response to selection (James 1970; Simões *et al.* 2008). For example, if certain biologically significant loci were already fixed or close to fixed in a given line, then that line would be forced to respond to selection via changes at other loci. The *Myh4*^{Minimsc} allele was present in the base population at a frequency of ∼7%, and so may have been absent in some lines (Garland *et al.* 2002), although the probability is low even for lines that were not observed to have the phenotype (∼0.07, based on calculation of posterior probabilities). Indeed, the phenotype was only ever observed in 1 C line and in 2 HR lines (e.g. see Fig. 1 in Garland *et al.* 2002).
- *Random genetic drift*. Following the founding of a small population, if the effective population size remains low, then drift may eliminate an allele despite some positive selection (or fix an allele despite some negative selection). This would be especially likely to occur for an allele that was present at a low frequency when the experiment began, such as *Myh4*^{Minimsc}. Thus, drift can exacerbate founder effects and constrain the genetic options available to a given population.
- *Epistatic effects*. If an allele with large epistatic effects (nonadditive interactions with alleles at other loci) increases in frequency within a given line, then substantial changes in allele frequencies at the epistatically related loci would be expected. For example, if allele A at the A locus positively affects wheel running, then alleles at other loci that increase wheel running only when allele A is present will be favored by selection only when allele A is present. If allele A were

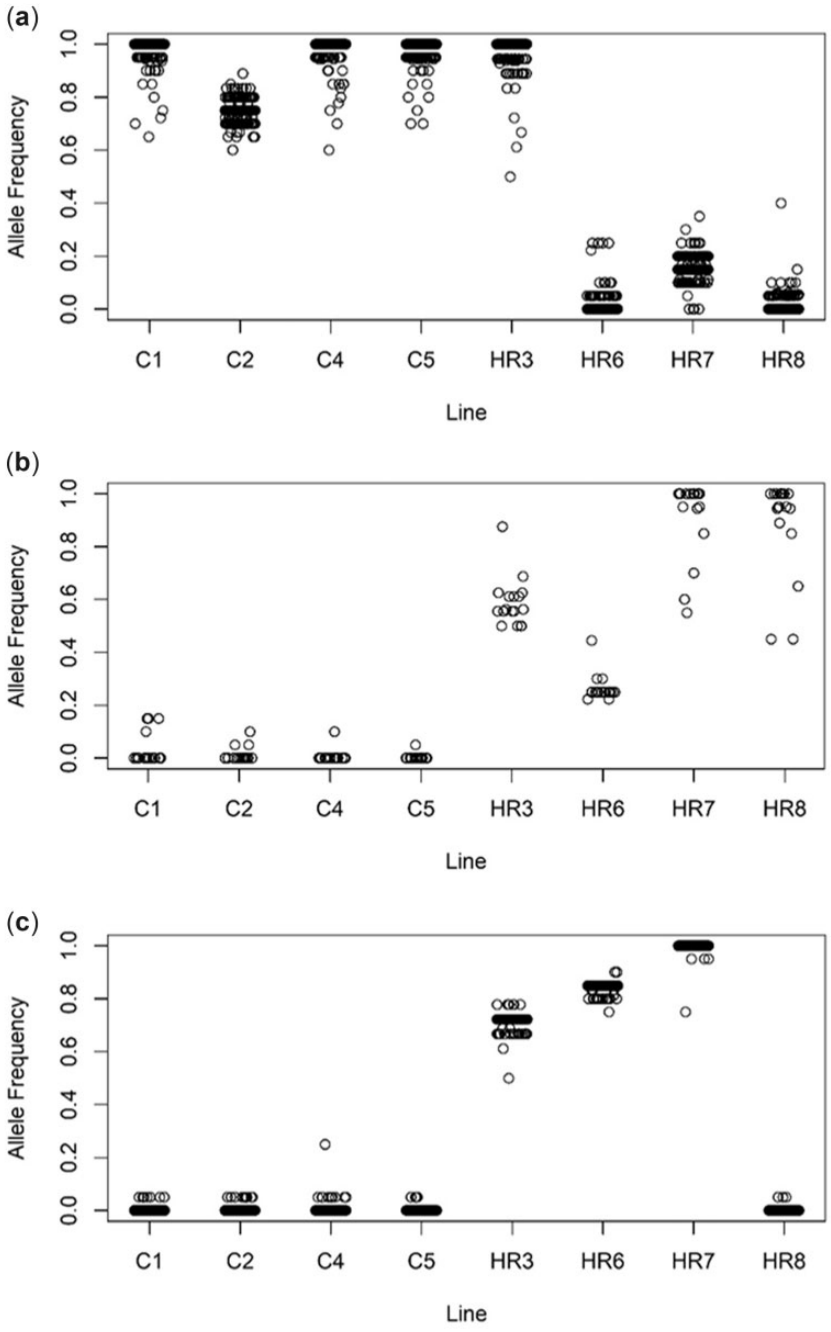

Fig. 4. Allele frequencies for individuals SNPs within example regions from Table 2. a) Chr1:155,052,375–157,767,127 illustrates a region that was detected as differentiated only after dropping line HR3, which has allele frequencies similar to the 4 control lines. b) Chr1:163,002,979–163,450,173 illustrates a region that was detected as differentiated only after dropping line HR6. c) Chr1: 189,994,733–190,372,872 illustrates a region that was detected as differentiated only after dropping line HR8.

present in only some populations under uniform selection, then the likelihood of multiple adaptive response would be increased.

- *Selection limits and constraints*. Suppose that mice are subject to a constraint on wheel running caused by joint pain: they stop running when the pain becomes intolerable. In this scenario, joint pain is sufficient to limit wheel running and it serves as a "weak link" in the physiological and neurobiological systems that are required for high levels of wheel running. Then suppose 10 alleles located at 10 independent biallelic loci, with entirely additive effects, are capable of increasing wheel running. Suppose further that only 5 such alleles are needed to achieve the amount of wheel running that causes intolerable joint pain. In such a scenario, fixation of the favorable allele at any 5 of the loci will coincide with a selection limit, but these alleles may be different among replicate lines.

Signatures of selection after dropping 1 line at a time

Although no loci were fixed for opposite alleles between the HR and C linetypes in Hillis *et al.* (2020) when considering all 8 lines, dropping any 1 HR or C line from the SNP analyses usually produced loci fixed for opposite alleles between the different linetypes (Table 3). These SNPs unsurprisingly tend to be clustered

Table 3. Fixation for opposite alleles after dropping a single line.

Dropped line	Chr	First SNP	Last SNP	Size (bp)	Total fixed	Identified in 8-line analyses	Nearby genes
C1	16	4,429,565	5,003,974	574,410	43	No	*Srl*, *Tfap4*, *Glis2*, *Pam16*, *Coro7*, *Vasn*, *Dnaja3*, *Nmral1*, *Hmox2*, *Cdip1*, *Ubald1*, *Mgrn1*, *Nudt16l1*, *Anks3*, *Sept12*
C2	5	105,140,011	105,818,176	678,166	5	Yes	*Gbp10*, *Gbp6*, *Gbp11*, *Lrrc8b*, *Lrrc8c*, *Lrrc8d*
C2	11	78,446,157	79,935,912	1,489,756	997	Yes	*Slc46a1*, *Sarm1*, *Vtn*, *Sebox*, *Tmem199*, *Poldip2*, *Tnfaip1*, *Ift20*, *Tmem97*, *Nlk*, *Fam58b*, *Lyrm9*, *Nos2*, *Lgals9*, *Ksr1*, *Wsb1*, *Nf1*, *Omg*, *Evi2*, *Evi2b*, *Evi2a*, *Rab11fip4*, *Utp*
C4	5	109,133,489	109,133,489	1	1	Yes	Vomeronasal region
HR3	3	51,572,179	51,606,225	34,047	149	Yes	*Setd7*
HR3	5	108,438,229	108,438,230	2	2	Yes	*Atp5k*
HR3	5	133,019,521	133,451,500	431,980	3	No	None
HR3	14	39,477,931	39,477,931	1	1	Yes	*Nrg3*
HR7	5	119,606,810	119,896,957	290,148	63	Yes	*Tbx3*, *Tbx5*
HR8	7	115,139,726	115,344,641	204,916	472	No	*Sox6*[a]
HR8	9	42,193,286	42,434,034	240,749	11	Yes	*Sc5d*, *Tecta*, *Tbcel*
HR8	16	14,215,165	14,473,951	258,787	2	Yes	*Myh11*, *Fopnl*, *Abcc1*

Frequency of fixation for opposite alleles in C and HR lines after dropping different lines. We do see unique regions when dropping other lines. Although more regions are identified when dropping HR3 than any other line, this is not by a large margin. However, as 8-line analyses (Hillis *et al.* 2020) identify 3 of these regions, at least some of these fixed regions are likely biologically significant across 3, if not all 4, HR lines.

[a] *Sox6* was mentioned by Hillis *et al.* (2020) as being a suggestive gene that the mixed-model analyses struggled to identify due to low within-line variance

Table 4. Type I error rates and estimated true positives.

P threshold	Expected	8-Line type I error[a]	8-Line true positives[b]	noHR3 type I error	noHR3 true positives	noHR6 type I error	noHR6 true positives	noHR7 type I error	noHR7 true positives	noHR8 type I error	noHR8 true positives
0.05	5,000	677	3,512.9	2,009	3,314.4	2,014	2,769.9	2,021	2,619.9	1,949	3,056.7
0.01	1,000	443	1,033.7	654	1,585.7	647	1,053.0	640	848.7	624	1,024.0
0.001	100	319	−35.1[c]	278	874.6	276	3.2	284	105.5	286	72.3
0.0001	10	67	28.5	64	203.2	60	45.1	61	51.8	78	62.1
0.00001	1	3	56.4	7	75	8	58.1	7	16.0	12	35.3
0.000001	0.1	0	16.6	1	7.1	0	4.6	1	1.8	1	35.7
1E-07	0.01	0	0	0	3.6	0	0.1	0	2.3	0	19.4
1E-08	0.001	0	0	0	3.5	0	0	0	2.1	0	10.2
1E-09	0.0001	0	0	0	2.6	0	0	0	1.1	0	8.2

[a] Type I error columns include the number of false positives for a given P threshold after sampling alleles from a given line and randomly shuffling lines between the linetypes (see *Materials and Methods*).

[b] Estimated true positive in the real data calculated by converting the number of loci found differentiated at a given P threshold into a ratio out of 100,000 loci and subtracting from these loci the number of estimated type I errors.

[c] Negative values (i.e. fewer total positives than predicted false positives).

into specific regions (separated by at least 1 mbp), some of which have been detected either in the present study or by Hillis *et al.* (2020). Most of the regions listed in Table 3 were also identified by the 8-line SNP analyses, which may suggest that the dropped line is not drastically different from the others within its linetype. However, 3 new regions emerge.

The first new region is seen after dropping C1 (chr16:4,429,565–5,003,974) and contains various genes whose knockouts have been associated with heart morphology (Yoshida *et al.* 2005; Hayashi *et al.* 2006; Cota *et al.* 2006; Dickinson *et al.* 2016). Since all of the HR lines fixed for the same allele, this would not be an example of different responses to selection, but an example of variation among control lines disrupting our ability to detect selection signatures in the HR lines.

The second and third regions were identified by dropping HR3 and HR8, respectively (Table 3). These regions might implicate different responses to selection among the HR lines. One of these regions contains the *Sox6* gene described above and by Hillis *et al.* (2020) for its effect in regulating muscle fiber type, hematopoiesis, bone growth and heart function (Smits *et al.* 2001; van Rooij *et al.* 2009). While 3 of the HR lines were fixed for the reference allele, HR8 became fixed for the alternate allele. The region identified when dropping HR3 (chr5:133,019,521–133,451,500) does not contain any annotated sequences. Some possible explanations include: a relevant gene being present but simply not yet annotated; this region serving an unknown regulatory role for other genes; or this region having undergone this fixation pattern purely by drift (i.e. it does not influence running behavior). One potential gene regulated by this region would be *Auts2* (approximately 480 kbp downstream of the region), which has been implicated in neurodevelopment (Oksenberg and Ahituv 2013). *Auts2* is also thought to be associated with the *Runx1* pathway: an

Table 5. Regions identified by analyses excluding HR3.

Chromosome	Low BP	High BP	SNPs	Haplotype	FHRPC	Identified in 8-line
1	156,378,655	156,798,297	x	x		
2	42,722,345	43,214,647	x	x	x	LM
3	35,754,848	36,551,567	x		x	LM
3	46,440,016	51,798,758	x	x	x	LM, Hap
3	148,414,742	148,638,738	x	x		
4	57,735,866	58,314,736	x		x	LM, FHRPC
4[a]	96,840,333	98,580,378	x	x	x	
4	155,480,343	155,654,426	x	x		LM
5	104,331,989	106,773,650	x		x	FHRPC
5	108,000,623	109,929,498	x	x	x	ALL
6	37,440,411	39,293,531		x	x	FHRPC
6	41,584,862	42,841,617	x	x	x	ALL
6	43,004,011	43,431,434	x	x		Hap
6	143,921,896	144,545,685	x		x	
7	29,603,841	30,672,131		x	x	ALL
7[a]	119,020,988	119,188,853	x	x	x	
8	121,821,514	126,241,684		x	x	
9	41,240,184	42,314,823	x	x	x	ALL
9[a]	48,043,076	50,172,437	x	x	x	
10	75,061,742	75,698,409		x	x	LM
10	101,672,276	105,807,985	x	x	x	LM, Hap
11[a]	74,340,021	74,838,159	x	x	x	
11	78,166,291	80,218,837		x	x	ALL
11	112,075,283	114,258,210	x	x	x	ALL
12	88,904,818	92,138,006	x		x	LM, FHRPC
13[a]	48,002,204	48,380,363	x	x	x	
13	56,280,106	56,618,730	x	x		
14	52,072,148	53,779,979		x		ALL
14	97,210,445	99,269,397		x	x	ALL
15	18,196,530	20,632,943	x	x	x	ALL
15	61,361,016	61,653,459		x	x	FHRPC
15	70,665,271	71,547,239	x	x	x	ALL
16	31,686,517	31,779,261		x	x	ALL
16	40,742,298	41,357,426		x		ALL[b]
16	87,080,074	88,667,433	x	x		
17	17,661,786	23,599,776	x	x	x	ALL
17[a]	44,063,088	44,685,529	x	x	x	
18[a]	56,147,892	60,494,440	x	x	x	

This table contains regions identified by at least 2 of the 3 analytical methods used. The column on the far right indicates which of these regions were identified by which method for the Hillis *et al.* (2020) 8-line analyses (LM = local maximum, implicating most significant SNPs of a suggestive region).

[a] These were regions implicated by all 3 analytical methods (mixed-model analyses of individual SNPs and haplotypes, as well as FixedHR/PolyC) after dropping HR3, but not by any method with all 8 lines.

[b] This region was accidentally not implicated as a consistent region by Hillis *et al.* (2020) but meets the criteria.

intriguing association when *Runx2* is found in a separate region identified when dropping HR3 (Table 6).

Variation in olfactory response to selection

Olfaction is known to play an important role in some motivated behaviors. Our previous 8-line analyses showed that several vomeronasal genes have responded consistently to selection (Hillis *et al.* 2020; Nguyen *et al.* 2020). The vomeronasal organ is part of the overall olfactory system and functions primarily to detect nonvolatile organic compounds. Table 2 includes regions with genes that have an olfactory, but not vomeronasal, function. These regions were identified when dropping lines HR3, HR6, or HR7, with a different region appearing important after dropping each of the lines (HR3 = chr11:73,267,237–74,873,424; HR6 = chr9:38,651,820–39,097,109; HR7 = chr14:51,204,847–54,600,493). We interpret this as evidence for multiple solutions occurring in a given physiological system, at 2 different levels. In other words, olfaction seems to be important in the evolution of the HR phenotype (Dewan *et al.* 2019), and this may occur by either vomeronasal or nonvomeronasal pathways (or both). Although multiple vomeronasal genes in multiple regions on multiple chromosomes were identified in the previous 8-line analyses, here we did not find evidence of differences among the HR lines for these genes. However, we did find that multiple nonvomeronasal olfactory genes seem to have been important in the response to selection, and with different genes being important in different HR lines.

Power and type I error simulations

An increase in type I error rate when dealing with low sample size is not a new observation for some types of genetic data (Baldi and Long 2001). In any case, the inflated type I error rate may draw into question some of our "significant" results for the 7-line analyses in Table 2. To gauge the magnitude of this problem, we subtracted the expected false positives (Table 4) from our total positives (scaled from Table 1). As shown in Table 4, dropping HR3 produces many more *P*-values of 0.001 or lower than expected under the null hypothesis, and more than when dropping any of the other lines. This observation increases our confidence that the genomic response to selection by line HR3 truly differs from that of the other 3 HR lines.

Chromosomal regions identified when excluding HR3

Despite the many expected similarities between the 8-line analyses and analyses dropping HR3, the present study identifies 7 genomic regions implicated by all 3 tests (SNPs, haplotype, and

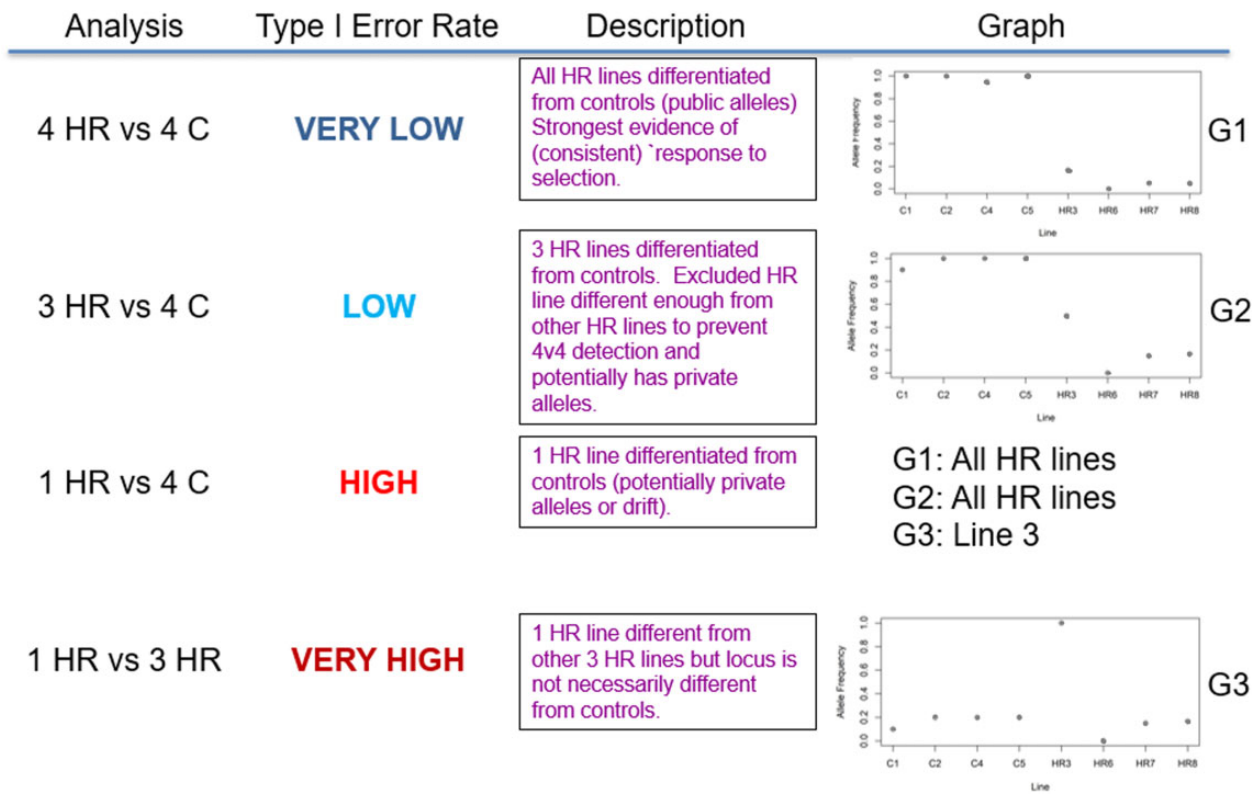

Fig. 5. Illustration of different analysis strategies for detection of "private" alleles. The 4 possibilities shown include all 4 HR lines vs all 4 C lines [as was done by Hillis *et al.* (2020)], 3 HR vs 4 C (as is done in the present study), as well as 1 HR vs 4 C and 1 HR vs 3 HR (both of which are expected to have increased type I error rate as compared to the previous 2 analyses).

Table 6. Genes identified by all analyses, after excluding HR3

Chromosome	Low BP	High BP	Genes within region
4	96,840,333	98,580,378	*Nfia*, *Tm2d1*, *Patj*
7	119,020,988	119,188,853	*Gpr139*, *Gprc5b* (downstream)
9	48,043,076	50,172,437	*Nxpe2*, *Nxpe4*, *Nxpe1-ps*, *Rexo2*, *Rbm7*, *Nnmt*, *Zbtb16*, *Htr3a*, *Htr3b*, *Usp28*, *Cldn25*, *Zw10*, *Tmprss5*, *Drd2*, *Ankk1*, *Ttc12*, *Ncam1*
11	74,340,021	74,838,159	*Olfr411*, *Olfr412*, *Rap1gap2*, *Ccdc92b*, *C1uh*, *Pafah1b1*, *Mettl16*, *Mnt*
13	48,002,204	48,380,363	*Id4*
17	44,063,088	44,685,529	*Enpp4*, *Enpp5*, *Clic5*, *Runx2*, *Runx2os3*
18	56,147,892	60,494,440	*Gramd3*, *Aldh7a1*, *Mir1258*, *Phax*, *Tex43*, *Lmnb1*, *Marchf3*, *C330018D20Rik*, *Megf10*, *Prrc1*, *Ctxn3*, *Ccdc192*, *Slc12a2*, *Fbn2*, *Slc27a6*, *Isoc1*, *Adamts19*, *Minar2*, *Chsy3*, *Mir6355*, *Iigp1*, *Smim3*

These regions were implicated by all 3 analyses (individual SNP, haplotype, and fixation in HR/polymorphic in C) only after dropping line 3 from the analyses. Brief descriptions of these genes are presented in Supplementary Table 3.

FixedHR/PolyC) that were not identified by any tests when analyzed with all 8 lines (Table 6). These regions contained nearly 61 genes; however, 3 in particular caught our attention, *Ncam1*, *Drd2*, and *Minar2*.

Ncam1 codes for a cell adhesion protein whose knockouts are associated with altered hippocampus, cerebellum, and olfactory bulb development (Tomasiewicz *et al.* 1993; Holst *et al.* 1998), as well as shortened circadian period (Shen *et al.* 2001). Differential circadian rhythms have been found by Koteja *et al.* (2003), who showed that HR mice have a shorter free-running period (tau) under both constant dark and constant light. In addition, human GWAS have implicated *Ncam1* in playing a role in heel bone mineral density (Kim 2018; Morris *et al.* 2019) and addictive behaviors, specifically smoking (Kichaev *et al.* 2019; Karlsson Linnér *et al.* 2019; Liu *et al.* 2019). Several bone differences between HR and C lines have been documented (particularly in limb bone size and shape). This includes a number of differences between mini- and normal-muscled mice (Kelly *et al.* 2006; Middleton *et al.* 2008, 2010; Wallace *et al.* 2010, 2012; Castro, Rabitoy, *et al.* 2021). Moreover, HR mice show withdrawal symptoms when wheel access is removed (Kolb *et al.* 2013).

Hippocampal function in the HR lines has been explored through indirect methods (Rhodes *et al.* 2003; Johnson *et al.* 2003; Bronikowski *et al.* 2004). For example, Bronikowski *et al.* (2004) found some genes related to transcription and translation that had increased expression in the hippocampus in HR vs C lines, whereas some associated with neuronal signaling and immune function had decreased expression in HR mice. The HR lines also had increased brain-derived neurotrophic factor in the hippocampus after having access to wheels for 7 days (Johnson *et al.* 2003). As for response to wheel running, the C lines showed a positive correlation between wheel running and neurogenesis in the

dentate gyrus of the hippocampus, whereas the HR mice did not, with all HR mice having a high level of neurogenesis (Rhodes *et al.* 2003). Moreover, wheel access improved learning in the Morris water maze for C mice but not for HR mice. With body mass as a covariate, Schmill (2021) found that the total volume of the hippocampus is larger in HR than in C mice, both for animals housed with wheels for 10 weeks and those housed without wheels.

Drd2 is a dopamine receptor that has been associated with a wide variety of disorders, addictions, and compulsive behaviors (Blum *et al.* 1995; Hung Choy Wong *et al.* 2000; Noble 2003; Bronikowski *et al.* 2004; Munafò *et al.* 2004; Foll *et al.* 2009). *Drd2* has also been tied to wheel running in mice based on differential expression in high- and low-running lines (C57L/J and C3H/HeJ, respectively) (Dawes *et al.* 2014). In addition, *Drd2* knockouts have altered wheel-running behavior (Roberts *et al.* 2017).

When line HR8 was compared to stock ICR mice, significant differences in expression of *Drd1a* and *Drd2* receptors (downregulated in HR8) were found in the dorsal striatum (Mathes *et al.* 2010). In addition, HR and C mice in the wheel-running response to cocaine (Rhodes *et al.* 2001). Though surprisingly, Drd2 receptor antagonist does not appear to cause a different response in the HR lines than control (Rhodes and Garland 2003); however, this study did not separate HR3 or other mini-muscle mice in the analyses.

Minar2 is a NOTCH2-associated receptor whose knockouts have been associated with altered bone structure (Dickinson *et al.* 2016), impaired coordination and gait (Ho *et al.* 2020), decreased body mass and length (Dickinson *et al.* 2016), and loss of dopaminergic neurons (Ho *et al.* 2020). Mice from the HR lines are generally smaller than the C lines, and differ in bone properties (see above), dopaminergic function (see above), and some aspects of gait during treadmill running (e.g. see Swallow *et al.* 1999; Rhodes *et al.* 2001; Girard *et al.* 2007; Garland *et al.* 2011; Claghorn *et al.* 2017).

Limitations of the present study and concluding remarks

Given the complexity of voluntary wheel-running behavior, identical evolutionary pathways in the 4 replicate HR lines would be highly unlikely. The fixation of the *Myh4*^{Minimsc} allele in just line HR3 is a clear example an alternative "solution" to selection that favors high-activity levels. Here, we show that the other 3 HR lines also show evidence of somewhat unique responses to selection (Table 2). However, HR3 seemingly stands out from the rest of the HR lines. As explained in the *Introduction*, a plausible explanation for this is that the *Myh4*^{Minimsc} allele has such large direct and pleiotropic effects (particularly in systems relevant for wheel running) that much of the rest of the genome has had to evolve differently in response. We would also note that HR3 has higher heterozygosity than any other line (including C lines) (Hillis *et al.* 2020).

Although the mixed-model method using multiple models and mivque variance estimation seems to be a relatively powerful method of analyzing these data (Xu and Garland 2017; Hillis *et al.* 2020), dropping lines negatively impacts power and inflates type I error rates (Table 4). More powerful analytical methods may need to be developed to better identify signatures of selection. One possibility may be to incorporate inferences similar to those described by Baldi and Long (2001) to offset the low sample size. Genomic data from generations closer to when the selection limit was reached may also reduce the type I errors produced by drift, allowing for better detection of true positive results. In addition, the present study does not perform any functional analyses of the suggested genes to establish a causal relationship between the gene and wheel-running behavior or other phenotypes suggested by KO studies (see above). Further studies are needed to establish these functional connections within the HR mice or at least to demonstrate that KO mice for these genes differ from wildtype in wheel running when measured under conditions similar to those used in the HR selection experiment.

A noteworthy question that the present study does not address is: why did HR3 become fixed for the *Myh4*^{Minimsc} allele while HR6 has remained heterozygous despite continued selection? Possible explanations for this include heterozygote advantage or epistatic interactions with loci unique to HR6. These ideas could be tested by genomic analyses of current or historical (e.g. see Kelly *et al.* 2013) samples and associating genotype with wheel-running and other relevant phenotypes. In addition, the differences between mini-muscle and normal-muscled individuals for some muscle properties are greater in HR3 than in HR6 (Guderley *et al.* 2006), suggesting that selection favoring this phenotype may have been stronger in HR3.

Despite the limitations discussed above, the present study was able to identify 7 new genomic regions of differentiation in 3 of the lines bred for high voluntary wheel running, as compared with the 4 nonselected control lines. These regions contain genes that are both intuitive for voluntary-exercise behavior and correlate to known phenotypic differences between the HR and control lines. These regions also highlight some of the genomic differences between HR3 and the other HR lines, enabling us to begin to address multiple solutions in response to uniform selection.

Selection experiments involving replicate lines have demonstrated both similar and varying responses to selection (Garland and Rose 2009). Supporting the latter possibility, Ernst Mayr (1961, p. 1505) once wrote that "Breeders and students of natural selection have discovered again and again that independent parallel lines exposed to the same selection pressures will respond at different rates and with different effects, none of them predictable." On the other hand, replicates involving asexually reproducing bacteria typically tend to implicate the same genes or pathways, although not necessarily the same SNPs (Long *et al.* 2015). For example, Tenaillon *et al.* (2012) demonstrated that evolving 115 populations of *Escherichia coli* for survival at increased temperatures resulted in replicates consistently implicating a limited number of genes. However, despite regular patterns in mutated genes, favorable mutations in the *rho* gene deterred the mutations that would normally have been favorable in the *rpoBC* gene, implicating a potential alternative solution.

Evolution of replicate *Drosophila* lines commonly results in similar responses to selection (Long *et al.* 2015). An example of this would be selection on *Drosophila melanogaster* wing venation (Cohan 1984a). Conversely, Cohan and Hoffmann (1986) identified different responses to selection for alcohol tolerance in *D. melanogaster*. The alcohol tolerance experiment began with different populations of flies taken from different geographic areas and so differences in starting genetic background is a potential explanation for these different responses. However, even with different populations, Cohan and Hoffmann (1986) concluded that genetic drift was no less a driving force in differential response to selection than genetic background. Furthermore, Cohan *et al.* (1989) later showed that models assuming large epistatic interactions were less consistent with response to selection than models assuming pure additivity. Epistatic interactions have commonly been found to influence outbred populations, potentially because recombination allows beneficial mutations to

be found in a variety of alleles and genetic backgrounds (Long *et al.* 2015).

Given the large size of the commercial breeding colony from which our base population of 224 mice derived and with 2 generations of random mating in our lab before being divided into 8 closed lines (Swallow *et al.* 1998; Carter *et al.* 1999; Girard *et al.* 2002), the replicate HR and C lines should have started with largely homogenous genetic backgrounds. However, even if most lines had the mini-muscle allele, only 3 of 8 ever had mice with the mini-muscle phenotype due to the low allele frequency and recessive nature (Garland *et al.* 2002). Potentially, only those HR lines that happened to express the mini-muscle phenotype before it was lost to drift had the opportunity for the mini-muscle allele to be favored by selection, thus altering the genetic background through various pleiotropic and epistatic effects. However, HR3 is not the only line to differ in its response to selection. As shown here, each of the HR lines reveal new potential selection signatures when dropped from the analyses (Tables 2 and 3), implicating variation in their response to the selection criterion. We encourage workers to focus more on the utility of replicate lines for the study of multiple solutions at all levels of biological organization (see also Garland 2003; Garland and Rose 2009).

Data availability

Original data were made available by Hillis *et al.* (2020) and can be found at https://doi.org/10.25386/genetics.12436649. Supplementary File 1 contains brief descriptions of supplemental tables. Supplementary Table 1 contains all regions with at least 10 SNPs with $P \leq 0.001$ for any analyses where an HR line was dropped. Supplementary Table 2 contains results of power analyses performed by sampling from a locus in a differentiated region and sampling alleles from each line for that locus with mixed-model analyses used to produce a test statistic for each of 100,000 repetitions of this sampling method (see *Materials and Methods*). Supplementary Table 3 contains a list of annotated genes in the new genomic regions identified only after dropping line HR3, with content from Entrez database related to current understanding of the genes' function.

Supplemental material is available at GENETICS online.

Acknowledgments

The authors thank Z. Jia and S. Xu for comments on the article.

Funding

This study was supported by the U.S. National Science foundation, most recently NSF grant IOS-2038528 to TG.

Author contributions

Conceptualization, DAH and TG; investigation, DAH and TG; software, DAH; formal analysis, DAH and TG; writing—original draft, DAH and TG; writing—review and editing, DAH and TG.

Conflicts of interest

None declared.

Literature cited

Agrawal A. F, Brodie ED, Rieseberg LH. Possible consequences of genes of major effect: transient changes in the G-matrix. Genetica. 2001;112–113:33–43. https://doi.org/10.1007/978-94-010–0585-2_3

Aluoch JR. Higher resistance to *Plasmodium falciparum* infection in patients with homozygous sickle cell disease in western Kenya. Trop Med Int Health. 1997;2(6):568–571. https://doi.org/10.1046/j.1365–3156.1997.d01-322.x

Baldi P, Long AD. A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. Bioinformatics. 2001;17(6):509–519. https://doi.org/10.1093/bioinformatics/17.6.509

Bennett AF. Experimental evolution and the Krogh Principle: generating biological novelty for functional and genetic analyses. Physiol Biochem Zool. 2003;76(1):1–11. https://doi.org/10.1086/374275

Bilodeau GM, Guderley H, Joanisse DR, Garland T, Jr. Reduction of type IIb myosin and IIB fibers in tibialis anterior muscle of mini-muscle mice from high-activity lines. J Exp Zool A Ecol Genet Physiol. 2009;311(3):189–198. https://doi.org/10.1002/jez.518

Blum K, Sheridan PJ, Wood RC, Braverman ER, Chen TJ, Comings DE. Dopamine D2 receptor gene variants: association and linkage studies in impulsive-addictive-compulsive behaviour. Pharmacogenetics. 1995;5(3):121–141. https://doi.org/10.1097/00008571–199506000-00001

Bock WJ. Preadaptation and multiple evolutionary pathways. Evolution. 1959;13(2):194–211. https://doi.org/10.1111/j.1558–5646.1959.tb03005.x

Bronikowski AM, Rhodes JS, Garland T, Prolla TA, Awad TA, Gammie SC. The evolution of gene expression in mouse hippocampus in response to selective breeding for increased locomotor activity. Evolution. 2004;58(9):2079–2086. https://doi.org/10.1111/j.0014–3820.2004.tb00491.x

Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE; Mouse Genome Database Group. Mouse Genome Database (MGD) 2019. Nucleic Acids Res. 2019;47(D1):D801–D806. https://doi.org/10.1093/nar/gky1056

Cadney MD, Hiramatsu L, Thompson Z, Zhao M, Kay JC, Singleton JM, Albuquerque RLd, Schmill MP, Saltzman W, Garland T, *et al.* Effects of early-life exposure to Western diet and voluntary exercise on adult activity levels, exercise physiology, and associated traits in selectively bred High Runner mice. Physiol Behav. 2021;234:113389. https://doi.org/10.1016/j.physbeh.2021.113389

Careau V, Wolak ME, Carter PA, Garland T, Jr. Limits to behavioral evolution: the quantitative genetics of a complex trait under directional selection. Evolution. 2013;67(11):3102–3119. https://doi.org/10.1111/evo.12200

Carter PA, Garland T, Jr, Dohm MR, Hayes JP. Genetic variation and correlations between genotype and locomotor physiology in outbred laboratory house mice (*Mus domesticus*). Comp Biochem Physiol A Mol Integr Physiol. 1999;123(2):155–162. https://doi.org/10.1016/S1095-6433(99)00044-6

Castro AA, Karakostis FA, Copes LE, McClendon HE, Trivedi AP, Schwartz NE, Garland T. Effects of selective breeding for voluntary exercise, chronic exercise, and their interaction on muscle attachment site morphology in house mice. J Anat. 2021;240:279–295. https://doi.org/10.1111/joa.13547.

Castro AA, Rabitoy H, Claghorn GC, Garland T. Rapid and longer-term effects of selective breeding for voluntary exercise behavior on skeletal morphology in house mice. J Anat. 2021;238(3):720–742. https://doi.org/10.1111/joa.13341

Claghorn GC, Thompson Z, Kay JC, Ordonez G, Hampton TG, Garland T. Selective breeding and short-term access to a running wheel alter stride characteristics in house mice. Physiol Biochem Zool. 2017;90(5):533–545. https://doi.org/10.1086/692909

Cohan FM. Genetic divergence under uniform selection. I. similarity among populations of *Drosophila melanogaster* in their responses to artificial selection for modifiers of *cid*. Evolution. 1984a;38(1): 55–71. https://doi.org/10.1111/j.1558–5646.1984.tb00260.x

Cohan FM. Can uniform selection retard random genetic divergence between isolated conspecific populations? Evolution. 1984b;38(3): 495–504. https://doi.org/10.1111/j.1558–5646.1984.tb00315.x

Cohan FM, Hoffmann AA. Genetic divergence under uniform selection. 11. Different responses to selection for knockdown resistance to ethanol among *Drosophila melanogaster* populations and their replicate lines. Genetics. 1986;114(1):145–164. https://doi.org/10.1093/genetics/114.1.145

Cohan FM, Hoffmann AA, Gayley TW. A test of the role of epistasis in divergence under uniform selection. Evolution. 1989;43(4): 766–774. https://doi.org/10.1111/j.1558–5646.1989.tb05175.x

Cota CD, Bagher P, Pelc P, Smith CO, Bodner CR, Gunn TM. Mice with mutations in *Mahogunin ring finger-1* (*Mgrn1*) exhibit abnormal patterning of the left-right axis. Dev Dyn. 2006;235(12): 3438–3447. https://doi.org/10.1002/dvdy.20992

Dawes M, Moore-Harrison T, Hamilton AT, Ceaser T, Kochan KJ, Riggs PK, Lightfoot JT. Differential gene expression in high- and low-active inbred mice. Biomed Res Int. 2014;2014: 361048–361049. https://doi.org/10.1155/2014/361048

Dewan I, Garland T, Jr, Hiramatsu L, Careau V. I smell a mouse: indirect genetic effects on voluntary wheel-running distance, duration and speed. Behav Genet. 2019;49(1):49–59. https://doi.org/10.1007/s10519-018–9930-2

Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, Meehan TF, Weninger WJ, Westerberg H, Adissu H, *et al.*; The International Mouse Phenotyping Consortium. High-throughput discovery of novel developmental phenotypes. Nature. 2016; 537(7621):508–514. https://doi.org/10.1038/nature19356

Didion JP, Morgan AP, Yadgary L, Bell TA, McMullan RC, Ortiz de Solorzano L, Britton-Davidian J, Bult CJ, Campbell KJ, Castiglia R, *et al. R2d2* drives selfish sweeps in the house mouse. Mol Biol Evol. 2016;33(6):1381–1395. https://doi.org/10.1093/molbev/msw036

Dlugosz EM, Chappell MA, McGillivray DG, Syme DA, Garland T, Jr. Locomotor trade-offs in mice selectively bred for high voluntary wheel running. J Exp Biol. 2009;212(Pt 16):2612–2618. https://doi.org/10.1242/jeb.029058

Foll BL, Gallo A, Strat YL, Lu L, Gorwood P. Genetics of dopamine receptors and drug addiction: a comprehensive review. Behav Pharmacol. 2009;20(1):1–17. https://doi.org/10.1097/FBP.0b013e3283242f05

Gammie SC, Hasen NS, Rhodes JS, Girard I, Garland T. Predatory aggression, but not maternal or intermale aggression, is associated with high voluntary wheel-running behavior in mice. Horm Behav. 2003;44(3):209–221. https://doi.org/10.1016/S0018-506X(03)00140-5

Garland T. Selection experiments: an under-utilized tool in biomechanics and organismal biology. In: Bels VL, Gasc JP, Casinos A, editors. Vertebrate Biomechanics and Evolution. Oxford, UK: BIOS Scientific Publishers; 2003. p. 23–56.

Garland TJr, Rose, MR (Eds.). Experimental Evolution: Concepts, Methods, and Applications of Selection Experiments. Berkeley: University of California Press; 2009.

Garland T, Kelly SA, Malisch JL, Kolb EM, Hannon RM, Keeney BK, Van Cleave SL, Middleton KM. How to run far: multiple solutions and sex-specific responses to selective breeding for high voluntary activity levels. Proc Biol Sci. 2011;278(1705):574–581. https://doi.org/10.1098/rspb.2010.1584

Garland T, Morgan MT, Swallow JG, Rhodes JS, Girard I, Belter JG, Carter PA. Evolution of a small-muscle polymorphism in lines of house mice selected for high activity levels. Evolution. 2002;56(6): 1267–1275. https://doi.org/10.1111/j.0014–3820.2002.tb01437.x

Gerrits L, Overheul GJ, Derks RC, Wieringa B, Hendriks WJAJ, Wansink DG. Gene duplication and conversion events shaped three homologous, differentially expressed myosin regulatory light chain (MLC2) genes. Eur J Cell Biol. 2012;91(8):629–639. https://doi.org/10.1016/j.ejcb.2012.02.001

Girard I, Rezende EL, Garland T, Jr. Leptin levels and body composition of mice selectively bred for high voluntary locomotor activity. Physiol Biochem Zool. 2007;80(6):568–579. https://doi.org/10.1086/521086

Girard I, Swallow JG, Carter PA, Koteja P, Rhodes JS, Garland T. Maternal-care behavior and life-history traits in house mice (*Mus domesticus*) artificially selected for high voluntary wheel-running activity. Behav Processes. 2002;57(1):37–50. https://doi.org/10.1016/S0376-6357(01)00206-6

Griffiths AJF, Wessler SR, Carroll SB, Doebley J. An Introduction to Genetic Analysis. New York, NY: W. H. Freeman & Company; 2015. p. 10010.

Guderley H, Houle-Leroy P, Diffee GM, Camp DM, Garland T. Morphometry, ultrastructure, myosin isoforms, and metabolic capacities of the "mini muscles" favoured by selection for high activity in house mice. Comp Biochem Physiol B Biochem Mol Biol. 2006;144(3):271–282. https://doi.org/10.1016/j.cbpb.2006.02.009

Guderley H, Joanisse DR, Mokas S, Bilodeau GM, Garland T. Altered fibre types in gastrocnemius muscle of high wheel-running selected mice with mini-muscle phenotypes. Comp Biochem Physiol B Biochem Mol Biol. 2008;149(3):490–500. https://doi.org/10.1016/j.cbpb.2007.11.012

Hannon RM, Kelly SA, Middleton KM, Kolb EM, Pomp D, Garland T. Phenotypic effects of the "Mini-Muscle" allele in a large HR x C57BL/6J mouse backcross. J Hered. 2008;99(4):349–354. https://doi.org/10.1093/jhered/esn011

Hayashi M, Imanaka-Yoshida K, Yoshida T, Wood M, Fearns C, Tatake RJ, Lee J-D. A crucial role of mitochondrial Hsp40 in preventing dilated cardiomyopathy. Nat Med. 2006;12(1):128–132. https://doi.org/10.1038/nm1327

Hillis DA, Yadgary L, Weinstock GM, Pardo-Manuel de Villena F, Pomp D, Fowler AS, Xu S, Chan F, Garland T. Genetic basis of aerobically supported voluntary exercise: results from a selection experiment with house mice. Genetics. 2020;216(3):781–804. https://doi.org/10.1534/genetics.120.303668

Hiramatsu L, Kay JC, Thompson Z, Singleton JM, Claghorn GC, Albuquerque RL, Ho B, Ho B, Sanchez G, Garland T, *et al.* Maternal exposure to Western diet affects adult body composition and voluntary wheel running in a genotype-specific manner in mice. Physiol Behav. 2017;179:235–245. https://doi.org/10.1016/j.physbeh.2017.06.008

Ho RX-Y, Amraei R, De La Cena KOC, Sutherland EG, Mortazavi F, Stein T, Chitalia V, Rahimi N. Loss of MINAR2 impairs motor function and causes Parkinson's disease-like symptoms in mice. Brain Commun. 2020;2(1):fcaa047. https://doi.org/10.1093/braincomms/fcaa047

Holst BD, Vanderklish PW, Krushel LA, Zhou W, Langdon RB, McWhirter JR, Edelman GM, Crossin KL. Allosteric modulation of AMPA-type glutamate receptors increases activity of the promoter for the neural cell adhesion molecule, N-CAM. Proc Natl Acad Sci U S A. 1998;95(5):2597–2602. https://doi.org/10.1073/pnas.95.5.2597

Houle-Leroy P, Guderley H, Swallow JG, Garland T. Artificial selection for high activity favors mighty mini-muscles in house mice. Am J Physiol Regul Integr Comp Physiol. 2003;284(2):R433–R443. https://doi.org/10.1152/ajpregu.00179.2002

James JW. The founder effect and response to artificial selection. Genet Res. 1970;16(3):241–250. https://doi.org/10.1017/S0016672300002500

Johnson RA, Rhodes JS, Jeffrey SL, Garland T, Mitchell GS. Hippocampal brain-derived neurotrophic factor but not neurotrophin-3 increases more in mice selected for increased voluntary wheel running. Neuroscience. 2003;121(1):1–7. https://doi.org/10.1016/S0306-4522(03)00422-6

Jónás I, Schubert KA, Reijne AC, Scholte J, Garland T, Gerkema MP, Scheurink AJW, Nyakas C, van Dijk G. Behavioral traits are affected by selective breeding for increased wheel-running behavior in mice. Behav Genet. 2010;40(4):542–550. https://doi.org/10.1007/s10519-010–9359-8

Kane SL, Garland T, Jr, Carter PA. Basal metabolic rate of aged mice is affected by random genetic drift but not by selective breeding for high early-age locomotor activity or chronic wheel access. Physiol Biochem Zool. 2008;81(3):288–300. https://doi.org/10.1086/587093

Kaplan NL, Hudson RR, Langley CH. The "hitchhiking effect" revisited. Genetics. 1989;123(4):887–899. https://doi.org/10.1093/genetics/123.4.887

Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nat Rev Genet. 2014;15(6):379–393. https://doi.org/10.1038/nrg3734

Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, Lebreton M, Tino SP, Abdellaoui A, Hammerschlag AR, *et al.*; 23and Me Research Team, eQTLgen Consortium, International Cannabis Consortium, Social Science Genetic Association Consortium. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet. 2019;51(2):245–257. https://doi.org/10.1038/s41588-018–0309-3

Kelly SA, Bell TA, Selitsky SR, Buus RJ, Hua K, Weinstock GM, Garland T, Pardo-Manuel de Villena F, Pomp D. A novel intronic single nucleotide polymorphism in the *Myosin heavy polypeptide 4* gene is responsible for the mini-muscle phenotype characterized by major reduction in hind-limb muscle mass in mice. Genetics. 2013;195(4):1385–1395. https://doi.org/10.1534/genetics.113.154476

Kelly SA, Czech PP, Wight JT, Blank KM, Garland T, Jr. Experimental evolution and phenotypic plasticity of hindlimb bones in high-activity house mice. J Morphol. 2006;267(3):360–374. https://doi.org/10.1002/jmor.10407

Kelly SA, Gomes FR, Kolb EM, Malisch JL, Garland T, Jr. Effects of activity, genetic selection and their interaction on muscle metabolic capacities and organ masses in mice. J Exp Biol. 2017;220(Pt 6):1038–1047. https://doi.org/10.1242/jeb.148759

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002;12(6):996–1006. https://doi.org/10.1101/gr.229102

Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, Schoech A, Pasaniuc B, Price AL. Leveraging polygenic functional enrichment to improve GWAS power. Am J Hum Genet. 2019;104(1):65–75. https://doi.org/10.1016/j.ajhg.2018.11.008

Kim SK. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. PLoS One. 2018;13(7):e0200785. https://doi.org/10.1371/journal.pone.0200785

Kim Y, Stephan W. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics. 2002;160(2):765–777. https://doi.org/10.1093/genetics/160.2.765

Klomberg KF, Garland T, Swallow JG, Carter PA. Dominance, plasma testosterone levels, and testis size in house mice artificially selected for high activity levels. Physiol Behav. 2002;77(1):27–38. https://doi.org/10.1016/S0031-9384(02)00767-9

Kolb EM, Kelly SA, Garland T. Mice from lines selectively bred for high voluntary wheel running exhibit lower blood pressure during withdrawal from wheel access. Physiology & Behavior. 2013;112-113:49–55. https://doi.org/10.1016/j.physbeh.2013.02.010

Koteja P, Swallow JG, Carter PA, Garland T. Different effects of intensity and duration of locomotor activity on circadian period. J Biol Rhythms. 2003;18(6):491–501. https://doi.org/10.1177/0748730403256998

Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, Datta G, Davila-Velderrain J, McGuire D, Tian C, *et al.*; 23and Me Research Team, HUNT All-In Psychiatry. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat. Genet. 2019;51(2):237–244. https://doi.org/10.1038/s41588-018–0307-5

Long A, Liti G, Luptak A, Tenaillon O. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat Rev Genet. 2015;16(10):567–582. https://doi.org/10.1038/nrg3937

Malaguarnera L, Musumeci S. The immune response to *Plasmodium falciparum* malaria. Lancet Infect Dis. 2002;2(8):472–478. https://doi.org/10.1016/S1473-3099(02)00344-4

Malisch JL, Kelly SA, Bhanvadia A, Blank KM, Marsik RL, Platzer EG, Garland T. Lines of mice with chronically elevated baseline corticosterone levels are more susceptible to a parasitic nematode infection. Zoology. 2009;112(4):316–324. https://doi.org/10.1016/j.zool.2008.09.004

Malisch JL, Saltzman W, Gomes FR, Rezende EL, Jeske DR, Garland T. Baseline and stress-induced plasma corticosterone concentrations of mice selectively bred for high voluntary wheel running. Physiol Biochem Zool. 2007;80(1):146–156. https://doi.org/10.1086/508828

Mathes WF, Nehrenberg DL, Gordon R, Hua K, Garland T, Pomp D, Jr. Dopaminergic dysregulation in mice selectively bred for excessive exercise or obesity. Behav Brain Res. 2010;210(2):155–163. https://doi.org/10.1016/j.bbr.2010.02.016

Mayr E. Cause and effect in biology. Science. 1961;134(3489):1501–1506. https://doi.org/10.1126/science.134.3489.1501

Mayr E. Systematics and the Origin of Species. New York: Columbia University Press; 1942.

Meek TH, Lonquich BP, Hannon RM, Garland T, Jr. Endurance capacity of mice selectively bred for high voluntary wheel running. J Exp Biol. 2009;212(18):2908–2917. https://doi.org/10.1242/jeb.028886

Middleton KM, Goldstein BD, Guduru PR, Waters JF, Kelly SA, Swartz SM, Garland T. Variation in within-bone stiffness measured by nanoindentation in mice bred for high levels of voluntary wheel running. J Anat. 2010;216(1):121–131. https://doi.org/10.1111/j.1469–7580.2009.01175.x

Middleton KM, Shubin CE, Moore DC, Carter PA, Garland T, Swartz SM, Jr. The relative importance of genetics and phenotypic plasticity in dictating bone morphology and mechanics in aged mice: evidence from an artificial selection experiment. Zoology. 2008;111(2):135–147. https://doi.org/10.1016/j.zool.2007.06.003

Morris JA, Kemp JP, Youlten SE, Laurent L, Logan JG, Chai RC, Vulpescu NA, Forgetta V, Kleinman A, Mohanty ST, *et al.*; 23and Me Research Team. An atlas of genetic influences on osteoporosis in humans and mice. Nat. Genet. 2019;51(2):258–266. https://doi.org/10.1038/s41588-018–0302-x

Munafò M, Clark T, Johnstone E, Murphy M, Walton R. The genetic basis for smoking behavior: a systematic review and meta-

analysis. Nicotine Tob Res. 2004;6(4):583–597. https://doi.org/10.1080/14622200410001734030

Natarajan C, Hoffmann FG, Lanier HC, Wolf CJ, Cheviron ZA, Spangler ML, Weber RE, Fago A, Storz JF. Intraspecific polymorphism, interspecific divergence, and the origins of function-altering mutations in deer mouse hemoglobin. Mol Biol Evol. 2015;32(4):978–997. https://doi.org/10.1093/molbev/msu403

Nguyen QAT, Hillis D, Katada S, Harris T, Pontrello C, Garland T, Haga-Yamanaka S. Coadaptation of the chemosensory system with voluntary exercise behavior in mice. PLoS One. 2020;15(11):e0241758. https://doi.org/10.1371/journal.pone.0241758

Noble EP. D2 dopamine receptor gene in psychiatric and neurologic disorders and its phenotypes. Am J Med Genet B Neuropsychiatr Genet. 2003;116B(1):103–125. https://doi.org/10.1002/ajmg.b.10005

Oksenberg N, Ahituv N. The role of AUTS2 in neurodevelopment and human evolution. Trends Genet. 2013;29(10):600–608. https://doi.org/10.1016/j.tig.2013.08.001

Partridge L, Gems D. Mechanisms of aging: public or private? Nat Rev Genet. 2002;3(3):165–175. https://doi.org/10.1038/nrg753

Pavlicev M, Wagner GP. A model of developmental evolution: selection, pleiotropy and compensation. Trends Ecol Evol. 2012;27(6):316–322. https://doi.org/10.1016/j.tree.2012.01.016

Rao CR. Estimation of variance and covariance components—MINQUE theory. J Multivar Anal. 1971;1(3):257–275. https://doi.org/10.1016/0047-259X(71)90001-7

Rezende EL, Gomes FR, Chappell MA, Garland T, Jr. Running behavior and its energy cost in mice selectively bred for high voluntary locomotor activity. Physiol Biochem Zool. 2009;82(6):662–679. https://doi.org/10.1086/605917

Rhodes JS, Garland T, Jr. Differential sensitivity to acute administration of Ritalin, apormorphine, SCH 23390, but not raclopride in mice selectively bred for hyperactive wheel-running behavior. Psychopharmacology (Berl). 2003;167(3):242–250. https://doi.org/10.1007/s00213-003-1399-9

Rhodes JS, Hosack GR, Girard I, Kelley AE, Mitchell GS, Garland T. Differential sensitivity to acute administration of cocaine, GBR 12909, and fluoxetine in mice selectively bred for hyperactive wheel-running behavior. Psychopharmacology (Berl). 2001;158(2):120–131. https://doi.org/10.1007/s002130100857

Rhodes JS, van Praag H, Jeffrey S, Girard I, Mitchell GS, Garland T, Gage FH. Exercise increases hippocampal neurogenesis to high levels but does not improve spatial learning in mice bred for increased voluntary wheel running. Behav Neurosci. 2003;117(5):1006–1016. https://doi.org/10.1037/0735–7044.117.5.1006

Roberts MD, Ruegsegger GN, Brown JD, Booth FW. Mechanisms associated with physical activity behavior: insights from rodent experiments. Exerc Sport Sci Rev. 2017;45(4):217–222. https://doi.org/10.1249/JES.0000000000000124

Schmill MP. Neuroanatomical, behavioral, and physiological correlates of high voluntary wheel running [thesis]; Riverside, California: University of California, Riverside. 2021.

Shen H, Watanabe M, Tomasiewicz H, Glass J. Genetic deletions of NCAM and PSA impair circadian function in the mouse. Physiol Behav. 2001;73(1–2):185–193. https://doi.org/10.1016/S0031-9384(01)00468-1

Simões P, Santos J, Fragata I, Mueller LD, Rose MR, Matos M. How repeatable is adaptive evolution? The role of geographical origin and founder effects in laboratory adaptation. Evolution. 2008;62(8):1817–1829. https://doi.org/10.1111/j.1558–5646.2008.00423.x

Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974;23(1):23–35. https://doi.org/10.1017/S0016672300014634

Smith JR, Hayman GT, Wang S-J, Laulederkind SJF, Hoffman MJ. The Year of the Rat: The Rat Genome Database at 20: a multi-species knowledgebase and analysis platform. Nucleic Acids Res. 2019. https://doi.org/10.1093/nar/gkz1041

Smits P, Li P, Mandel J, Zhang Z, Deng JM, Behringer RR, de Crombrugghe B, Lefebvre V. The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. Dev Cell. 2001;1(2):277–290. https://doi.org/10.1016/s1534-5807(01)00003-x

Stinchcombe JR, Weinig C, Heath KD, Brock MT, Schmitt J. Polymorphic genes of major effect: consequences for variation, selection and evolution in *Arabidopsis thaliana*. Genetics. 2009;182(3):911–922. https://doi.org/10.1534/genetics.108.097030

Storz JF. Gene duplication and evolutionary innovations in hemoglobin-oxygen transport. Physiology. 2016;31(3):223–232. https://doi.org/10.1152/physiol.00060.2015

Swallow JG, Garland T, Jr, Carter PA, Zhan W-Z, Sieck GC. Effects of voluntary activity and genetic selection on aerobic capacity in house mice (*Mus domesticus*). J Appl Physiol (1985). 1998;84(1):69–76. https://doi.org/10.1152/jappl.1998.84.1.69

Swallow JG, Koteja P, Carter PA, Garland T, Jr. Artificial selection for increased wheel-running activity in house mice results in decreased body mass at maturity. J Exp Biol. 1999;202(Pt 18):2513–2520. https://doi.org/10.1242/jeb.202.18.2513

Swallow JG, Rhodes JS, Garland T, Jr. Phenotypic and evolutionary plasticity of organ masses in response to voluntary exercise in house mice. Integr Comp Biol. 2005;45(3):426–437. https://doi.org/10.1093/icb/45.3.426

Syme DA, Evashuk K, Grintuch B, Rezende EL, Garland T. Contractile abilities of normal and "mini" triceps surae muscles from mice (*Mus domesticus*) selectively bred for high voluntary wheel running. J Appl Physiol (1985). 2005;99(4):1308–1316. https://doi.org/10.1152/japplphysiol.00369.2005

Tanksley SD. Mapping polygenes. Annu Rev Genet. 1993;27:205–233. https://doi.org/10.1142/9789812835857_0035

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. The molecular diversity of adaptive convergence. Science. 2012;335(6067):457–461. https://doi.org/10.1126/science.1212986

Tomasiewicz H, Ono K, Yee D, Thompson C, Goridis C, Rutishauser U, Magnuson T. Genetic deletion of a neural cell adhesion molecule variant (N-CAM-180) produces distinct defects in the central nervous system. Neuron. 1993;11(6):1163–1174. https://doi.org/10.1016/0896–6273(93)90228-J

van Rooij E, Quiat D, Johnson BA, Sutherland LB, Qi X, Richardson JA, Kelm RJ, Olson EN. A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. Dev Cell. 2009;17(5):662–673. https://doi.org/10.1016/j.devcel.2009.10.013

Wallace IJ, Middleton KM, Lublinsky S, Kelly SA, Judex S, Garland T, Demes B. Functional significance of genetic variation underlying limb bone diaphyseal structure. Am J Phys Anthropol. 2010;143(1):21–30. https://doi.org/10.1002/ajpa.21286

Wallace IJ, Tommasini SM, Judex S, Garland T, Demes B. Genetic variations and physical activity as determinants of limb bone morphology: an experimental approach using a mouse model. Am J Phys Anthropol. 2012;148(1):24–35. https://doi.org/10.1002/ajpa.22028

Wone BWM, Yim WC, Schutz H, Meek TH, Garland T. Mitochondrial haplotypes are not associated with mice selectively bred for high voluntary wheel running. Mitochondrion. 2019;46:134–139. https://doi.org/10.1016/j.mito.2018.04.002

Wong AH, Buckle CE, Van Tol HH. Polymorphisms in dopamine receptors: what do they tell us? Eur J Pharmacol. 2000;410(2–3):183–203. https://doi.org/10.1016/S0014-2999(00)00815-3

Xu S, Garland T. A mixed model approach to genome-wide association studies for selection signatures, with application to mice bred for voluntary exercise behavior. Genetics. 2017;207(2):785–799. https://doi.org/10.1534/genetics.117.300102

Yoshida M, Minamisawa S, Shimura M, Komazaki S, Kume H, Zhang M, Matsumura K, Nishi M, Saito M, Saeki Y, *et al.* Impaired Ca2+ store functions in skeletal and cardiac muscle cells from sarcalumenin-deficient mice. J Biol Chem. 2005;280(5):3500–3506. https://doi.org/10.1074/jbc.M406618200

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28(24):3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Communicating editor: J. Gleason