Title
Genomics of High Voluntary Running Behavior in Mice

Permalink
https://escholarship.org/uc/item/1722x41q

Author
Hillis, David Anthony

Publication Date
2022

Supplemental Material
https://escholarship.org/uc/item/1722x41q#supplemental

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Genomics of High Voluntary Running Behavior in Mice

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics, and Bioinformatics

by

David Anthony Hillis

December 2022

Dissertation Committee:
        Dr. Theodore Garland, Jr., Chairperson
        Dr. Zhenyu (Arthur) Jia
        Dr. Shizhong Xu

The Dissertation of David Anthony Hillis is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to thank my advisor Dr. Theodore Garland for all of his help and guidance through the creation of this dissertation. I have learned a great deal of about biology, statistics, teaching, and writing from him. Thank you to the numerous graduate students in the Garland lab and the Genetics, Genomics, and Bioinformatics program who have helped me with my research, lab work, and teaching and provided me with cherished friendships throughout these years and into the future. Thank you to God, family, and friends who have provided consent support and encouragement through the years.

Thank you to collaborators without whom the data used in this dissertation would not be available. The copyright material is held jointly with co-authors for each chapter and specific contributions can be found at the beginning of each chapter. Chapters One and Two of this dissertation has been adapted from published articles in Genetics (full citation below):

Hillis, David A, Liran Yadgary, George M Weinstock, Fernando Pardo-Manuel de Villena, Daniel Pomp, Alexandra S Fowler, Shizhong Xu, Frank Chan, and Theodore Garland. "Genetic Basis of Aerobically Supported Voluntary Exercise: Results from a Selection Experiment with House Mice." *Genetics* 216, no. 3 (November 1, 2020): 781–804. https://doi.org/10.1534/genetics.120.303668.

Hillis, David A., and Theodore Garland Jr. "Multiple Solutions at the Genomic Level in Response to Selective Breeding for High Locomotor Activity." *Genetics*, October 28, 2022. https://doi.org/10.1093/genetics/iyac165.

ABSTRACT OF THE DISSERTATION

Genomics of High Voluntary Running Behavior in Mice

by

David Anthony Hillis

Doctor of Philosophy, Graduate Program in Genetics, Genomics, and Bioinformatics
University of California, Riverside, December 2022
Dr. Theodore Garland Jr., Chairperson

Physical activity is an essential component of the life history for most animals, and it also promotes both physical and mental health. Activity is a complex trait, affected by both genetics and numerous environmental factors, and the result of both motivation and physical ability. To elucidate the evolution of physical activity, the High Runner selection experiment was begun with 4 lines of mice bred for high voluntary wheel running (HR lines) and 4 non-selected control (C) lines. Although the HR lines evolved to run ~3 times as much as C lines daily, and numerous physiological and morphological differences have been documented, little is known about the genetic factors that differentiate HR and C lines.

The first chapter utilizes whole-genome sequence data from 79 individuals from the 8 lines at generation 61 to identify signatures of selection. Three analytical methods agreed in identifying 13 genomic regions. These regions included genes associated with

reward pathways and neural development, limb development, and other intuitive functions for wheel running.

The second chapter uses the same genomic data and performs similar analyses, except dropping one line at a time. This identifies several new selection signatures and highlights how the replicate HR lines have responded to selection via "multiple solutions." The greatest change comes from dropping line HR3, which became fixed for a gene of major effect (i.e., the mini-muscle allele) that substantially alters the genetic background.

The third chapter analyzes generation 22 allele frequencies obtained from sequencing pooled samples of approximately 10 males and 10 females from each line. Analyses identified not only many more selection signatures than at generation 61, but also very different genomic regions, with many of the strongest signatures in one generation being only weakly supported in the other. Simulations demonstrated that a hypothetical physiological constraint on wheel running reduces the power to detect selection and increases the likelihood of detectable signatures changing as selection limits are reached and passed.

Each chapter identifies candidate genes for wheel-running behavior, related to both motivation and ability. Overall, these results enhance our understanding of the genetics and evolution of complex traits.

Table of Contents

List of Figures

List of Tables

INTRODUCTION

Complex traits

One of the major fields of research in genetics is the study of inheritance of complex traits. Complex traits are composed of various lower-level traits that may not always contribute to the higher-level trait in an intuitive manner (Garland, Jr. *et al.* 2016). For example, a person's sense of taste might influence their growth rate and body weight by altering dietary choices. Moreover, lower-level traits, such as circulating concentrations of hormones, may serve functions for multiple complex traits, which can lead to both phenotypic and genetic correlations among traits. For example, circulating levels of glucocorticoids may affect both motivation to perform certain behaviors and the ability of muscles to carry out behavior (Garland, Jr. *et al.* 2016). Additionally, complex traits are controlled by numerous genetic and environmental factors that may interact in various ways.

One example of a complex trait is human height. Underlying traits associated with height will include the size of skeletal components (e.g., limb bones), as well as the aspects of metabolism and the endocrine system that control growth rate. Environmental contributing factors would include available diet or opportunities for physical exercise. These factors can interact in various ways. For example, an individual's diet choices will dictate how much total energy they have available, and then that energy will be partitioned among various components, including physical activity (Garland, Jr. *et al.* 2011a). Additionally, underlying these lower-level traits are thousands of genetic

components (Wood *et al.* 2014) that may influence, for example, bone development, preferred diet and foods, or even the propensity to engage in physical activity, including voluntary exercise.

Exercise behavior as a complex trait

Physical activity is a complex behavioral trait, controlled by various environmental factors and biological factors (Lightfoot *et al.* 2018). The environment can affect the motivation for physical activity through such factors as the availability of food, the need to avoid predators, etc. These can be further complicated by availability of resources, social interactions with members of its own species or other species, and much more.

Numerous biological components contribute to physical activity behavior, including various systems that affect motivated behaviors, such as dopamine and serotonin signaling (Freed and Yamamoto 1985; Simonen *et al.* 2003; Mathes *et al.* 2010; Leinninger *et al.* 2011; Claghorn *et al.* 2016; Cordeiro *et al.* 2017). Aside from motivation, the expression of any behavior, including physical activity, depends on ability, which is determined by bone morphology, skeletal muscle physiology, aerobic capacity (VO_{2Max}), metabolism, and more (Lightfoot *et al.* 2018).

Evolutionary and medical relevance of physical activity

Physical activity is of particular interest due to its relevance for both the behavioral ecology and evolutionary history of many species and human health and wellbeing. Physical activity has notable implications for evolution in that new methods

of locomotion (e.g., flight) can avail new ways to forage, new landscapes to be traversed, novel resources to be acquired, and more. Aside from the evolution of new modes of locomotion, almost all animals, at some point during their lifecycle, need to move to survive. The need for physical activity is expected to lead to strong selection on locomotor behavior, ability, and all of the lower-level (subordinate) traits that are involved. Indeed, Dickinson et al. (2000) claimed that "Locomotion, movement through the environment, is the behavior that most dictates the morphology and physiology of animals." Even within a given mode of locomotion, improved physical activity can evolve by various mechanisms, thus allowing for the emergence of "multiple solutions" to the same evolutionary need (Garland, Jr. *et al.* 2011b). For example, a predator could evolve faster running to outrun its prey in a high-speed pursuit or more energy efficient running to be able to catch the prey after a long-distance pursuit.

The medical implications of physical activity in humans are expansive. Physical activity or exercise (physical activity for the purpose of recreation or health benefits) promotes skeletal development, reduces the risk of heart disease and some cancers, facilitates weight control, and positively affects mental health (Manley 1996; Booth *et al.* 2008, 2012; Lee *et al.* 2012). Despite the fact that many people know the numerous benefits of exercise, few people get sufficient levels of exercise. Guthold et al. (2018) conducted analyses of 358 surveys across 168 countries and found that many countries had a large proportion of people who did not achieve sufficient levels of exercise. The lack of sufficient exercise is predicted to have cost the United States more than 100

billion dollars annually between the years of 2006 and 2011 (Carlson *et al.* 2015), thus adding economic incentives to the health benefits.

The High Runner (HR) mouse artificial selection experiment

To better understand physical activity and exercise behavior the High Runner mouse selection experiment was began in 1993 (Swallow *et al.* 1998a). This experiment started with 224 outbred Hsd:ICR mice. These mice were randomly bred for 2 generations, after which individuals were randomly chosen to be part of one of 8 closed lines, each with 10 breeding pairs. Of the 8 lines, four were randomly chosen to serve as non-selected controls lines (C1, C2, C4, and C5) and four to serve as selected High Runner lines (HR3, HR6, HR7, and HR8). With each generation, mice are weaned at 3 weeks of age and given wheel access (which they could interact with voluntarily) at ~6-8 weeks of age for 6 days with *ad lib* food and water. Wheel running is measured daily in terms of number of revolutions. For control lines, two males and females from each family are chosen to breed independent of wheel running levels. In the High Runner (HR) lines, the male and female from each family with the highest wheel running on days 5 and 6 are chosen as breeders (no sib-mating is allowed within either HR or control lines). For logistical purposes, mice were measured in 3 (or more) batches, where a mouse in any batch has a clean cage and bedding, with fresh food and water, but mice after batch 1 have a wheel that has not been cleaned from the previous mouse or mice (which could explain some results seen in all three chapters).

Statistically significant differences in wheel-running behavior were observed as early as generation 6 (T. Garland, Jr. personal communication). Additionally, numerous

physiological and morphological differences between the HR and control lines have been documented (Rhodes *et al.* 2005; Swallow *et al.* 2009; Garland, Jr. *et al.* 2011a; Wallace and Garland, Jr. 2016). These include traits associated with motivation to run, such as changes in dopamine (Rhodes *et al.* 2001; Mathes *et al.* 2010), serotonin (Waters *et al.* 2013), and endocannabinoid signaling (Thompson *et al.* 2017), as well as changes in brain size and structure (Kolb *et al.* 2013a). Changes associated with ability to run have also been demonstrated, including endurance capacity during forced treadmill exercise (Meek *et al.* 2009), maximal aerobic capacity (VO_{2Max}) (e.g., Swallow *et al.* 1998b; Kolb *et al.* 2010; Dlugosz *et al.* 2013; Cadney *et al.* 2021), heart size (Kolb *et al.* 2010, 2013b; Kelly *et al.* 2017), skeletal muscle physiology (Dumke *et al.* 2001; Syme *et al.* 2005; Guderley *et al.* 2008; Castro *et al.* in press), and bone morphology (Garland, Jr. and Freeman 2005; Kelly *et al.* 2006; Middleton *et al.* 2008, 2010; Wallace *et al.* 2010, 2012; Castro and Garland, Jr. 2018; Copes *et al.* 2018; Schwartz *et al.* 2018).

Previous genetic work on the High Runner mice

The HR and control lines have been the subject of several genetic analyses, using various approaches, ranging from quantitative genetics, through line crosses, mapping of QTL and eQTL, SNP chips to identify divergent chromosomal regions, and whole-genome sequencing. For example, Careau et al. (2013) applied the "animal model" to the first 31 generations and estimated the generation at which each HR line reached a selection limit (plateau) for wheel running (1-27, depending on sex and line). They also estimated selection differentials and selection showed that although narrow-sense

heritability declined across generations in the HR lines, this decline was not sufficient to explain the selection limits in all of the HR lines. Among other results, they also documented strong seasonal variation in running, which suggests the presence of an endogenous annual clock.

Another study identified the mini-muscle locus on the *Myo4* gene (Kelly *et al.* 2013), which is associated with a rare recessive point mutation that causes a drastic reduction in Type IIB muscle fibers among other pleiotropic effects in mice homozygous for the allele (see Chapter 2). Other genetic studies include identification of eQTLs in the brain (Kelly *et al.* 2012) and right triceps surae (Kelly *et al.* 2014) using advance intercrossed lines between one HR line and C57BL/6J. These studies by Kelly et al. (2012, 2014) identified various candidate genes for expression in the brain (*Insig2*, *Socs2*, *DBY*, *Arrdc4*, *Prcp*, *IL15*) and in skeletal muscle (*Insig2*, *Prcp*, *Sparc*). Additionally, these eQTL studies found that much of the gene regulation was via trans-acting regulators (regulator >10 mbp away from affected gene).

The first study that explored differentiation in allele frequencies between the HR and control lines was performed by Xu and Garland (2017). This study used MegaMUGA technology to determine individual mouse genotypes for 25,318 SNPs (single nucleotide polymorphisms) for each of 80 mice (10 mice for each of the 8 lines) from generation 61. With the individual mouse data, more powerful statistical tests could be performed to identify the differentiated loci between the two linetypes, as compared with pooled sequence data. This study showed that the mixed model analyses with mivque (minimum variance quadratic unbiased estimation) estimation method was more

powerful than regularized T-tests implemented on pooled sequence data. However, Xu and Garland (2017) did not go into detail regarding the biological implications of the genomic regions identified.

In this dissertation, I utilize whole-genome sequencing of individual mice at generation 61 and pooled sequencing for each line at generation 22. With these data, I identify numerous chromosomal regions differentiated between the HR and control linetypes, and consider the biological significance of genes in these regions, which can serve as targets for future functional studies (Chapter 1). Additionally, I use results of analyses dropping individual lines to demonstrate divergent responses to selection among the HR lines (Chapter 2). I also show how the regions detected as differentiated can change over several generations, even with continued selection beyond a selection limit (Chapter 3). Overall, this dissertation contributes to our understanding of the genetics and evolution of a complex behavioral trait that is relevant for human health.

REFERENCES

Booth F. W., M. J. Laye, S. J. Lees, R. S. Rector, and J. P. Thyfault, 2008 Reduced physical activity and risk of chronic disease: the biology behind the consequences. Eur. J. Appl. Physiol. 102: 381–390. https://doi.org/10.1007/s00421-007-0606-5

Booth F. W., C. K. Roberts, and M. J. Laye, 2012 Lack of exercise is a major cause of chronic diseases, in *Comprehensive Physiology*, edited by Terjung R. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Cadney M. D., L. Hiramatsu, Z. Thompson, M. Zhao, J. C. Kay, *et al.*, 2021 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. 234: 113389. https://doi.org/10.1016/j.physbeh.2021.113389

Careau V., M. E. Wolak, P. A. Carter, and T. Garland, Jr., 2013 Limits to behavioral evolution: the quantative genetics of a complex trait under directional selection. Evolution 67: 3102–3119. https://doi.org/10.1111/evo.12200

Carlson S. A., J. E. Fulton, M. Pratt, Z. Yang, and E. K. Adams, 2015 Inadequate physical activity and health care expenditures in the United States. Prog. Cardiovasc. Dis. 57: 315–323. https://doi.org/10.1016/j.pcad.2014.08.002

Castro A. A., and T. Garland, Jr., 2018 Evolution of hindlimb bone dimensions and muscle masses in house mice selectively bred for high voluntary wheel-running behavior. J. Morphol. 279: 766–779. https://doi.org/10.1002/jmor.20809

Castro A. A., T. Garland Jr, S. Ahmed, and N. Holt, in press Trade-offs in muscle physiology in selectively bred High Runner mice. J. Exp. Biol.

Claghorn G. C., I. A. T. Fonseca, Z. Thompson, C. Barber, and T. Garland, Jr., 2016 Serotonin-mediated central fatigue underlies increased endurance capacity in mice from lines selectively bred for high voluntary wheel running. Physiol. Behav. 161: 145–154. https://doi.org/10.1016/j.physbeh.2016.04.033

Copes L. E., H. Schutz, E. M. Dlugsoz, S. Judex, and T. Garland, Jr., 2018 Locomotor activity, growth hormones, and systemic robusticity: An investigation of cranial vault thickness in mouse lines bred for high endurance running. Am. J. Phys. Anthropol. 166: 442–458. https://doi.org/10.1002/ajpa.23446

Cordeiro L. M. S., P. C. R. Rabelo, M. M. Moraes, F. Teixeira-Coelho, C. C. Coimbra, *et al.*, 2017 Physical exercise-induced fatigue: The role of serotonergic and

dopaminergic systems. Braz. J. Med. Biol. Res. 50. https://doi.org/10.1590/1414-431x20176432

Dickinson M. H., C. T. Farley, R. J. Full, M. A. R. Koehl, R. Kram, *et al.*, 2000 How animals move: an integrative view. Science 288: 100–106. https://doi.org/10.1126/science.288.5463.100

Dlugosz E. M., H. Schutz, T. H. Meek, W. Acosta, C. J. Downs, *et al.*, 2013 Immune response to a Trichinella spiralis infection in house mice from lines selectively bred for high voluntary wheel running. J. Exp. Biol. 216: 4212–4221. https://doi.org/10.1242/jeb.087361

Dumke C. L., J. S. Rhodes, T. Garland, E. Maslowski, J. G. Swallow, *et al.*, 2001 Genetic selection of mice for high voluntary wheel running: effect on skeletal muscle glucose uptake. J. Appl. Physiol. 91: 1289–1297. https://doi.org/10.1152/jappl.2001.91.3.1289

Freed C., and B. Yamamoto, 1985 Regional brain dopamine metabolism: a marker for the speed, direction, and posture of moving animals. Science 229: 62–65. https://doi.org/10.1126/science.4012312

Garland, Jr. T., and P. W. Freeman, 2005 Selective breeding for high endurance running increases hindlimb symmetry. Evolution 59: 1851–1854.

Garland, Jr. T., H. Schutz, M. A. Chappell, B. K. Keeney, T. H. Meek, *et al.*, 2011a The biological control of voluntary exercise, spontaneous physical activity and daily energy expenditure in relation to obesity: human and rodent perspectives. J. Exp. Biol. 214: 206–229. https://doi.org/10.1242/jeb.048397

Garland, Jr. T., S. A. Kelly, J. L. Malisch, E. M. Kolb, R. M. Hannon, *et al.*, 2011b How to run far: multiple solutions and sex-specific responses to selective breeding for high voluntary activity levels. Proc. R. Soc. B Biol. Sci. 278: 574–581. https://doi.org/10.1098/rspb.2010.1584

Garland, Jr. T., M. Zhao, and W. Saltzman, 2016 Hormones and the evolution of complex traits: insights from artificial selection on behavior. Integr. Comp. Biol. 56: 207–224. https://doi.org/10.1093/icb/icw040

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

Guthold R., G. A. Stevens, L. M. Riley, and F. C. Bull, 2018 Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358

population-based surveys with 1·9 million participants. Lancet Glob. Health 6: e1077–e1086. https://doi.org/10.1016/S2214-109X(18)30357-7

Kelly S. A., P. P. Czech, J. T. Wight, K. M. Blank, and T. Garland, Jr., 2006 Experimental evolution and phenotypic plasticity of hindlimb bones in high-activity house mice. J. Morphol. 267: 360–374. https://doi.org/10.1002/jmor.10407

Kelly S. A., D. L. Nehrenberg, K. Hua, T. Garland, Jr., and D. Pomp, 2012 Functional genomic architecture of predisposition to voluntary exercise in mice: expression QTL in the brain. Genetics 191: 643–654. https://doi.org/10.1534/genetics.112.140509

Kelly S. A., T. A. Bell, S. R. Selitsky, R. J. Buus, K. Hua, *et al.*, 2013 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 195: 1385–1395. https://doi.org/10.1534/genetics.113.154476

Kelly S. A., D. L. Nehrenberg, K. Hua, T. Garland, Jr., and D. Pomp, 2014 Quantitative genomics of voluntary exercise in mice: transcriptional analysis and mapping of expression QTL in muscle. Physiol. Genomics 46: 593–601. https://doi.org/10.1152/physiolgenomics.00023.2014

Kelly S. A., F. R. Gomes, E. M. Kolb, J. L. Malisch, and T. Garland, Jr., 2017 Effects of activity, genetic selection and their interaction on muscle metabolic capacities and organ masses in mice. J. Exp. Biol. 220: 1038–1047. https://doi.org/10.1242/jeb.148759

Kolb E. M., S. A. Kelly, K. M. Middleton, L. S. Sermsakdi, M. A. Chappell, *et al.*, 2010 Erythropoietin elevates VO2,max but not voluntary wheel running in mice. J. Exp. Biol. 213: 510–519. https://doi.org/10.1242/jeb.029074

Kolb E. M., E. L. Rezende, L. Holness, A. Radtke, S. K. Lee, *et al.*, 2013a Mice selectively bred for high voluntary wheel running have larger midbrains: support for the mosaic model of brain evolution. J. Exp. Biol. 216: 515–523. https://doi.org/10.1242/jeb.076000

Kolb E. M., S. A. Kelly, and T. Garland, Jr., 2013b Mice from lines selectively bred for high voluntary wheel running exhibit lower blood pressure during withdrawal from wheel access. Physiol. Behav. 112–113: 49–55. https://doi.org/10.1016/j.physbeh.2013.02.010

Lee I.-M., E. J. Shiroma, F. Lobelo, P. Puska, S. N. Blair, *et al.*, 2012 Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden

of disease and life expectancy. The Lancet 380: 219–229.
https://doi.org/10.1016/S0140-6736(12)61031-9

Leinninger G. M., D. M. Opland, Y.-H. Jo, M. Faouzi, L. Christensen, *et al.*, 2011 Leptin
Action via Neurotensin Neurons Controls Orexin, the Mesolimbic Dopamine
System and Energy Balance. Cell Metab. 14: 313–323.
https://doi.org/10.1016/j.cmet.2011.06.016

Lightfoot J. T., E. J. C. De Geus, F. W. Booth, M. S. Bray, M. Den Hoed, *et al.*, 2018
Biological/genetic regulation of physical activity level: Consensus from
GenBioPAC. Med. Sci. Sports Exerc. 50: 863–873.
https://doi.org/10.1249/MSS.0000000000001499

Manley A. F., 1996 *Physical activity and health: a report of the Surgeon General*.
DIANE Publishing.

Mathes W. F., D. L. Nehrenberg, R. Gordon, K. Hua, T. Garland, Jr., *et al.*, 2010
Dopaminergic dysregulation in mice selectively bred for excessive exercise or
obesity. Behav. Brain Res. 210: 155–163.
https://doi.org/10.1016/j.bbr.2010.02.016

Meek T. H., B. P. Lonquich, R. M. Hannon, and T. Garland, Jr., 2009 Endurance capacity
of mice selectively bred for high voluntary wheel running. J. Exp. Biol. 212:
2908–2917. https://doi.org/10.1242/jeb.028886

Middleton K. M., C. E. Shubin, D. C. Moore, P. A. Carter, T. Garland, Jr., *et al.*, 2008
The relative importance of genetics and phenotypic plasticity in dictating bone
morphology and mechanics in aged mice: Evidence from an artificial selection
experiment. Zoology 111: 135–147. https://doi.org/10.1016/j.zool.2007.06.003

Middleton K. M., B. D. Goldstein, P. R. Guduru, J. F. Waters, S. A. Kelly, *et al.*, 2010
Variation in within-bone stiffness measured by nanoindentation in mice bred for
high levels of voluntary wheel running. J. Anat. 216: 121–131.
https://doi.org/10.1111/j.1469-7580.2009.01175.x

Rhodes J. S., G. R. Hosack, I. Girard, A. E. Kelley, G. S. Mitchell, *et al.*, 2001
Differential sensitivity to acute administration of cocaine, GBR 12909, and
fluoxetine in mice selectively bred for hyperactive wheel-running behavior.
Psychopharmacology (Berl.) 158: 120–131.
https://doi.org/10.1007/s002130100857

Rhodes J. S., S. C. Gammie, and T. Garland Jr, 2005 Neurobiology of mice selected for
high voluntary wheel-running activity. Integr. Comp. Biol. 45: 438–455.
https://doi.org/10.1093/icb/45.3.438

Schwartz N. L., B. A. Patel, T. Garland, Jr., and A. M. Horner, 2018 Effects of selective breeding for high voluntary wheel-running behavior on femoral nutrient canal size and abundance in house mice. J. Anat. 233: 193–203. https://doi.org/10.1111/joa.12830

Simonen R. L., T. Rankinen, L. Pérusse, A. S. Leon, J. S. Skinner, *et al.*, 2003 A dopamine D2 receptor gene polymorphism and physical activity in two family studies. Physiol. Behav. 78: 751–757. https://doi.org/10.1016/S0031-9384(03)00084-2

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. https://doi.org/10.1023/A:1021479331779

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. https://doi.org/10.1152/jappl.1998.84.1.69

Swallow J. G., J. P. Hayes, P. Koteja, and T. Garland, 2009 Selection experiments and experimental evolution of performance and physiology, pp. 301–351 in *Experimental evolution: concepts, methods, and applications of selection experiments*, University of California Press, Berkeley.

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

Thompson Z., D. Argueta, T. Garland, Jr., and N. DiPatrizio, 2017 Circulating levels of endocannabinoids respond acutely to voluntary exercise, are altered in mice selectively bred for high voluntary wheel running, and differ between the sexes. Physiol. Behav. 170: 141–150. https://doi.org/10.1016/j.physbeh.2016.11.041

Wallace I. J., K. M. Middleton, S. Lublinsky, S. A. Kelly, S. Judex, *et al.*, 2010 Functional significance of genetic variation underlying limb bone diaphyseal structure. Am. J. Phys. Anthropol. 143: 21–30. https://doi.org/10.1002/ajpa.21286

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

Wallace I. J., and T. Garland, Jr., 2016 Mobility as an emergent property of biological organization: Insights from experimental evolution: Mobility and biological

organization. Evol. Anthropol. Issues News Rev. 25: 98–104.
https://doi.org/10.1002/evan.21481

Waters R. P., R. B. Pringle, G. L. Forster, K. J. Renner, J. L. Malisch, *et al.*, 2013
Selection for increased voluntary wheel-running affects behavior and brain
monoamines in mice. Brain Res. 1508: 9–22.
https://doi.org/10.1016/j.brainres.2013.01.033

Wood A. R., The Electronic Medical Records and Genomics (eMERGE) Consortium,
The MIGen Consortium, The PAGE Consortium, The LifeLines Cohort Study, *et
al.*, 2014 Defining the role of common variation in the genomic and biological
architecture of adult human height. Nat. Genet. 46: 1173–1186.
https://doi.org/10.1038/ng.3097

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

Chapter 1

Genetic Basis of Aerobically Supported Voluntary Exercise: Results from a Selection Experiment with House Mice

David A. Hillis, Liran Yadgary, George M. Weinstock, Fernando Pardo-Manuel de Villena, Daniel Pomp, Alexandra S. Fowler, Shizhong Xu, Frank Chan, and Theodore Garland, Jr.

Author contributions: Conceptualization, D.A.H., L.Y., F.P.M.dEV., D.P., S.X., F.C., T.G.; investigation, D.A.H., L.Y., G.M.W., F.P.M.dEV., D.P., A.S.F., F.C., T.G.; software, D.A.H., L.Y., S.X.; formal analysis, D.A.H., L.Y., A.S.F., S.X., F.C., T.G.; writing – original draft, D.A.H., L.Y., A.S.F., F.C., T.G.; writing review and editing, D.A.H., L.Y., G.M.W., F.P.M.dEV., D.P., A.S.F., S.X., F.C., T.G.

ABSTRACT

The biological basis of exercise behavior is increasingly relevant for maintaining healthy lifestyles. Various quantitative genetic studies and selection experiments have conclusively demonstrated substantial heritability for exercise behavior in both humans and laboratory rodents. In the "High Runner" selection experiment, 4 replicate lines of *Mus domesticus* were bred for high voluntary wheel running (HR), along with 4 non-selected control (C) lines. After 61 generations, the genomes of 79 mice (9-10 from each line) were fully sequenced and single nucleotide polymorphisms (SNPs) were identified. We used nested ANOVA with MIVQUE estimation and other approaches to compare allele frequencies between the HR and C lines for both SNPs and haplotypes. Approximately 85 genomic regions, across all somatic chromosomes, showed evidence of differentiation. Twelve of these regions were differentiated by all methods of analysis. Gene function was inferred largely using Panther gene ontology terms and KO phenotypes associated with genes of interest. Some of the differentiated genes are known to be associated with behavior/motivational systems and/or athletic ability, including *Sorl1, Dach1*, and *Cdh10*. Sorl1 is a sorting protein associated with cholinergic neuron morphology, vascular wound healing, and metabolism. *Dach1* is associated with limb bud development and neural differentiation. *Cdh10* is a calcium ion binding protein associated with phrenic neurons. Overall, these results indicate that selective breeding for high voluntary exercise has resulted in changes in allele frequencies for multiple genes associated with both motivation and ability for endurance exercise, providing candidate genes that may explain phenotypic changes observed in previous studies.

15

INTRODUCTION

Most traits of interest in biology are complex, modulated by numerous genetic and environmental factors, and comprised of multiple lower-level (subordinate) traits that often influence higher-level traits in nonintuitive ways (Garland, Jr. *et al.* 2016; Sella and Barton 2019). Examples of complex traits include human height, which is influenced by more than 9,500 quantitative trait loci (QTL) (Wood *et al.* 2014), as well as one's susceptibility to various psychological diseases (Horwitz *et al.* 2019).

One complex trait of great interest to medicine is exercise behavior. Exercise has been linked to numerous health benefits, including muscle and bone strength, weight control, reduced cardiac disease, and improved mental health(Manley 1996; Lightfoot *et al.* 2018). Nonetheless, the majority of Americans are not getting sufficient exercise and this problem is common world-wide (Guthold *et al.* 2018). Not only does insufficient exercise contribute to such health issues as obesity and diabetes (Booth *et al.* 2002; Cornier *et al.* 2008; Myers *et al.* 2017), but it also increases healthcare costs in the United States, e.g., by more than $100 billion annually between the years of 2006-2011 (Carlson *et al.* 2015). Conversely, higher levels of physical activity promote physical fitness and cardiovascular health, while lowering risk for depression, anxiety-related disorders, obesity, Type 2 diabetes, and mortality (Blair and Morris 2009; Matta Mello Portugal *et al.* 2013; Mok *et al.* 2019).

A variety of human studies have been conducted to determine the genes or chromosomal regions that modulate various components of exercise behavior, including both motivation and/or capability to exercise (Lightfoot *et al.* 2018). Many of these

studies use observational methods to compare humans who engage in either frequent and/or strenuous exercise with those who are less active (Kostrzewa and Kas 2014; Lin *et al.* 2017). Historically, the most common approach to measuring human exercise levels was by use of questionnaires, which can be of dubious reliability, but an increasing number of studies use accelerometers (Prince *et al.* 2008; Dyrstad *et al.* 2014). Detecting QTL in these studies is generally done with genome-wide association studies (GWAS), which rely on phenotypic and genetic data from many individuals within a population and can identify particularly strong correlations between the phenotype and key genetic markers and loci.

Various QTL identified in humans are associated with motivation, e.g., dopaminergic regulation. Dopamine is a well-established modulator of exercise motivation or reward (Garland, Jr. *et al.* 2011b). Various genes associated with the dopamine pathway are associated with exercise behavior in humans (Simonen *et al.* 2003; Loos *et al.* 2005; De Moor *et al.* 2009). The large body of evidence that dopamine signaling is a major component of exercise motivation dwarfs other motivational systems that have been associated with exercise, including serotonin and endocannabinoids (Dietrich 2004; Cordeiro *et al.* 2017), though serotonin has been implicated in GWAS of hyperactivity disorders (Aebi *et al.* 2016).

Other human studies have detected QTL associated with physical traits related to exercise abilities, including maximal oxygen consumption (VO_{2max}) (Williams *et al.* 2017), bone density (Herbert *et al.* 2019), and more (Lin *et al.* 2017). The list of possible biological traits affiliated with exercise and their associated QTL is extensive (Sarzynski

et al. 2016; Lightfoot *et al.* 2018).

Observational studies of human exercise behavior are limited by measurement error and environmental cofactors that cannot always be accounted for in statistical models (Garland, Jr. *et al.* 2011b; Lightfoot *et al.* 2018). An alternative way to study this is to use animal models in selective breeding experiments. Use of animal models in selective breeding experiments (Garland, Jr. and Rose 2009) can alter proportions of alleles that affect a trait of interest, thus allowing for easier detection of such alleles (Britton and Koch 2001; Konczal *et al.* 2016).

To elucidate the biological basis of voluntary aerobic exercise behavior, a selection experiment was begun in 1993 using a base population of outbred Hsd:ICR mice. Four replicate lines have been bred for high voluntary wheel-running behavior and another four bred without regard to their wheel running as controls for founder effects and random genetic drift (Swallow *et al.* 1998). Since the beginning of this experiment, over 150 papers have been published that document a variety of phenotypic differences between the High Runner (HR) and Control (C) lines. These previous studies establish morphological and physiological differences in bone, kidney, heart, skeletal muscle, brain, and other organs and systems (Rhodes *et al.* 2005; Swallow *et al.* 2005; Kolb *et al.* 2013b; Wallace and Garland, Jr. 2016) and, more generally, reaffirm the diversity of the systems involved in voluntary exercise behavior (Garland, Jr. *et al.* 2011b; Lightfoot *et al.* 2018). The previous studies also give potential directions for informed analyses of the genome. For example, we would expect divergence in allele frequencies related to the reward system in the brain and to muscle function. The HR selection experiment is the

world's "largest" involving a behavioral trait in rodents in terms of the number of lines and generations. Therefore, addressing the genomic differences between the HR and C mice is expected to provide novel insights into the underpinnings of exercise behavior.

Previously, Xu and Garland (2017) used a mixed model (nested ANOVA) with minimum variance quadratic unbiased estimation (MIVQUE) to analyze medium-density single nucleotide polymorphism (SNP) data for the HR and control lines sampled from generation 61 (Xu and Garland 2017). This statistical method proved more powerful than the commonly used regularized F test and Generalized Linear Mixed Model (GLMM) methods when incorporating permutation-based multiple testing correction. The data used included 7-10 females from each of eight lines (four HR and four C). Genotypes were determined with the MegaMUGA SNP-chip (Morgan and Welsh 2015). After removing markers with missing data, 25,318 markers were analyzed with the mixed models, finding 152 markers to be significantly differentiated between the HR and C linetypes (i.e. test group). Although Xu and Garland (2017) demonstrated numerous differentiated SNP loci between the HR and control lines, biological interpretations were not presented. Additionally, as demonstrated by the whole-genome sequence (WGS) data addressed in this paper, various differentiated loci were not detected in the previous SNP-chip analysis.

Here, we apply the mixed model with MIVQUE estimation method to WGS data obtained from the same individuals as in Xu and Garland (2017). We analyze both SNP and haplotype data to take full advantage of the information provided by each data type (Shim *et al.* 2009; Taliun *et al.* 2016). We also use simulations to explore some of the

statistical properties of the MIVQUE estimation method for this application, and we

implement procedures aimed at improving model fit and potentially statistical power.

We detect numerous differentiated SNP and haplotype loci between the HR and C lines.

Many of these can be tied to specific lower-level traits that should influence exercise

behavior, through use of gene ontology terms and KO phenotype analyses of nearby

genes.

Using information on known morphological and physiological differences

between the HR and control lines, we were able to perform both broad and directed

strategies to detecting significantly differentiated loci. We show that the method of Xu

and Garland (2017) can be improved by allowing for different among- and within-line

variance structures. We identified several differentiated genes associated with bone,

heart, and brain morphology. We also identified a few candidates with potential large-

scale influences on the HR mice, including *Sorl1, Dach1*, and *Cdh10*.

MATERIALS AND METHODS

High Runner Mouse Model

As described previously (Swallow *et al.* 1998; Careau *et al.* 2013), 112 males and 112 females of the outbred Hsd:ICR strain were purchased from Harlan Sprague Dawley in 1993. These mice were randomly bred in our laboratory for 2 generations. Ten males and 10 females were then randomly chosen as founders for each of 8 closed lines (generation 0). Four of these lines were randomly picked to be "High Runner" (HR) lines, in which mice would be selected for breeding based on voluntary wheel running. The remaining 4 lines were used as Control (C) lines, without any selection. At approximately 6-8 weeks of age, all mice were given access to wheels for six days. The amount of running (total revolutions) on days 5 and 6 was used as the selection criterion. For the non-selected C lines, one male and one female from each of 10 families were chosen as breeders to propagate the line. For the HR lines, the highest-running male and female from within each of 10 families were chosen as breeders (within-family selection). Sib-mating was disallowed in all lines (Swallow *et al.* 1998).

Whole-genome Sequencing

80 xx ~male mice (10 from each line), from generation 61, were subject to whole genome sequencing 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 12X per mouse. SNPs were filtered based on genotype quality ("GQ") >5, read depth >3, MAF < 0.0126 for all samples, and Mapping Quality ("MQ") > 30. One of the 80 mice

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 two of the 80 mice were also removed from analysis.

Heterozygosity Calculations

Individual mouse heterozygosity (multi-locus heterozygosity) was calculated by dividing the number of heterozygous loci for each mouse by the total number of segregating loci across all 80 mice (n=5,932,124). Heterozygosity per line is the average of the heterozygosity of all sequenced mice within that line.

SNP Analysis

Individual Single Nucleotide Polymorphisms (SNPs) were initially analyzed using a mixed model approach with the Minimum Variance Quadratic Unbiased Estimation of variance (MIVQUE) method of estimating variance parameters as described in Xu and Garland (2017). However, rather than removing loci or mice with missing data, code was modified to remove only the missing values themselves. The MIVQUE analysis provides a p-value for each locus for rejecting the null hypothesis of no differentiation between the HR and C lines. Xu and Garland had performed the analysis using two different encoding schemes to represent genotypes as 0, 0.5 and 1 vs. as twin vectors of 0-0, 0-1 and 1-1. We have since determined that the twin vectors encoding was preferable, and we report only those results (File S1.7).

Multi-Model Analysis of SNP Data from Whole-genome Sequences

The analyses performed in Xu and Garland (2017) used a single statistical model in R for all loci (our comparable SAS model being "Simple" in Table 1.1). This model did not allow for several possibilities that might be expected a priori and that were in fact observed, such as differing variances among the 4 replicate HR and C lines (designated "SepVarLines" in Table 1.1), as is the case for wheel-running behavior (Garland, Jr. *et al.* 2011a). Beyond this, the amount of variation among individual mice within the replicate lines might differ for the HR and C lines ("Full" model). Interpretation of these different models is presented in the Discussion. In total, we applied four alternate models to the data for each locus, and followed a model selection procedure for the one with the lowest the Aikake Information Criterion, corrected for small sample sizes (AICc), and retained the p-value for its linetype effect (differentiation between the HR and C lines). All Multi-Model analyses were performed in SAS using PROCEDURE MIXED with the mivque0 method (File S1.10). We elected to prioritize SAS over R for its performance gains over large number of loci. For a direct comparison, we reanalyzed the MegaMUGA data in Xu and Garland (2017) the multi-model method (Figures S1.1 and S1.2).

Loci that contained no within-line variance (i.e. each line was fixed for one allele or the other) could not be analyzed with the foregoing procedures. We analyzed these loci by counting the net number of alternatively fixed lines among the HR and C linetypes. Those loci with greater difference in allele frequency between the HR and C linetypes are regarded as being more "significant."

Multiple Testing Correction

Permutations for MegaMUGA Data

This approach is based on the permutation method used by Xu and Garland (2017), but modified to account for the multiple models. All permutations were performed using SAS PROC MIXED as described above in the section on multi-model approach. The mouse IDs, line, and linetype were randomly permuted as a block to break their original associations with the allelic data but not with each other. The permuted data for each locus were then analyzed with each of the four models listed in Table 1.1 (i.e., for the MegaMUGA SNP data, 4 X 25,332 analyses were performed). For each of the four models, the AICc was recorded, and the corresponding F-statistics were retained. From these 25,332 loci (for the MegaMUGA data), the F-statistic corresponding to the model with the lowest AICc was saved. The foregoing process was repeated 5,000 times, the resulting F-statistics were sorted from largest to smallest, and the 250th largest F-statistic was used to establish the critical value for the 5% FWER.

Permutations for Haplotype Data

Permutations done for haplotypes were performed separately for 2-allele haplotype blocks and 3-allele blocks, using 1,000 permutations to keep computational times manageable. As in the unpermuted haplotype analyses, blocks with three alleles (n=5,869) were analyzed with two dummy variables, each individual dummy variable was tested using the multi-model method, and the two p-values generated were combined using Fisher's method (Fisher 1925). However, some permutations of the 3-allele blocks

produced erroneous low p-values (apparently due to numerical issues), which, if included

in subsequent calculations would have caused an artifactual reduction of the critical value

needed to obtain the true 5% FWER. The permutations of the 2-allele blocks (n=11,032)

did not produce any artifactually low p-values. Given the problems with the 3-allele

haplotype permutations, we elected to apply the MeguMUGA permutation threshold

(P<0.00526) to the haplotype blocks because of their similar sample size

(MegaMUGA=25,332; Haplotypes=16,901) and the fact that they should be highly

correlated.

Local Maxima Selection for WGS Data

In the original paper, which analyzed 25,332 SNPs from a commercial chip, a

permutation procedure was used to control the family-wise Type I error rate (FWER) at

5% (Xu and Garland 2017). Those procedures were not computationally practical for the

5,932,124 SNPs from the whole-genome sequences, nor are linked SNPs within a

haplotype block truly independent from each other. Accordingly, significant loci were

chosen via a combination of -logP cutoff and local maximum (LM) determination, the

latter acting as a filter to focus on actual selected loci over their hitchhikers. Similar

methods have been previously described (Nicod *et al.* 2016). Briefly, suggestive loci

with -logP >3.0 were clustered with a maximum gap of 1 Mbp. For each such cluster, the

global peak, and a set of local maxima were determined for every 500 kbp spanned by the

cluster. The set of local maxima were chosen as peaks separated by dips in the signal

below the median -logP in the cluster. These LM SNPs were annotated using R libraries

GenomicFeatures and VariantAnnotation, with the mm10 knownGene.sqlite database provided by the Genome Browser team at the University of California, Santa Cruz.

Haplotype Determination

From the whole-genome sequences, haplotypes were determined using JMP 11 and JMP Scripting Language (SAS Institute Inc., Cary, NC). To construct haplotypes, we first defined the genomic block segments as consecutive 20 kbp windows that did not transition between homozygous and heterozygous states. For each block region, we performed a hierarchical clustering analysis using SNP genotype data (of homozygous regions only) as input. Preliminary haplotype analysis showed that the HR population at generation 61 rarely had more than 3 alleles in a given haplotype. Therefore, the analysis was restricted to a maximum of three clusters (haplotype alleles) per block (File S1.5).

Haplotype Analysis

As for the SNP data, haplotype data were analyzed using the multi-model method described above. Haplotype blocks with only two alleles (n=11,032) were analyzed the same way as for the SNP data (File S1.10). Blocks with three alleles (n=5,869) were analyzed with two dummy variables, with the base allele chosen as the most common one, and then two dummy variables coding for presence of the other two alleles. Each individual dummy variable was tested using the multi-model method. The two p-values generated from the two dummy variables were combined using Fisher's method

(Fisher 1925). Different models potentially were used for each dummy variable based on AICc, allowing for up to two models to contribute to the final p-value of a locus (File S1.6).

SNPs Fixed in One Treatment but Polymorphic in the Other

As noted previously with the SNP chip data (Xu and Garland 2017), we observed no loci that were fixed for one allele in all four HR lines while being fixed for the alternate allele in all four C lines (see Results). We did, however, observe loci fixed for a given allele in all 4 HR lines, which is symptomatic of a complete selective sweep (caused by directional selection) as described by Burke (2012), while remaining polymorphic in all 4 C lines. All loci that were fixed in the HR mice and simultaneously polymorphic in all C lines (FixedHR/PolyC) were extracted from the multi-model results and grouped such that those fixed loci that were within 100,000bp of other fixed loci would be part of the same group. This process was then repeated for loci fixed in the Control lines but polymorphic in all HR lines (FixedC/PolyHR).

General Ontology Analysis

Transcribed regions (N = 56, as indicated in Table 1.2) found to contain LM based on the whole-genome sequence analyses were analyzed using The Gene Ontology Resource (GO). GO analyses were performed based on biological process, molecular function, and cellular component. Ontologies reported as significant at raw $P < 0.05$ for any of these three categories are reported here. Analysis of these genes was also performed using the

Database for Annotation, Visualization and Integrated Discovery (DAVID). The results of these analyses did not vary greatly from the GO results.

Targeted Ontology Analysis

Previous papers show that the HR lines of mice have diverged from the C lines for many different phenotypes (reviews in Rhodes *et al.* 2005; Garland, Jr. *et al.* 2011b; Wallace and Garland, Jr. 2016). Many of these phenotypes can be tied to specific neurobiological or physiological functions. In such cases, a logical approach is to analyze separately some candidate genes known to be affiliated with relevant functions and find differentiated SNPs for those genes. We used this approach for several ontologies. Specifically, lists of genes affiliated with dopamine, serotonin, brain, bone, cardiac muscle, and skeletal muscle were extracted from the Mouse Genome Informatics website. SNPs found within these genes were separated from the full WGS data and the most differentiated among these were recorded.

Data Availability Statement

Any additional intermediary or results file are available upon request. Supplemental files are available at FigShare (note that all filenames on FigShare will exlcude "1." from the numbering). File S1.1 contains supplemental figures and brief descriptions of all other supplemental files and tables. File S1.2 contains allelic SNP data. File S1.3 contains mouse data with line and lintype. File S1.4 contains all results for analyses of individual SNPs. File S1.5 contains all haplotype data. Files S1.6 contains all results for analyses

of haplotype data. File S1.7 contains justification for use of allelic coding of alleles. File

S1.8 includes simulations of Type I error rates for Mixed Model analyses using MIVQUE

variance estimation. File S1.9 expands on the discussion of genes in consistent regions

(see Results). File S1.10 includes all R and SAS code used for the SNP and haplotype

analyses. Table S1.1 includes local maxima associated genes. Table S1.2 contains

groups of loci fixed in all lines of one lintype but polymorphic in all lines of the other.

Table S1.3 includes heterozygosity for each individual mouse. Table S1.4 includes top

ten genes for each of the targeted ontologies analyses. Table S1.5 includes allele

frequency by line of each loci identified as a local maximum. Table S1.6 includes

genomic regions identified as suggestive ($p<0.001$) by the SNP analyses.

RESULTS

Variation in Genetic Diversity

After 61 generations of the High Runner mouse selection experiment, and based on a sample of 79 mice, we found SNPs segregating at 5,932,124 loci (~2.2 SNPs per kbp or 0.22%) across the entire set of lines (i.e., at least 2 mice containing an alternate allele were found across the 79 mice sequenced) with at least 1.5% minor allele frequency. Individual lines contained 2.04 – 2.82M SNPs (34–48% of the total diversity) (Table 1.3), with no appreciable loss in diversity for the HR lines compared to the Control replicates (Mann-Whitney U-test, W=6; p-value=0.6857). SNP heterozygosity ranged from 10.3% to 20.6% among individual mice (Table S1.3) and averaged 12.7% to 18.1% per line (Table 1.3).

Initial haplotype analysis demonstrated that there were rarely more than three alleles for any given haplotype block (region with little to no discernable recombination events within the 79 mice analyzed). Therefore, for the final haplotype analysis, hierarchical clustering was performed with a limit of 3 clusters. 16,901 of these blocks remained variable across the 8 lines in generation 61. As would be expected, the number of haplotypes that have not gone to fixation in each line appears to be proportional to the number of SNPs that have not gone to fixation (Table 1.3). Heterozygosity for the haplotypes ranged from 12.2% to 25.5% for individual mice (Table S1.3), and 14.7% to 19.6% when averaged per line (Table 1.3). Heterozygosity for the haplotype data were not significantly different between HR and C lines (Mann-Whitney U-test, W=8; p-value=1.0 and W=6; p-value=0.6857, respectively).

Multi-Model vs Single-Model Comparisons

As expected, we found that many, indeed most, loci were better fit by models other than the "Simple" model used by Xu and Garland (2017). Generally, the "Full" model was the most preferred, followed by the "Simple" model (Table 1.4). In general, differences between the p-values determined by the single and multi-model methods were negligible (Figure S1.2).

When analyzing data generated under the null hypothesis, the mixed models with MIVQUE estimation for both single and multi-model produced a deflated Type I error rate for $\alpha = 0.05$ (File S1.8). The multi-model approach helped to correct this, but the Type I error rate did not improve greatly with the multi-model approach alone. We attempted to utilize the Kenward Rogers method of determining degrees of freedom to correct this low Type I error rate, but this did not bring Type I error rate to 0.05 and effectively dropped the nested line effect for many loci. We did not want to drop the nested line effect because this ignores the fundamental experimental design of the selection experiment. However, the permutation and local maxima methods of determining loci of interest are robust to this deflated Type I error rate (File S1.8), so we proceeded with our analyses using conservative results produced by the MIVQUE variance estimation method.

Three Major Analyses

Whole-Genome Haplotype

No haplotypes were identified as being fixed in all HR lines for one allele and fixed in all

C lines for the opposite allele. The multi-model haplotype analysis produced 102 blocks

of significant differentiation at the $p<0.005$ (permutations) level. Significant blocks

could be found on 13 chromosomes (Figure 1.1). We consider haplotype blocks within

1,000,000 bp of each other to be linked and therefore part of the same haplotype group:

28 such groups were determined (Table 1.5). These groups include a total of 154

transcribed sequences recognized by the Panther database for gene ontology. The largest

of these groups was found on chromosome 14:52,100,155-54,334,868 bp (Table 1.5).

Whole-Genome SNP

Similarly to haplotypes, no individual SNPs were identified as being fixed in alternative

alleles across all HR on one hand and all C lines on the other. At the $p<8.4E-09$ critical

level (Bonferroni-corrected), only two SNPs in chromosome 5 were identified to be

significantly differentiated across the entire genome (Figure 1.2), both in an intron of an

uncharacterized gene (GM34319). The syntenic/orthologous region of both the human

and cat genomes correspond to a coding region (exon 3) of the MYL5 gene (Myosin light

chain 5). Due to the small number of significant SNPs under Bonferroni and the

computational difficulties of using permutations with the multi-model method, we focus

on local maxima SNPs.

In the local maxima (LM) analyses, the suggestive cutoff ($-logP>3.0$) produced

38,065 SNPs for analysis. 44 clusters were found, ranging in size from 1 SNP to 3,787 SNPs (Chr9: 41,303,824-42,478,817 bp). The largest single group in terms of genome spanned is on chr17: 17,846,983-23,586,163 bp (Table 1.6). From these groups, a total of 84 LM were determined. 31 of these SNPs were associated with 27 unique transcribed regions. 26 of the 27 genes could be utilized for GO analysis. Although chromosome 3 had no LM fall into specific genes (despite clear significance based on the Manhattan plot), the cluster on chr3 (chr3:51,190,735-52,498,029 bp) includes about 10 validated coding genes and various predicted genes, but none of the LMs fall in these. However, all three LMs in this group are upstream of Setd7, a methyltransferase.

The most significant SNPs with no within-line variance fell into three regions. One of these regions is on chromosome 5 (105-109 mbp), which is close to the LM identified in this chromosome. Another is on chromosome 16 (44 mbp), about 2.5 million base pair from the LM on chromosome 16 containing Lsamp, a gene which codes for a neuron-associated membrane protein. However, the last region falls in chromosome 7 (115 mbp), a chromosome which contained no LM. This location is downstream of Sox6, a developmental regulator broadly associated with muscle fiber type composition (van Rooij *et al.* 2009), hematopoiesis, bone growth and heart function (Smits *et al.* 2001).

SNPs Fixed in One Treatment and Polymorphic in the Other

SNPs that were fixed in all HR lines and polymorphic in all C lines (FixedHR/PolyC) were combined into 95 regions, based on their being separated by at least 100kbp (Table

S1.2). Some of these regions are probably not independently segregating (i.e. chr17: 17,895,909-22,546,405 bp) and might therefore be combined further. Regions varied in size from 1 to 1,626,783 bp. These regions include or are proximal to (in the case of 1 bp regions) 135 transcribed regions, including genes, miRNA, and predicted genes. SNPs that were fixed in all C lines and polymorphic in all HR lines (FixedC/PolyHR) were combined into 64 regions. The size of each region varies from 1 to 753,066 bp. We expect the 1 bp loci may be spurious but chose to include them in results for completeness, especially given that the mini-muscle locus involves only a single base pair (Kelly *et al.* 2013). These regions include or are proximal to 63 transcribed regions, again including genes, miRNA, and predicted genes. FixedHR/PolyC regions were also identified in haplotypes. These haplotype blocks overlapped with the SNP regions identified by FixedHR/PolyC; however, some of the single unlinked loci that met these criteria were not identified using haplotypes.

Ontology Analyses

General Ontology

GO analysis of biological process for the haplotype data reveal "sensory perception of chemical stimulus" to be a major term of interest (Table 1.7). This appears to be caused by various clusters of olfactory and vomeronasal genes. Many of the most prominent terms (highlighted in green) appear to be correlated to these olfactory and vomeronasal gene clusters. Although a single, large group of closely linked olfactory genes may overrepresent olfactory's role in selection, we were able to identify two distinct genomic

regions of vomeronasal genes and three such regions of olfactory genes.

The biological process GO terms for LM include many results that are consistent with our previous findings involving the HR mice, including cardiac and myoblast related terms (Table 1.8). Regulation of locomotion is among the most statistically significant GO terms.

The FixedHR/PolyC GO analyses indicate terms: complement receptor mediated signaling pathway and response to pheromone. These terms were significant with a false discovery rate correction (FDR<0.05), p=7.11E-04 and p=2.40E-07, respectively) (Table 1.9). For FixedC/PolyHR, no GO terms were significantly enriched with FDR correction, some novel GO terms were deemed most significant. Included in these results is also CDP-choline pathway, which had also been implicated in the haplotype data. The full list of regions for both FixedHR/PolyC and FixedC/PolyHR can be found in (Table S1.2).

Targeted Ontology

The gene search for specific ontologies produced 45-820 genes and 7,315-143,507 SNPs associated with each search (Table 1.10). The top ten genes were chosen based on the most significant SNP within the gene (Table S1.4). The most significantly differentiated SNPs were generally found in genes associated with the brain, followed by bone and muscle related genes. Surprisingly, the reward-related ontologies (dopamine and serotonin) did not contain as strong evidence for differentiation as the others.

Consistent Regions Identified Across Multiple Analyses

The major analyses (LM, haplotype, and FixedHR/PolyC) individually implicate about 80, 24, and 46 differentiated genomic regions, respectively. Combined, 85 unique regions across the genome are indicated, including at least one region on every chromosome. Of these 85 regions, 12 are found in all three analyses (Table 1.11). These 12 consistent regions span just over 27.4 MBP and include 300 validated and predicted genes. Of the 300 genes, 77 are either olfactory or vomeronasal genes, which are predominantly located in two large regions on chromosomes 14 and 17. Surprisingly, many of these regions do not contain many of the most differentiated SNPs according to the multi-model MIVQUE analyses, but do have at least one SNP with $p \leq 0.001$ by the LM criteria.

DISCUSSION

Variation in Genetic Diversity

For the present sample of 79 mice from generation 61, based on the polymorphic SNPs within each line (Table 1.2), each of the lines continues to retain approximately 34-48% of the total diversity across all 8 lines. Such a drop in genetic diversity would be expected after 61 generation with ~10 breeding pairs per generation per each line. We found no evidence that HR and C lines had differing levels of genetic diversity, averaged across the whole genome.

Consistent Regions from Multiple Analyses

Many of the identified regions span too many genes to allow ready identification of a candidate. However, a few of the regions contain a limited number of genes for which the reported functions make sense in the context of directional selection for high voluntary wheel-running behavior (from first principles of physiology and neurobiology) and/or given previously identified differences between the HR and C lines (see Introduction). Given the rich phenotyping literature on the HR mouse selection experiment (more than 150 publications), we discuss a relatively large number of genes. Additional regions are covered in supplemental material (File S1.9).

The region identified on chromosome 5 includes 16 genes (excluding predicted and non-coding), three of which were previously identified as differentially expressed in the striatum of the HR and C mice (Saul *et al.* 2017). These genes include *Tmed5, Gak,* and *Mfsd7a*. *Tmed5* is a trafficking protein associated with cell proliferation and

WNT7B expression in HeLa cells (Yang *et al.* 2019). Mice knockouts in Gak are generally lethal to adult and developing mice causing various abnormal symptoms, including altered brain development (Lee *et al.* 2008). *Mfsd7a* (aka *Slc49a3*) has been associated with ovarian cancer, but much remains unknown about this gene (Khan and Quigley 2013).

The region on chromosome 6 includes *Trpv5* and *Kel*, both of which are associated with KO phenotypes that may be tied to known differences between the HR and C lines. Trpv5 KO is associated with phenotypes related to structural changes in the femur and kidney physiology (Hoenderop *et al.* 2003; Loh *et al.* 2013), both of which differ between HR and C lines (Swallow *et al.* 2005; Castro and Garland, Jr. 2018). *Trpv5* is also associated with calcium homeostasis (Hoenderop *et al.* 2003; Loh *et al.* 2013). Kel is a blood group antigen with KO phenotypes affiliated with weakness, gait and motor coordination, neurological development, and heart function (Zhu *et al.* 2009, 2014). Previous experiments have shown the HR and C mice to have differences in heart physiology (Kolb *et al.* 2013a), gait and motor coordination (Claghorn *et al.* 2017), and brain development (Kolb *et al.* 2013b).

The region on chromosome 9 contains various predicted genes and miRNA, but also one large gene of interest, *Sorl1* (aka *SorlA*). This gene is also implicated in our targeted search for genes related to the brain (Table 1.10). *Sorl1* codes for a sorting receptor that has been associated with various neural and metabolic diseases (Schmidt *et al.* 2017). Although some of the associated phenotypes, such as obesity, may have some correlation to phenotypic differences between HR and C mice, such as difference in body

fat (Swallow *et al.* 2001; Vaanholt *et al.* 2008; Hiramatsu and Garland, Jr. 2018), this does not directly answer the question of how *Sorl1* influences running behavior. Mouse knockouts in this gene have not shown changes in running gait (Rohe 2008), whereas differences in gait do exist between HR and C mice (Claghorn *et al.* 2017). However, these treadmill tests do not address exercise motivation, which might be influenced by such a neurobiologically relevant gene. Additionally, a more significantly differentiated haplotype can be found over 150,000bp downstream of *Sorl1*, containing various predicted genes and miRNA. Therefore, further studies will be required to determine precisely the elements of this region that modulate wheel running. Although *Tbcel* is near this consistent region rather than included in it, it is the most differentiated gene in the genome (based on median p-value of included SNPs, p= 4.01E-07). This gene is known to regulate tubulin activity in sperm and the nervous system (Nuwal *et al.* 2012; Frédéric *et al.* 2013).

One region on chromosome 11 contains numerous genes of potential interest. One LM within this region is proximal to a handful of genes that may be influencing the HR phenotype, including: *Tefm*, *Adap2*, *Crlf3*, and *Suz12*. These genes are associated with KO phenotypes including enlarged heart and decreased body weight (Jiang *et al.* 2019), blood cell concentration (White *et al.* 2013), and brain morphology (Miro *et al.* 2009). All of these phenotypes have been found to differ between HR and C mice (Kolb *et al.* 2013b; Thompson 2017; Singleton and Garland, Jr. 2019).

One region on chromosome 14 includes almost exclusively *Dach1*, which is an important regulator for various early developmental genes. *Dach1* is a regulator of

muscle satellite cell proliferation and differentiation (Pallafacchina *et al.* 2010).

Although knockouts of *Dach1* in mice do not appear to disrupt limb development (Davis *et al.* 2001), *Dach1* mutants sometimes have stunted leg development in *Drosophila* (Mardon *et al.* 1994). Furthermore, *Dach1* has been shown to localize around limb budding regions and interact with known limb patterning genes in both mice and poultry (Horner *et al.* 2002; Kida 2004; Salsi *et al.* 2008). Studies of skeletal muscle (Garland, Jr. *et al.* 2002; Bilodeau *et al.* 2009) and of the peripheral skeleton show several differences between HR and C lines of mice (Garland, Jr. and Freeman 2005; Kelly *et al.* 2006; Castro and Garland, Jr. 2018; Schwartz *et al.* 2018). This gene has also been implicated in the development and function of the kidneys (Köttgen *et al.* 2010), which have been shown to be larger in the HR lines than C lines in some studies (Swallow *et al.* 2005).

A region on chromosome 15 includes *Cdh10* among a few predicted genes. GO links *Cdh10* to both "calcium ion binding" and "glutamatergic synapse," terms that occasionally produced suggestive p-values for enrichment searches in our differentiation analyses (Table 1.7, Table 1.9). These terms could have various implications for the HR mice. *Cdh10* specifically is a cadherin with extensive expression in the brain (Liu *et al.* 2006; Matsunaga *et al.* 2015). This gene has been shown to have increased expression in phrenic neurons (Machado *et al.* 2014), potentially modulating diaphragm movement, and increased functionality of the diaphragm could partly underlie the elevated maximal rate of oxygen consumption during exercise (VO$_2$max) observed in HR lines (Kolb *et al.* 2010; Hiramatsu *et al.* 2017; Singleton and Garland, Jr. 2019). *Cdh10* is also known to

have increased expression of genes associated with olfactory system development (Akins *et al.* 2007), which could be corroborated by the other two consistent regions associated with olfactory and vomeronasal (see Results, General Ontology). The other region detected on chromosome 15 currently only contains *Fam135b* among its annotations. Few studies have been conducted involving the function of *Fam135b*, but evidence indicates it has an important role in spinal motor neurons based on a > 10,000-fold decrease in expression in spinal and bulbar muscular atrophy models (Sheila *et al.* 2019).

Ontology

General Ontology

The GO analyses in this paper serve two functions. The first includes determining pathways that have been influenced by the selective breeding protocol. Additionally, the vast publications and data on various morphological and physiological differences between the HR and C lines provide insight into differentiated biological processes.

The Haplotype and Fixed/Poly methods of identifying differentiated genes had considerable overlap between genes and regions identified, which seems to result in similar GO terms for these analyses. The term "sensory perception of chemical stimulus" is expected, given the large number olfactory and vomeronasal genes present in some of these regions. Selection for such genes is likely in response to how the mice are tested for wheel running. For logistical reasons, approximately 2/3 of the mice tested in a given generation were measured on wheels that had not been washed since the previous mouse was on that same wheel, although the attached cages were fresh (Dewan *et al.* 2019).

The scent of the previous mouse would potentially elicit different running behavior, dependent on these vomeronasal and olfactory genes (e.g., see Drickamer and Evans 1996). GO terms related to postsynaptic neurotransmitters were largely incited by three genes (*Cplx1*, *Dlg1*, and *Shisa6*). Such terms would be expected due to observations of the HR mice having larger brain and altered reward mechanisms (Belke and Garland, Jr. 2007; Mathes *et al.* 2010; Garland, Jr. *et al.* 2011b; Keeney *et al.* 2012; Kolb *et al.* 2013b; Thompson *et al.* 2017).

The local maxima GO results are generally quite different from the haplotype and Fixed/Poly analyses. This is partially attributable to less overlapping of identified genomic regions. Additionally, LM is useful for gene culling to reduce influence of hitchhiking genes in the GO analyses. Many of the top terms for LM genes are associated with heart development and function. Heart ventricle mass is greater in the HR mice (Kolb *et al.* 2013a; Kelly *et al.* 2017; Kay *et al.* 2019) and correlates with VO$_2$max in both HR and C mice (Rezende *et al.* 2006). The genes most associated with cardiac development include *Pkp2*, *Myh11*, and *Tbx5* (also a forelimb regulator). Forelimb development may be altered in the HR mice, while humerus sizes do not seem to differ (Copes *et al.* 2018), differences have been found in metatarsal and metacarpal lengths (Young *et al.* 2009).

Targeted Ontology

As the target ontologies were chosen based on structures and systems known to have been altered by the selection experiment, we would expect to find at least one gene of

each ontology that would contain a differentiated SNP. Of these ontologies, "serotonin" and "dopamine" are associated with some of our less impressive p-values (Table 1.10), with many of the top dopamine-related genes (*Fpr1*, *Fpr2*, *Fpr3*, and *Fpr-rs4*) being present likely because of linkage to highly differentiated vomeronasal genes (Table 1.10). The most significantly differentiated loci in a dopamine-related gene are in *Gnb1*, part of the Gβγ complex, which activates Girk2 in dopamine neuron membranes (Wang *et al.* 2016). We are surprised not to have found more impressive results for dopamine-related genes, given clear differences in dopamine function between the HR and C mice (Rhodes *et al.* 2001, 2005; Rhodes and Garland, Jr. 2003; Bronikowski *et al.* 2004; Mathes *et al.* 2010). A possible explanation for is that trans-regulating sites for these genes have been more influenced by the HR selection regime (Kelly *et al.* 2012; Nica and Dermitzakis 2013).

The remaining ontologies (bone, cardiac, skeletal muscle, and brain) all have at least one gene containing a SNP with p <0.0001 (Table 1.10). Some of these are included with our LM genes, such as *Myh11* (a myosin gene affiliated with the "cardiac" tag) and *Sorl1* ("Brain" tag). However, some of these are not present among the LM list. Kel, described above as influencing various phenotypes relevant for high running behavior, may appear to be a confusing "miss" for the LM detection process, with a p-value = 1.49E-05. However, the region does have two local maxima, neither of which land in genes, but one is about 15,000 bases upstream of *Kel*. This might be taken as evidence that the LM approach to determining affected genes ought to be modified to better catch nearby genes that could be affected.

Comparison with Previous Studies

Exercise behavior and the QTL that help to modulate it have been studied in various other GWA and gene expression studies of mice, as well as comparisons of inbred strains. Below are some examples of the listed in chronological order.

A study involving a cross between high- and low-running inbred strains located several markers on chromosome 9 (Lightfoot *et al.* 2008). Although none of these markers correspond to our significant region (about 41,000,000 to 42,000,000), one of them is only about 500,000bp from the gene *Leo1*. However, with only one significant locus (raw p-value = 0.00186) in the region, evidence for this gene being a modulator of exercise in the HR mice is not strong.

The findings of Lightfoot et al. (2010), which identified loci associated with wheel running levels among 38 inbred strains of mice, suggested very few QTL similar to our findings. The best example of similarity includes a region on chromosome 8 that includes *Galntl6*, which was found as suggestive in the current study. Additionally, Lightfoot et al. (2010) identify a region on chromosome 12 very close to *Nrxn3*. Both LM and FixedHR/PolyC methods indicated this gene as a strong candidate. This was not listed as a consistent region because the haplotype results did not produce a significant haplotype near *Nrxn3*. *Nrxn3* creates particular interest in that it is associated with various addictive behaviors (Zheng *et al.* 2018), which is consistent with evidence that the HR mice are to some extent addicted to running (Rhodes *et al.* 2005; Kolb *et al.* 2013b). Exercise addiction is not a new concept, but remains controversial (Nogueira *et*

al. 2018).

QTL mapping of the G_4 intercross of C57BL/6J with one of the four HR lines of mice from the present selection experiment paper produced a region implicating olfactory/vomeronasal influence on chromosome 7 (Kelly *et al.* 2010). The current study also identifies vomeronasal (though different from our region on Chr 17).

Muscle and brain gene expression studies have been utilized to better understand the molecular basis of exercise. Using the same G_4 intercross mice as previously described (Kelly *et al.* 2010), Kelly et al. (2012, 2014) identified and highlighted *Insig2* (brain and muscle), *Socs2* (brain), *Dby* (Brain), *Arrdc4* (Brain), *Prcp* (brain and muscle), *Il15* (brain), and *Sparc* (muscle). However, none of those genes were determined by the present study to be local to differentiated SNPs.

Dawes et al. (2014) found *Actn2*, *Actn3*, *Casq1*, *Drd2*, *Lepr*, *Mc4r*, *Mstn*, *Papss2*, and *Glut4* to have differential expression in skeletal muscle and brain tissue based on a comparison of two inbred strains, one high- and the other low-wheel running (C57L/J and C3H/HeJ, respectively). None of these genes were found by the present study to contain significant SNPs. However, *Drd2* is about 8 Mbp from of one of the most differentiated regions of the genome (on chromosome 9).

Saul et al. (2017) performed expression analysis using the striatum of the HR and C lines from generation 66. Some of the highlighted differentially expressed genes include: *Htr1b*, *Slc38a2*, *Tmed5*, *5031434O11Rik*, *Gak*, *Mfsd7a*, and *Gpr3*. *Tmed5*, *Gak*, and *Mfsd7a* are all found within a highly differentiated region in the SNP data (median p=4.85E-04 for all three genes). Although *5031434O11Rik* and the associated *Setd7* are

not found within the consistent regions, they both contain many of the most differentiated loci of individual SNP analyses (median p=3.78E-05). Knockouts of *Setd7* (aka *Set9*) have been associated with altered lung development and morphology (Elkouris *et al.* 2016). Lung differences in the HR and C lines have not been greatly explored. Three studies have reported no statistical difference in lung mass (Meek *et al.* 2009; Kolb *et al.* 2010; Dlugosz *et al.* 2013), but an unpublished study of males from generation 21 found that HR lines tended to have higher pulmonary diffusion capacity and capillary surface area determined via morphometry (T. Garland, Jr., and S. F. Perry, personal communication) and a study of females from generation 37 reported a trend for HR mice to have higher dry lung mass (Meek *et al.* 2009; Kelly *et al.* 2017).

Overall, previous studies of mouse wheel-running behavior mostly disagree with the current study results. Studies involving HR and C57BL/6J intercross (Kelly *et al.* 2010, 2012) did not find many similarities. This may be expected, considering the loci whose frequencies have changed consistently across all four HR lines would not necessarily be expected to correspond with those that affect wheel running in a population derived from crossing one HR line with a distantly related inbred strain, nor with those that differ between two other inbred strains, neither of which has a history of experiencing selection for activity levels (Dawes *et al.* 2014). Similarly, few agreements exist between the present study and Lightfoot et al's (2008) findings. This is likely because comparing "High Running" alleles to control alleles is not the same as comparing these HR alleles to "Low Running" alleles, as a low-running phenotype may be correlated to dysfunctions in running ability or motivation, or that are not directly

associated with either motivation or ability. The greatest congruence is between the

present study and Saul et al. (2017). This agreement is encouraging because that study

compared the HR and C lines at a similar generation (66) to the present study (61).

Mini-Muscle Allele

The mini-muscle phenotype was discovered in the HR selection experiment and is

associated with alterations in various organs, especially skeletal muscle, but also

including heart, kidney, and overall body mass of the mice (Swallow *et al.* 2005; Meek *et

al.* 2009; Kolb *et al.* 2013a; Talmadge *et al.* 2014; Kay *et al.* 2019) as well as behaviors

(Kelly *et al.* 2006; Singleton and Garland, Jr. 2019). This phenotype is caused by a

single recessive SNP mutation located in an *Myh4* (myosin heavy polypeptide 4) gene

(Kelly *et al.* 2013). Mice expressing the mini-muscle phenotype have often been found

to run faster and sometimes for longer distances than other HR mice (Kolb *et al.* 2013a).

This polymorphism was lost, presumably via random genetic drift, from all lines except

for HR lines 3 (where it went to fixation) and line 6 (where it remains polymorphic with

the wildtype allele). Population-genetic analyses indicate that the allele was under

positive selection in the HR lines (Garland, Jr. *et al.* 2002). The current WGS data show

(generation 61) that the mutation is still only present in lines 3 (fixed) and 6, with allele

frequency of 0.65 in line 6. As the mini-muscle phenotype appears to enable faster

overall running on wheels at the cost of running duration, it has been regarded as an

alternative "solution" to the selection criterion (Garland, Jr. *et al.* 2011a), not unlike the

concept of "private" alleles (Martin *et al.* 1996). Such a mutation is expected to change

the genetic background of line 3 (and to a lesser extent, line 6) giving rationale to

analyzing these lines separately for possible QTL, in future studies.

Allele Frequency Implications

The general pattern of allele frequencies across the replicate lines can be used to infer

patterns of selection. Table 1.12 includes some of the potential profiles that could

possibly be observed and (for the most part) were observed in the WGS data.

Profile 1. No observed genetic variation. For our 79 mice, this accounts for about 99.8%

of the genome (Table 1.2).

Profile 2 Fixation for alternate alleles in the two selection treatments would imply

opposing directional selection, as might occur in experiments with replicate lines selected

for high versus low values of a trait. The HR mouse selection experiment includes high-

selected and control treatments, but not a low-selected treatment. Thus, fixation for

alternate alleles in the HR and C lines would not necessarily be expected, and indeed was

never observed for either the WGS data or the MegaMUGA data reported previously (Xu

and Garland 2017). Importantly, even data from selection experiments that include high-

and low-selected treatments are not showing much evidence of fixation for alternate

alleles (Burke *et al.* 2010; Lillie *et al.* 2019).

Profile 3. Stabilizing selection or random drift for one group and directional selection for

the other. This was the focus of the scans for loci fixed in all lines of one linetype and

polymorphic in all lines of the other (Fixed/Poly) in our own haplotype and WGS data

and produced several prospective regions of interest. The fixed allele can either be

entirely the reference (0) or alternative (1).

Profile 4. Selection for test group 2 but evidence of drift for group 1 (likely caused by little to no selection). Some of the loci of the WGS SNP data meet this profile. For example, Chromosome 11: 96,332,082 (p=0.051).

Profile 5. Random genetic drift for both test groups. Such loci will be among those analyzed, but this pattern of differentiation is unlikely to result from the selective breeding regimen.

In general, as with any population that is relatively well adapted to the prevailing environmental conditions, breeding colonies of laboratory house mice maintained under standard vivarium housing conditions should experience continuing stabilizing selection at many loci. Under standard housing conditions, an allele with a strong positive influence on wheel running, or activity in cages without wheels, might be disfavored if it were negatively associated with such aspects of the life history as litter size or maternal care. In contrast, under the conditions of the HR mouse selection experiment, an allele with a strong positive influence on wheel running might be expected to go to fixation rapidly in all HR lines in a manner consistent with a "complete sweep" (Burke 2012). Thus, to fix an allele, directional selection in the HR lines must be strong enough to overcome a presumed prevailing background of stabilizing selection and possibly negative selection. Regions that are FixedHR/PolyC (profile 3) should, therefore, be indicative of relatively strong directional selection in the HR lines.

Alternatively, some loci may have come under stabilizing selection in the HR lines, e.g., due to heterozygote advantage or epistatic interactions with other loci,

preventing them from going to fixation. Hence, we also examined loci polymorphic in all

HR lines but fixed in all C lines (FixedC/PolyHR). Surprisingly, many of these are

immediately adjacent to FixedHR/PolyC regions (Table S1.2). The GO analyses of the

included genes in these regions were consistently less significant (raw p≥0.0026 for all

implicated terms). However, such terms as "synapse assembly" and those related to

glycerolipids emerged may merit further exploration.

Interpretation of the Four Models

The four models in the multi-model analysis were included to allow for different variance

structures within and between the HR and C linetypes. The within-line variance is the

variability of allele frequency among the ~10 mice within each line. This variance is zero

when a line is fixed for one allele or another, but maximized when 5 mice within each

line are homozygous for one allele while 5 mice are homozygous for the other. The

among-line variance indicates how different the replicate lines within a linetype are from

each other. This variance component is minimized when all four lines within a linetype

are fixed for the same allele, but maximized when two lines are fixed for one allele while

two lines are fixed for the other.

In principle, both the within-line and among-line variances can differ between the

two selection treatments (linetypes); hence, the Full model includes separate estimates of

both within- and among-line variances. For wheel running in later generations of the

selection experiment, a full model has been shown to fit well (Garland, Jr. *et al.* 2011a).

The SepVarInd model includes only the within-line variance. The SepVarLine model

includes only the among-line variance. Lastly, the Simple model does not include either

of these two variances, and corresponds to the single model used by Xu and Garland

(2017).

As expected, we found many loci that were better fit by models other than the

Simple model used by Xu and Garland (2017) (Table 1.4). Figure 1.3 gives examples. In

A, the Full model is implemented because C lines exhibit very little within- and among-

line variance while HR lines exhibit both. In B, the SepVarInd model is used because C

lines have high within-line variance (while HR lines are comparatively low), but both

have similar among-line variance. In C, SepVarLines model is used because nearly all

lines contain very little within-line variance (6 are fixed for a single allele), but C lines,

being fixed for opposing alleles, creates different among-line variance. D identifies a

Simple model locus because these variances are roughly the same for the different

linetypes. E represents a locus with no within-line variance and thus could not be

analyzed with the mixed model ANOVA like other loci. However, use of multiple

models did not increase the number of loci identified as statistically significant based on

repeat analyses of the MEGAMuga data with both methods (Figure 1.1).

CONCLUSIONS

Exercise, or the lack of exercise, has far-reaching medical and financial implications (Manley 1996; Carlson *et al.* 2015). Numerous studies have provided strong evidence for the existence of genetic underpinnings of exercise behavior and physical activity (Kostrzewa and Kas 2014; Lightfoot *et al.* 2018), including in the High Runner mouse selection experiment (Careau *et al.* 2013; Saul *et al.* 2017; Xu and Garland 2017). Here we have used three different analytical methods with whole-genome sequence data to address the genetic basis of the 3-fold increase in running observed in the four replicate selectively bred HR lines of mice. These methods include haplotype and SNP statistical analysis, as well as non-statistical analysis of fixation patterns in HR and C lines.

The intersection of multiple analyses indicated 61 genomic regions of differentiation, with 12 regions identified as of particular interest. These regions include genes known to influence systems that have already been demonstrated to differ between HR and Control mice, such as response to conspecific odors, brain development, body weight, and relative heart size. However, they also contain genes whose role in voluntary running behavior is as yet unknown.

Importantly, none of the analytical approaches we used address the possibility of "private alleles" (Martin *et al.* 1996) in one or more of the HR lines that may influence exercise behavior, thus representing "multiple solutions" to the selective breeding regime (Garland, Jr. *et al.* 2011a), but this will be an important possibility to consider in future studies. We already know of one private allele of major effect (mini-muscle) that has far-reaching effects on mouse muscle and organ development (Swallow *et al.* 2005;

52

McGillivray *et al.* 2009; Kelly *et al.* 2013), as well as many other aspects of the phenotype, and has been favored by the selection protocol (Garland, Jr. *et al.* 2002). Determination of such alleles will be an important area for future research.

ACKNOWLEDGEMENTS

FUNDING

CONFLICT OF INTEREST

The authors have no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, D.A.H., L.Y., F.P.M.dEV., D.P., S.X., F.C., T.G.; investigation, D.A.H., L.Y., G.M.W., F.P.M.dEV., D.P., A.S.F., S.X., F.C., T.G.; software, D.A.H., L.Y., S.X.; formal analysis, D.A.H., L.Y., F.P.M.dEV., D.P., A.S.F., S.X., F.C., T.G.; writing – original draft, D.A.H., L.Y., A.S.F., F.C., T.G.; writing – review and editing, D.A.H., L.Y., G.M.W., F.P.M.dEV., D.P., A.S.F., S.X., F.C., T.G.

TABLES
Table 1.1

Model	d.f.	Covariance Parameters	Description	HR and C different among-line variance	HR and C different within-line variance	HR and C same among-line variance	HR and C same within-line variance	SAS Code
Full	6	4	Random effects for replicate line within selection treatment (linetype) and for mouse within line and linetype, allowing for separate variance estimates for both lines within linetype and mouse within line and linetype	x	x			proc mixed data=locus method=mivque0; class pop sub mouse; model COL1 =pop/solution; random sub(pop) /group=pop; random mouse(sub pop) /group=pop;
SepVarLines	6	3	Random effects for replicate line within selection treatment (linetype) and for mouse within line and linetype, allowing for separate variance estimates for line within linetype	x			x	proc mixed data=locus method=mivque0; class pop sub mouse; model COL1=pop/solution; random sub(pop) /group=pop; random mouse(sub pop);
SepVarInd	6	3	Random effects for replicate line within selection treatment (linetype) and for mouse within line and linetype, allowing for		x	x		proc mixed data=locus method=mivque0; class pop sub mouse; model

					separate variance estimates for mouse within line and linetype	COL1=pop/solution;
						random sub(pop);
						random mouse(sub pop) /group=pop;

Simple	6	2	x	x	Random effects for replicate line within selection treatment (linetype) and for mouse within line and linetype (as used by Xu and Garland 2017)	proc mixed
						data=locus
						method=mivque0;
						class pop sub mouse;
						model
						COL1=pop/solution;
						random sub(pop);
						random mouse(sub pop);

Multiple models used to analyze the allelic SNP data (two values per mouse) for whole-genome sequences from 79 mice. For each model, we used SAS Procedure Mixed with MIVQUE estimation (Xu and Garland 2017) to obtain the test statistic (F), significance level (P), and AICc (d.f. method was containment).

[a] For some loci, the within-line variance was zero for all 8 lines. In those cases, we used direct enumeration to calculate a significance level, i.e., the probability of observing the pattern versus the 23 possible combinations. See text for further details.

Table 1.2

Dataset	Total "Loci"	Significant Loci	Critical Threshold	Significant Genes
MegaMUGA	25,332	162[a]	P<0.00526 (5% FWER)	174[b]
Whole-Genome SNPs	5,932,124	84	P<0.001 (Local Maximum)	27
Haplotypes	16,901	102[c] (28 regions)	P<0.00526 (See text)	154[b]
All HR Fixed, All C Polymorphic	5,932,124	2,562 (46 regions)	See text	135[b]

Basic descriptive statistics for the primary analyses

[a] In Xu and Garland (2017), 152 SNPs were identified as statistically significant with a single model and the MIVQUE procedure, after use of a permutation procedure to control the family-wise Type I error rate (FWER) at 5% (P < 0.00343).
[b] These are not genes that SNPs fell into. These are genes close to significant SNPs or haplotypes.
[c] From 28 closely linked groups.

Table 1.3

Line	Polymorphic SNP loci	SNP %	Polymorphic Haplotypes	Haplotype %	SNP Het	Haplotype Het
C1	2,333,951	39.3%	7,773	46.0%	14.7%	17.8%
C2	2,436,225	41.1%	7,652	45.3%	13.7%	16.6%
C3	2,602,007	43.9%	7,841	46.4%	15.8%	17.8%
C5	2,102,405	35.4%	7,160	42.4%	12.7%	16.5%
HR3	2,819,828	47.5%	8,717	51.6%	18.1%	19.6%
HR6	2,220,487	37.4%	7,060	41.8%	13.5%	16.2%
HR7	2,042,309	34.4%	6,304	37.3%	13.0%	14.7%
HR8	2,226,282	37.5%	7,315	43.3%	14.4%	16.6%

Summary of polymorphism and heterozygosity by line

Table 1.4

Model	MegaMUGA[a]	WGS[a]	Hap 2-allele[b]	Hap 3-allele[b]
Full	9,875 (39.0%)	2,441,601 (41.2%)	4,512 (40.9%)	5,510 (46.9%)
SepVarLine	3,105 (12.3%)	504,946 (8.5%)	1,052 (9.5%)	1,583 (13.5%)
SepVarInd	2,983 (11.8%)	716,265 (12.1%)	726 (6.6%)	748 (6.4%)
Simple	8,654 (34.2%)	2,186,803 (36.9%)	4,594 (41.6%)	3,615 (30.8%)
# with no within-line variance	715 (2.8%)	82,533 (1.4%)	148 (1.3%)	282 (2.4%)

Model preference by data set, test, and allele counts
[a]Number of SNPS whose lowest AICc match the indicated model
[b]Number of haplotype blocks whose lowest AICc match the indicated model (one for each dummy variable for 3-allele blocks)

Table 1.5

Group	Chr	Start (BP)	End (BP)	Size (BP)	P-value
1	2	43,100,041	43,214,647	114,606	4.42E-03
2	3	51,580,020	51,659,891	79,871	2.25E-06
3	4	89,300,145	89,357,884	57,739	4.92E-03
4	4	155,480,343	155,654,426	174,083	3.94E-04
5	5	108,000,623	108,679,807	679,184	4.85E-04
6	5	118,824,587	119,299,787	475,200	2.15E-03
7	5	132,540,807	133,720,551	1,179,744	1.12E-03
8	6	37,440,411	37,659,588	219,177	3.47E-03
9	6	41,584,862	43,431,434	1,846,572	1.47E-05
10	7	29,640,243	29,697,093	56,850	5.67E-04
11	9	41,240,184	42,275,833	1,035,649	4.90E-07
12	10	75,061,742	75,456,261	394,519	3.99E-03
13	10	103,363,232	104,139,953	776,721	3.94E-03
14	10	105,220,041	105,699,704	479,663	3.72E-03
15	11	79,724,263	81,409,849	1,685,586	1.89E-04
16	11	114,466,946	114,489,018	22,072	2.69E-03
17	14	52,100,155	54,334,868	2,234,713	5.62E-04
18	14	98,380,090	98,679,965	299,875	2.22E-03
19	15	18,960,135	19,759,996	799,861	1.09E-03
20	15	69,120,025	70,219,737	1,099,712	4.53E-03
21	15	71,480,090	71,559,595	79,505	9.91E-04
22	15	86,541,805	86,599,823	58,018	3.55E-03
23	16	31,540,757	33,178,952	1,638,195	2.79E-04
24	16	40,742,298	41,357,426	615,128	1.01E-03
25	17	18,020,933	18,039,390	18,457	3.54E-04
26	17	20,700,046	20,939,819	239,773	3.54E-04

| 27 | 17 | 23,000,233 | 23,586,163 | 23,599,776 | 599,543 | 3.54E-04 |
| 28 | 17 | 65,458,617 | 65,738,255 | | 279,638 | 1.46E-03 |

Significant haplotype groups

Table 1.6

Chr	Start	End	Size	Lowest P
17	17,846,983	23,586,163	5,739,180	7.54E-05
10	103,429,623	105,529,701	2,100,078	3.73E-05
16	31,440,034	33,128,268	1,688,234	7.05E-06
15	18,958,730	20,635,226	1,676,496	8.49E-05
16	16,235,542	17,805,005	1,569,463	7.04E-04

Top 5 largest suggestive regions

Table 1.7

GO Term	Total Genes	Input Genes	Expected	Fold Enrichment	Raw P-Value
detection of chemical stimulus involved in sensory perception of smell	3	1	0.02	47.88	2.74E-02
sensory perception of smell	1128	27	7.85	3.44	2.46E-08
sensory perception of chemical stimulus	1228	34	8.55	3.98	5.71E-12
sensory perception	1641	36	11.42	3.15	7.12E-10
detection of chemical stimulus involved in sensory perception	59	7	0.41	17.04	3.65E-07
detection of stimulus involved in sensory perception	136	8	0.95	8.45	7.40E-06
detection of stimulus	236	9	1.64	5.48	5.40E-05
detection of chemical stimulus	85	7	0.59	11.83	3.53E-06
G protein-coupled receptor signaling pathway	1853	37	12.9	2.87	4.86E-09
regulation of systemic arterial blood pressure by aortic arch baroreceptor feedback	1	1	0.01	>100	1.38E-02
system process	2594	42	18.06	2.33	2.12E-07
multicellular organismal process	7307	74	50.87	1.45	1.43E-04
nervous system process	2085	39	14.51	2.69	9.97E-09
sensory perception of sour taste	5	1	0.03	28.73	4.08E-02
sensory perception of taste	71	7	0.49	14.16	1.15E-06
detection of chemical stimulus involved in sensory perception of bitter taste	47	6	0.33	18.34	1.74E-06
sensory perception of bitter taste	51	6	0.36	16.9	2.69E-06
detection of chemical stimulus involved in sensory perception of taste	51	6	0.36	16.9	2.69E-06

Top Biological process terms from GO analysis for Haplotype

Table 1.8

GO Term	Total Genes	Input Genes	Expected	Fold Enrichment	Raw P-Value
locomotory exploration behavior	16	1	0.02	53.6	1.96E-02
locomotory behavior	**240**	**4**	**0.28**	**14.29**	**1.72E-04**
behavior	**685**	**6**	**0.8**	**7.51**	**1.17E-04**
positive regulation by host of viral release from host cell	5	1	0.01	>100	6.97E-03
positive regulation of viral release from host cell	15	1	0.02	57.17	1.85E-02
regulation of viral release from host cell	31	1	0.04	27.66	3.66E-02
regulation of locomotion	**1040**	**7**	**1.21**	**5.77**	**1.47E-04**
negative regulation of cardiac muscle cell proliferation	**17**	**2**	**0.02**	**>100**	**2.20E-04**
negative regulation of cell population proliferation	684	3	0.8	3.76	4.46E-02
negative regulation of cardiac muscle tissue growth	**29**	**2**	**0.03**	**59.14**	**5.94E-04**
regulation of cardiac muscle tissue growth	74	2	0.09	23.18	3.53E-03
regulation of cardiac muscle tissue development	98	2	0.11	17.5	6.02E-03
regulation of striated muscle tissue development	160	2	0.19	10.72	1.52E-02
regulation of muscle tissue development	163	2	0.19	10.52	1.57E-02
regulation of muscle organ development	164	2	0.19	10.46	1.59E-02
regulation of heart growth	80	2	0.09	21.44	4.09E-03
regulation of organ growth	114	2	0.13	15.04	8.02E-03
negative regulation of cardiac muscle tissue development	40	2	0.05	42.88	1.09E-03
negative regulation of striated muscle tissue development	64	2	0.07	26.8	2.67E-03
negative regulation of muscle organ development	66	2	0.08	25.99	2.83E-03
negative regulation of muscle tissue development	67	2	0.08	25.6	2.92E-03
negative regulation of heart growth	**29**	**2**	**0.03**	**59.14**	**5.94E-04**
bundle of His cell-Purkinje myocyte adhesion involved in cell communication	6	1	0.01	>100	8.13E-03
bundle of His cell to Purkinje myocyte communication	13	1	0.02	65.96	1.62E-02
cell communication involved in cardiac conduction	32	1	0.04	26.8	3.78E-02
multicellular organismal signaling	109	2	0.13	15.73	7.37E-03
cardiac muscle cell-cardiac muscle cell adhesion	7	1	0.01	>100	9.28E-03
cell-cell adhesion	389	3	0.45	6.61	1.04E-02
cell adhesion	**789**	**6**	**0.92**	**6.52**	**2.50E-04**
biological adhesion	**799**	**6**	**0.93**	**6.44**	**2.68E-04**
negative regulation of cellular extravasation	8	1	0.01	>100	1.04E-02
negative regulation of leukocyte migration	41	2	0.05	41.83	1.14E-03
regulation of leukocyte migration	209	2	0.24	8.21	2.49E-02
regulation of cell migration	912	5	1.06	4.7	3.71E-03
regulation of cell motility	963	5	1.12	4.45	4.67E-03
negative regulation of cell migration	**276**	**4**	**0.32**	**12.43**	**2.91E-04**
negative regulation of cell motility	**289**	**4**	**0.34**	**11.87**	**3.46E-04**
negative regulation of cellular component movement	**323**	**4**	**0.38**	**10.62**	**5.24E-04**
definitive hemopoiesis	**21**	**2**	**0.02**	**81.67**	**3.25E-04**

Table 1.9

GO Term	Total Genes	Input Genes	Expected	Fold Enrichment	Raw P-value
response to pheromone	104	8	0.63	12.7	3.93E-07
complement receptor mediated signaling pathway	13	4	0.08	50.82	2.81E-06
phospholipase C-activating G protein-coupled receptor signaling pathway	91	5	0.55	9.07	2.89E-04
exocytic insertion of neurotransmitter receptor to postsynaptic membrane	8	3	0.05	61.93	3.40E-05
regulation of postsynaptic membrane neurotransmitter receptor levels	62	3	0.38	7.99	7.09E-03
neurotransmitter receptor transport to postsynaptic membrane	20	3	0.12	24.77	3.46E-04
neurotransmitter receptor transport to plasma membrane	21	3	0.13	23.59	3.93E-04
vesicle-mediated transport to the plasma membrane	90	3	0.54	5.51	1.87E-02
neurotransmitter receptor transport	40	3	0.24	12.39	2.21E-03
establishment of protein localization to postsynaptic membrane	21	3	0.13	23.59	3.93E-04
protein localization to postsynaptic membrane	44	3	0.27	11.26	2.85E-03
protein localization to synapse	76	3	0.46	6.52	1.21E-02
receptor localization to synapse	51	3	0.31	9.72	4.23E-03
calcium ion import across plasma membrane	9	2	0.05	36.7	1.91E-03
calcium ion import into cytosol	10	2	0.06	33.03	2.28E-03
calcium ion transport into cytosol	69	3	0.42	7.18	9.40E-03
positive regulation of cytosolic calcium ion concentration	292	7	1.77	3.96	2.26E-03
regulation of cytosolic calcium ion concentration	340	8	2.06	3.89	1.25E-03
cellular calcium ion homeostasis	446	10	2.7	3.7	4.48E-04
calcium ion homeostasis	463	10	2.8	3.57	5.95E-04

Top GO results for FixedHR/PolyC implicated genes

Table 1.10

Search Term	Total Genes	Total SNPs	Top Genes	Top P-value
Dopamin*	254	43,890	Gnb1, Fpr*, Adora2a	1.33E-04
Serotonin	45	7,315	Htr7, Chrm2, Btbd9	9.33E-03
Osteo*	491	56,091	Noct, Nf1, Mmp14	3.76E-05
Cardiac	820	143,507	Myh11, Tbx5, Dlg1	7.25E-06
"Skeletal Muscle"	295	39,383	Kel, Foxp1, Nf1	5.23E-06
Brain	667	123,416	Sorl1, Gak, Fbxo45	1.92E-07

Summary of ontology search.

[a] Genes are listed from most significant to least significant by SNP with lowest p-value

[b] Includes: *Fpr1, Fpr2, Fpr3, Fpr-rs4* (all closely linked)

Table 1.11

Chr	First BP	Last BP	Included Genes
5	108,000,623	108,679,807	Tmed5, Ccdc18, Pigg, Mfsd7a, Gak, Tmem175, Slc26a1
6	41,584,862	41,918,440	Trpv5, Trpv6, Ephb6, Kel, Llcfc1, Olfr459
7	29,603,841	29,697,093	Catsperg2
9	41,240,184	42,275,833	Sorl1, Mir100hg, Mir100, Mir125b-1, Mirlet7a-2, Tbcel[a]
11	79,724,263	80,090,780	Atad5, Suz12, Utp6, Crlf3
11	112,227,183	114,489,018	BC006965, Sox9
14	52,072,148	53,779,979	Olfr[b], Trav[b]
14	97,645,171	98,679,965	Dach1
15	18,960,135	20,609,074	Cdh10, Gm35496
15	71,023,429	71,559,595	Fam135b
16	31,540,757	33,178,952	Gm536, Rnf168, Ubxn7, Fbxo45, Tnk2, Tnk2os
17	17,895,909	22,396,753	Vmn2r

Genomic regions implicated by LM, haplotype, and FixedHR/PolyC analyses
[a] Tbcel is most differentiated gene in genome based on median p-value
[b] Several genes in this gene family were represented in this region

Table 1.12

	Test Group 1				Test Group 2			
Profile	Rep 1	Rep 2	Rep 3	Rep 4	Rep 1	Rep 2	Rep 3	Rep 4
1	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1
3	Het	Het	Het	Het	0	0	0	0
4	0	0	1	1	0	0	0	0
5	0	0	1	1	0	0	1	1

FIGURES

Figure 1.1
Manhattan plot for haplotype data. Red line indicates p-value <0.005 (see Methods and Materials), which yielded 28 haplotype groups (see Table 5).
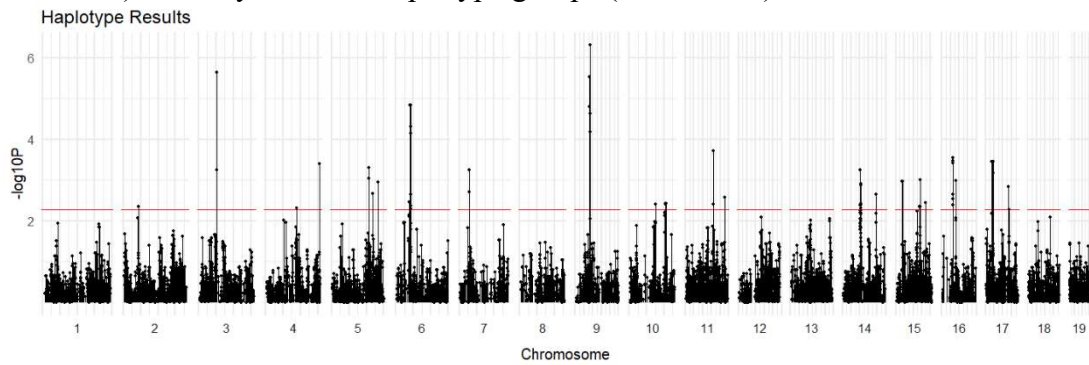
Figure 1.2
Manhattan plot for WGS SNP data. Red dots represent local maxima (N = 84).
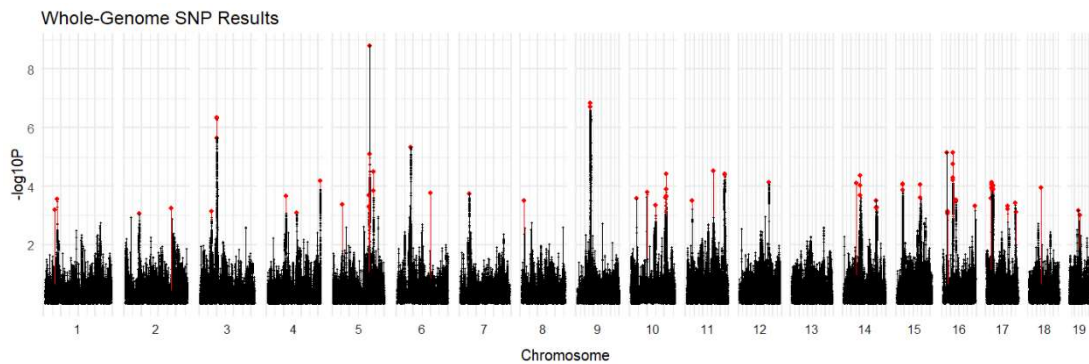
Figure 1.3
These are images of different variance structures depicted by actual examples from the MegaMUGA data (Xu and Garland 2017). This includes example data that were best fit by the "Full" model (A), "SepVarInd" model (B), "SepVarLines" model (C), and the "Simple" model (D). E shows a locus that had no within-line variance. P-values are significance levels for comparing the HR and C lines.

A "Full" Model
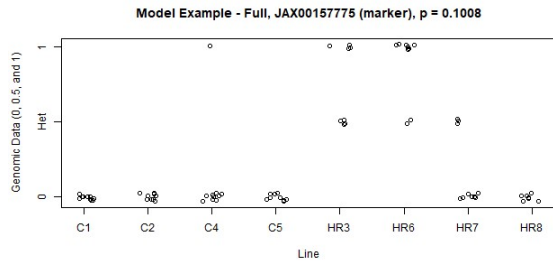
Model Example - Full, JAX00157775 (marker), p = 0.1008

B "SepVarInd" Model

Model Example - SepVarInd, UNC25816949 (marker), p = 0.006754

C "SepVarLines" Model

Model Example - SepVarLines, JAX00170011 (marker), p = 0.04885

D "Simple" Model

Model Example - Simple, UNC080521812 (marker), p = 0.2687

E No Within-line Variance

Example - No Within-Line Variance, backupJAX00272176 (marker), p = 0.5
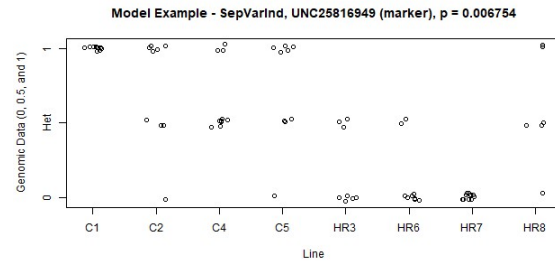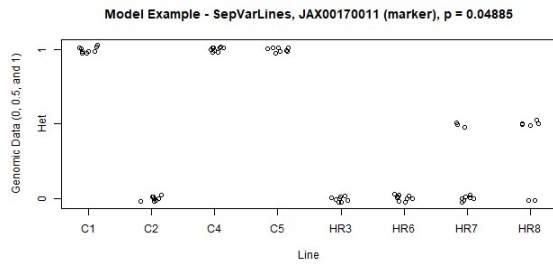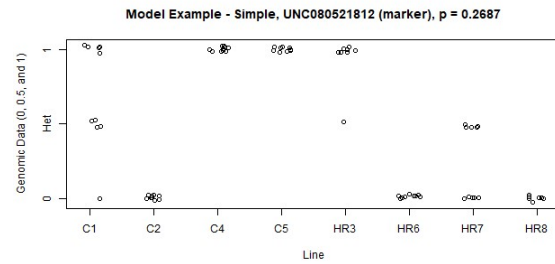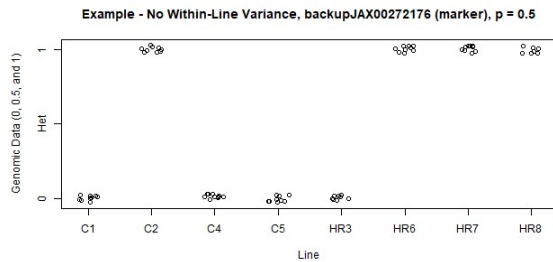
REFERENCES

Aebi, M., M. M. J. van Donkelaar, G. Poelmans, J. K. Buitelaar, E. J. S. Sonuga-Barke *et al.*, 2016 Gene-set and multivariate genome-wide association analysis of oppositional defiant behavior subtypes in attention-deficit/hyperactivity disorder. Am. J. Med. Genet. B Neuropsychiatr. Genet. 171: 573–588.

Akins, M. R., D. L. Benson, and C. A. Greer, 2007 Cadherin expression in the developing mouse olfactory system. J. Comp. Neurol. 501: 483–497.

Belke, T. W., and T. Garland, Jr., 2007 A brief opportunity to run does not function as a reinforcer for mice selected for high daily wheel-running rates. J. Exp. Anal. Behav. 88: 199–213.

Bilodeau, G. M., H. Guderley, D. R. Joanisse, and T. Garland, Jr., 2009 Reduction of type IIb myosin and IIB fibers in tibialis anterior muscle of mini-muscle mice from high-activity lines. J. Exp. Zool. Part Ecol. Genet. Physiol. 311A: 189–198.

Blair, S. N., and J. N. Morris, 2009 Healthy Hearts—and the universal benefits of being physically active: physical activity and health. Ann. Epidemiol. 19: 253–256.

Booth, F. W., M. V. Chakravarthy, S. E. Gordon, and E. E. Spangenburg, 2002 Waging war on physical inactivity: using modern molecular ammunition against an ancient enemy. J. Appl. Physiol. 93: 3–30.

Britton, S. L., and L. G. Koch, 2001 Animal genetic models for complex traits of physical capacity: Exerc. Sport Sci. Rev. 29: 7–14.

Bronikowski, A. M., J. S. Rhodes, T. Garland, Jr., T. A. Prolla, T. A. Awad *et al.*, 2004 The evolution of gene expression in mouse hippocampus in response to selective breeding for increased locomotor activity. Evolution 58: 2079–2086.

Burke, M. K., 2012 How does adaptation sweep through the genome? Insights from long-term selection experiments. Proc. R. Soc. B Biol. Sci. 279: 5029–5038.

Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 467: 587–590.

Careau, V., M. E. Wolak, P. A. Carter, and T. Garland, Jr., 2013 Limits to behavioral evolution: the quantative genetics of a complex trait under directional selection. Evolution 67: 3102–3119.

Carlson, S. A., J. E. Fulton, M. Pratt, Z. Yang, and E. K. Adams, 2015 Inadequate physical activity and health care expenditures in the United States. Prog. Cardiovasc. Dis. 57: 315–323.

Castro, A. A., and T. Garland, Jr., 2018 Evolution of hindlimb bone dimensions and muscle masses in house mice selectively bred for high voluntary wheel-running behavior. J. Morphol. 279: 766–779.

Claghorn, G. C., Z. Thompson, J. C. Kay, G. Ordonez, T. G. Hampton *et al.*, 2017 Selective breeding and short-term access to a running wheel alter stride characteristics in house mice. Physiol. Biochem. Zool. 90: 533–545.

Copes, L. E., H. Schutz, E. M. Dlugsoz, S. Judex, and T. Garland, Jr., 2018 Locomotor activity, growth hormones, and systemic robusticity: An investigation of cranial vault thickness in mouse lines bred for high endurance running. Am. J. Phys. Anthropol. 166: 442–458.

Cordeiro, L. M. S., P. C. R. Rabelo, M. M. Moraes, F. Teixeira-Coelho, C. C. Coimbra *et al.*, 2017 Physical exercise-induced fatigue: The role of serotonergic and dopaminergic systems. Braz. J. Med. Biol. Res. 50:.

Cornier, M.-A., D. Dabelea, T. L. Hernandez, R. C. Lindstrom, A. J. Steig *et al.*, 2008 The metabolic syndrome. Endocr. Rev. 29: 777–822.

Davis, R. J., W. Shen, Y. I. Sandler, M. Amoui, P. Purcell *et al.*, 2001 Dach1 mutant mice bear no gross abnormalities in eye, limb, and brain development and exhibit postnatal lethality. Mol. Cell. Biol. 21: 1484–1490.

Dawes, M., T. Moore-Harrison, A. T. Hamilton, T. Ceaser, K. J. Kochan *et al.*, 2014 Differential gene expression in high- and low-active inbred mice. BioMed Res. Int. 2014: 1–9.

De Moor, M. H. M., Y.-J. Liu, D. I. Boomsma, J. Li, J. J. Hamilton *et al.*, 2009 Genome-wide association study of exercise behavior in dutch and american adults: Med. Sci. Sports Exerc. 41: 1887–1895.

Dewan, I., T. Garland, Jr., L. Hiramatsu, and V. Careau, 2019 I smell a mouse: indirect genetic effects on voluntary wheel-running distance, duration and speed. Behav. Genet. 49: 49–59.

Didion, J. P., A. P. Morgan, L. Yadgary, T. A. Bell, R. C. McMullan *et al.*, 2016 *R2d2* drives selfish sweeps in the house mouse. Mol. Biol. Evol. 33: 1381–1395.

Dietrich, A., 2004 Endocannabinoids and exercise. Br. J. Sports Med. 38: 536–541.

Dlugosz, E. M., H. Schutz, T. H. Meek, W. Acosta, C. J. Downs *et al.*, 2013 Immune response to a Trichinella spiralis infection in house mice from lines selectively bred for high voluntary wheel running. J. Exp. Biol. 216: 4212–4221.

Drickamer, L. C., and T. R. Evans, 1996 Chemosignals and activity of wild stock house mice, with a note on the use of running wheels to assess activity in rodents. Behav. Processes 36: 51–66.

Dyrstad, S. M., B. H. Hansen, I. M. Holme, and S. A. Anderssen, 2014 Comparison of self-reported versus accelerometer-measured physical activity. Med. Sci. Sports Exerc. 46: 99–106.

Elkouris, M., H. Kontaki, A. Stavropoulos, A. Antonoglou, K. C. Nikolaou *et al.*, 2016 SET9-mediated regulation of TGF-β signaling links protein methylation to pulmonary fibrosis. Cell Rep. 15: 2733–2744.

Fisher, R. A., 1925 Statistical methods for research workers, pp. 25–235 in *Biological monographs and manuals*, edited by F. A. E. Crew and D. W. Cutler. Oliver and Boyd (Edinburgh).

Frédéric, M. Y., V. F. Lundin, M. D. Whiteside, J. G. Cueva, D. K. Tu *et al.*, 2013 Identification of 526 conserved metazoan genetic innovations exposes a new role for cofactor E-like in neuronal microtubule homeostasis (A. D. Chisholm, Ed.). PLoS Genet. 9: e1003804.

Garland, Jr., T., and P. W. Freeman, 2005 Selective breeding for high endurance running increases hindlimb symmetry. Evolution 59: 1851–1854.

Garland, Jr., T., S. A. Kelly, J. L. Malisch, E. M. Kolb, R. M. Hannon *et al.*, 2011a How to run far: multiple solutions and sex-specific responses to selective breeding for high voluntary activity levels. Proc. R. Soc. B Biol. Sci. 278: 574–581.

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

Garland, Jr., T., and M. R. Rose, 2009 *Experimental evolution: concepts, methods, and applications of selection experiments, Chapter 1*. University of California Press.

Garland, Jr., T., H. Schutz, M. A. Chappell, B. K. Keeney, T. H. Meek *et al.*, 2011b The biological control of voluntary exercise, spontaneous physical activity and daily energy expenditure in relation to obesity: human and rodent perspectives. J. Exp. Biol. 214: 206–229.

Garland, Jr., T., M. Zhao, and W. Saltzman, 2016 Hormones and the evolution of complex traits: insights from artificial selection on behavior. Integr. Comp. Biol. 56: 207–224.

Guthold, R., G. A. Stevens, L. M. Riley, and F. C. Bull, 2018 Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob. Health 6: e1077–e1086.

Herbert, A. J., A. G. Williams, P. J. Hennis, R. M. Erskine, C. Sale *et al.*, 2019 The interactions of physical activity, exercise and genetics and their associations with bone mineral density: implications for injury risk in elite athletes. Eur. J. Appl. Physiol. 119: 29–47.

Hiramatsu, L., and T. Garland, Jr., 2018 Mice selectively bred for high voluntary wheel-running behavior conserve more fat despite increased exercise. Physiol. Behav. 194: 1–8.

Hiramatsu, L., J. C. Kay, Z. Thompson, J. M. Singleton, G. C. Claghorn *et al.*, 2017 Maternal exposure to Western diet affects adult body composition and voluntary wheel running in a genotype-specific manner in mice. Physiol. Behav. 179: 235–245.

Hoenderop, J. G. J., J. P. T. M. van Leeuwen, B. C. J. van der Eerden, F. F. J. Kersten, A. W. C. M. van derKemp *et al.*, 2003 Renal Ca2+ wasting, hyperabsorption, and reduced bone thickness in mice lacking TRPV5. J. Clin. Invest. 112: 1906–1914.

Horner, A., L. Shum, J. A. Ayres, K. Nonaka, and G. H. Nuckolls, 2002 Fibroblast growth factor signaling regulates Dach1 expression during skeletal development. Dev. Dyn. 225: 35–45.

Horwitz, T., K. Lam, Y. Chen, Y. Xia, and C. Liu, 2019 A decade in psychiatric GWAS research. Mol. Psychiatry 24: 378–389.

Jiang, S., C. Koolmeister, J. Misic, S. Siira, I. Kühl *et al.*, 2019 TEFM regulates both transcription elongation and RNA processing in mitochondria. EMBO Rep. 20:.

Kay, J. C., G. C. Claghorn, Z. Thompson, T. G. Hampton, and T. Garland, Jr., 2019 Electrocardiograms of mice selectively bred for high levels of voluntary exercise: Effects of short-term exercise training and the mini-muscle phenotype. Physiol. Behav. 199: 322–332.

Keeney, B. K., T. H. Meek, K. M. Middleton, L. F. Holness, and T. Garland, Jr., 2012 Sex differences in cannabinoid receptor-1 (CB1) pharmacology in mice

selectively bred for high voluntary wheel-running behavior. Pharmacol. Biochem. Behav. 101: 528–537.

Kelly, S. A., T. A. Bell, S. R. Selitsky, R. J. Buus, K. Hua *et al.*, 2013 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 195: 1385–1395.

Kelly, S. A., P. P. Czech, J. T. Wight, K. M. Blank, and T. Garland, Jr., 2006 Experimental evolution and phenotypic plasticity of hindlimb bones in high-activity house mice. J. Morphol. 267: 360–374.

Kelly, S. A., F. R. Gomes, E. M. Kolb, J. L. Malisch, and T. Garland, Jr., 2017 Effects of activity, genetic selection and their interaction on muscle metabolic capacities and organ masses in mice. J. Exp. Biol. 220: 1038–1047.

Kelly, S. A., D. L. Nehrenberg, K. Hua, T. Garland, Jr., and D. Pomp, 2012 Functional genomic architecture of predisposition to voluntary exercise in mice: expression QTL in the brain. Genetics 191: 643–654.

Kelly, S. A., D. L. Nehrenberg, K. Hua, T. Garland, Jr., and D. Pomp, 2014 Quantitative genomics of voluntary exercise in mice: transcriptional analysis and mapping of expression QTL in muscle. Physiol. Genomics 46: 593–601.

Kelly, S. A., D. L. Nehrenberg, J. L. Peirce, K. Hua, B. M. Steffy *et al.*, 2010 Genetic architecture of voluntary exercise in an advanced intercross line of mice. Physiol. Genomics 42: 190–200.

Khan, A. A., and J. G. Quigley, 2013 Heme and FLVCR-related transporter families SLC48 and SLC49. Mol. Aspects Med. 34: 669–682.

Kida, Y., 2004 Chick Dach1 interacts with the Smad complex and Sin3a to control AER formation and limb development along the proximodistal axis. Development 131: 4179–4187.

Kolb, E. M., S. A. Kelly, and T. Garland, Jr., 2013a Mice from lines selectively bred for high voluntary wheel running exhibit lower blood pressure during withdrawal from wheel access. Physiol. Behav. 112–113: 49–55.

Kolb, E. M., S. A. Kelly, K. M. Middleton, L. S. Sermsakdi, M. A. Chappell *et al.*, 2010 Erythropoietin elevates VO2,max but not voluntary wheel running in mice. J. Exp. Biol. 213: 510–519.

Kolb, E. M., E. L. Rezende, L. Holness, A. Radtke, S. K. Lee *et al.*, 2013b Mice selectively bred for high voluntary wheel running have larger midbrains: support for the mosaic model of brain evolution. J. Exp. Biol. 216: 515–523.

Konczal, M., P. Koteja, P. Orlowska-Feuer, J. Radwan, E. T. Sadowska *et al.*, 2016 Genomic response to selection for predatory behavior in a mammalian model of adaptive radiation. Mol. Biol. Evol. 33: 2429–2440.

Kostrzewa, E., and M. J. Kas, 2014 The use of mouse models to unravel genetic architecture of physical activity: a review: Unravel genetic architecture of physical activity. Genes Brain Behav. 13: 87–103.

Köttgen, A., C. Pattaro, C. A. Böger, C. Fuchsberger, M. Olden *et al.*, 2010 New loci associated with kidney function and chronic kidney disease. Nat. Genet. 42: 376–384.

Lee, D., X. Zhao, Y.-I. Yim, E. Eisenberg, and L. E. Greene, 2008 Essential role of cyclin-G–associated kinase (auxilin-2) in developing and mature mice (R. Parton, Ed.). Mol. Biol. Cell 19: 2766–2776.

Lightfoot, J. T., E. J. C. De Geus, F. W. Booth, M. S. Bray, M. Den Hoed *et al.*, 2018 Biological/genetic regulation of physical activity level: Consensus from GenBioPAC. Med. Sci. Sports Exerc. 50: 863–873.

Lightfoot, J. T., L. Leamy, D. Pomp, M. J. Turner, A. A. Fodor *et al.*, 2010 Strain screen and haplotype association mapping of wheel running in inbred mouse strains. J. Appl. Physiol. 109: 623–634.

Lightfoot, J. T., M. J. Turner, D. Pomp, S. R. Kleeberger, and L. J. Leamy, 2008 Quantitative trait loci for physical activity traits in mice. Physiol. Genomics 32: 401–408.

Lillie, M., C. F. Honaker, P. B. Siegel, and Ö. Carlborg, 2019 Bidirectional selection for body weight on standing genetic variation in a chicken model. Genes|Genomes|Genetics g3.400038.2019.

Lin, X., C. B. Eaton, J. E. Manson, and S. Liu, 2017 The genetics of physical activity. Curr. Cardiol. Rep. 19:.

Liu, Q., R. J. Duff, B. Liu, A. L. Wilson, S. G. Babb-Clendenon *et al.*, 2006 Expression of cadherin10, a type II classic cadherin gene, in the nervous system of the embryonic zebrafish. Gene Expr. Patterns 6: 703–710.

Loh, N. Y., L. Bentley, H. Dimke, S. Verkaart, P. Tammaro *et al.*, 2013 Autosomal dominant hypercalciuria in a mouse model due to a mutation of the epithelial calcium channel, TRPV5 (Y. Ishimaru, Ed.). PLoS ONE 8: e55412.

Loos, R. J. F., T. Rankinen, A. Tremblay, L. Pérusse, Y. Chagnon *et al.*, 2005 Melanocortin-4 receptor gene and physical activity in the Québec Family Study. Int. J. Obes. 29: 420–428.

Machado, C. B., K. C. Kanning, P. Kreis, D. Stevenson, M. Crossley *et al.*, 2014 Reconstruction of phrenic neuron identity in embryonic stem cell- derived motor neurons. Development 141: 784–794.

Manley, A. F., 1996 *Physical activity and health: a report of the Surgeon General*. DIANE Publishing.

Mardon, G., N. M. Solomon, and G. M. Rubin, 1994 *dachshund* encodes a nuclear protein required for normal eye and leg development in *Drosophila*. Development 120: 3473–3486.

Martin, G. M., S. N. Austad, and T. E. Johnson, 1996 Genetic analysis of ageing: role of oxidative damage and environmental stresses. Nat. Genet. 13: 25–34.

Mathes, W. F., D. L. Nehrenberg, R. Gordon, K. Hua, T. Garland, Jr. *et al.*, 2010 Dopaminergic dysregulation in mice selectively bred for excessive exercise or obesity. Behav. Brain Res. 210: 155–163.

Matsunaga, E., S. Nambu, M. Oka, and A. Iriki, 2015 Complex and dynamic expression of cadherins in the embryonic marmoset cerebral cortex. Dev. Growth Differ. 57: 474–483.

Matta Mello Portugal, E., T. Cevada, R. Sobral Monteiro-Junior, T. Teixeira Guimarães, E. da Cruz Rubini *et al.*, 2013 Neuroscience of exercise: from neurobiology mechanisms to mental health. Neuropsychobiology 68: 1–14.

McGillivray, D. G., T. Garland, Jr., E. M. Dlugosz, M. A. Chappell, and D. A. Syme, 2009 Changes in efficiency and myosin expression in the small-muscle phenotype of mice selectively bred for high voluntary running activity. J. Exp. Biol. 212: 977–985.

Meek, T. H., B. P. Lonquich, R. M. Hannon, and T. Garland, Jr., 2009 Endurance capacity of mice selectively bred for high voluntary wheel running. J. Exp. Biol. 212: 2908–2917.

Miro, X., X. Zhou, S. Boretius, T. Michaelis, C. Kubisch *et al.*, 2009 Haploinsufficiency of the murine polycomb gene Suz12 results in diverse malformations of the brain and neural tube. Dis. Model. Mech. 2: 412–418.

Mok, A., K.-T. Khaw, R. Luben, N. Wareham, and S. Brage, 2019 Physical activity trajectories and mortality: population based cohort study. BMJ l2323.

Morgan, A. P., and C. E. Welsh, 2015 Informatics resources for the Collaborative Cross and related mouse populations. Mamm. Genome 26: 521–539.

Myers, A., C. Gibbons, G. Finlayson, and J. Blundell, 2017 Associations among sedentary and active behaviours, body fat and appetite dysregulation: investigating the myth of physical inactivity and obesity. Br. J. Sports Med. 51: 1540–1544.

Nica, A. C., and E. T. Dermitzakis, 2013 Expression quantitative trait loci: present and future. Philos. Trans. R. Soc. B Biol. Sci. 368: 20120362.

Nicod, J., R. W. Davies, N. Cai, C. Hassett, L. Goodstadt *et al.*, 2016 Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing. Nat. Genet. 48: 912–918.

Nogueira, A., O. Molinero, A. Salguero, and S. Márquez, 2018 Exercise addiction in practitioners of endurance sports: A literature review. Front. Psychol. 9:.

Nuwal, T., M. Kropp, S. Wegener, S. Racic, I. Montalban *et al.*, 2012 The *Drosophila* homologue of tubulin-specific chaperone E–like protein is required for synchronous sperm individualization and normal male fertility. J. Neurogenet. 26: 374–381.

Pallafacchina, G., S. François, B. Regnault, B. Czarny, V. Dive *et al.*, 2010 An adult tissue-specific stem cell in its niche: A gene profiling analysis of in vivo quiescent and activated muscle satellite cells. Stem Cell Res. 4: 77–91.

Prince, S. A., K. B. Adamo, M. Hamel, J. Hardt, S. Connor Gorber *et al.*, 2008 A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int. J. Behav. Nutr. Phys. Act. 5: 56.

Rezende, E. L., F. R. Gomes, J. L. Malisch, M. A. Chappell, and T. Garland, Jr., 2006 Maximal oxygen consumption in relation to subordinate traits in lines of house mice selectively bred for high voluntary wheel running. J. Appl. Physiol. 101: 477–485.

Rhodes, J. S., S. C. Gammie, and T. Garland Jr, 2005 Neurobiology of mice selected for high voluntary wheel-running activity. Integr. Comp. Biol. 45: 438–455.

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

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

Rohe, M., 2008 Role of SORLA in the brain and its relevance for Alzheimer disease [Ph.D. Dissertation]: Freien Universität Berlin, 110 p.

van Rooij, E., D. Quiat, B. A. Johnson, L. B. Sutherland, X. Qi *et al.*, 2009 A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. Dev. Cell 17: 662–673.

Salsi, V., M. A. Vigano, F. Cocchiarella, R. Mantovani, and V. Zappavigna, 2008 Hoxd13 binds in vivo and regulates the expression of genes acting in key pathways for early limb and skeletal patterning. Dev. Biol. 317: 497–507.

Sarzynski, M. A., R. J. F. Loos, A. Lucia, L. Pérusse, S. M. Roth *et al.*, 2016 Advances in exercise, fitness, and performance genomics in 2015: Med. Sci. Sports Exerc. 48: 1906–1916.

Saul, M. C., P. Majdak, S. Perez, M. Reilly, T. Garland, Jr. *et al.*, 2017 High motivation for exercise is associated with altered chromatin regulators of monoamine receptor gene expression in the striatum of selectively bred mice: Striatal transcriptome of mice born to run. Genes Brain Behav. 16: 328–341.

Schmidt, V., A. Subkhangulova, and T. E. Willnow, 2017 Sorting receptor SORLA: cellular mechanisms and implications for disease. Cell. Mol. Life Sci. 74: 1475–1483.

Schwartz, N. L., B. A. Patel, T. Garland, Jr., and A. M. Horner, 2018 Effects of selective breeding for high voluntary wheel-running behavior on femoral nutrient canal size and abundance in house mice. J. Anat. 233: 193–203.

Sella, G., and N. H. Barton, 2019 Thinking about the evolution of complex traits in the era of genome-wide association studies. Annu. Rev. Genomics Hum. Genet. 20: 461–493.

Sheila, M., G. Narayanan, S. Ma, W. L. Tam, J. Chai *et al.*, 2019 Phenotypic and molecular features underlying neurodegeneration of motor neurons derived from spinal and bulbar muscular atrophy patients. Neurobiol. Dis. 124: 1–13.

Shim, H., H. Chun, C. D. Engelman, and B. A. Payseur, 2009 Genome-wide association studies using single-nucleotide polymorphisms versus haplotypes: an empirical comparison with data from the North American Rheumatoid Arthritis Consortium. BMC Proc. 3: S35.

Simonen, R. L., T. Rankinen, L. Pérusse, A. S. Leon, J. S. Skinner *et al.*, 2003 A dopamine D2 receptor gene polymorphism and physical activity in two family studies. Physiol. Behav. 78: 751–757.

Singleton, J. M., and T. Garland, Jr., 2019 Influence of corticosterone on growth, home-cage activity, wheel running, and aerobic capacity in house mice selectively bred for high voluntary wheel-running behavior. Physiol. Behav. 198: 27–41.

Smits, P., P. Li, J. Mandel, Z. Zhang, J. M. Deng *et al.*, 2001 The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. Dev. Cell 277–290.

Swallow, J. G., T. Garland, Jr., P. A. Carter, W.-Z. Zhan, and G. C. Sieck, 1998 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, and T. Garland Jr., 2001 Food consumption and body composition in mice selected for high wheel-running activity. J. Comp. Physiol. [B] 171: 651–659.

Swallow, J. G., J. S. Rhodes, and T. Garland Jr, 2005 Phenotypic and evolutionary plasticity of organ masses in response to voluntary exercise in house mice. Integr. Comp. Biol. 45: 426–437.

Taliun, D., J. Gamper, U. Leser, and C. Pattaro, 2016 Fast sampling-based whole-genome haplotype block recognition. IEEE/ACM Trans. Comput. Biol. Bioinform. 13: 315–325.

Talmadge, R. J., W. Acosta, and T. Garland, Jr., 2014 Myosin heavy chain isoform expression in adult and juvenile mini-muscle mice bred for high-voluntary wheel running. Mech. Dev. 134: 16–30.

Thompson, Z., 2017 The neurobiological basis of voluntary exercise in selectively-bred high runner mice: University of California, Riverside, 150 p.

Thompson, Z., D. Argueta, T. Garland, Jr., and N. DiPatrizio, 2017 Circulating levels of endocannabinoids respond acutely to voluntary exercise, are altered in mice selectively bred for high voluntary wheel running, and differ between the sexes. Physiol. Behav. 170: 141–150.

Vaanholt, L. M., I. Jonas, M. Doornbos, K. A. Schubert, C. Nyakas *et al.*, 2008 Metabolic and behavioral responses to high-fat feeding in mice selectively bred for high wheel-running activity. Int. J. Obes. 32: 1566–1575.

Wallace, I. J., and T. Garland, Jr., 2016 Mobility as an emergent property of biological organization: Insights from experimental evolution: Mobility and biological organization. Evol. Anthropol. Issues News Rev. 25: 98–104.

Wang, W., K. K. Touhara, K. Weir, B. P. Bean, and R. MacKinnon, 2016 Cooperative regulation by G proteins and Na+ of neuronal GIRK2 K+ channels. eLife 5:.

White, J. K., A.-K. Gerdin, N. A. Karp, E. Ryder, M. Buljan *et al.*, 2013 Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. Cell 154: 452–464.

Williams, C. J., M. G. Williams, N. Eynon, K. J. Ashton, J. P. Little *et al.*, 2017 Genes to predict VO2max trainability: a systematic review. BMC Genomics 18:.

Wood, A. R., The Electronic Medical Records and Genomics (eMERGE) Consortium, The MIGen Consortium, The PAGE Consortium, The LifeLines Cohort Study *et al.*, 2014 Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46: 1173–1186.

Xu, S., and T. Garland, 2017 A mixed model approach to genome-wide association studies for selection signatures, with application to mice bred for voluntary exercise behavior. Genetics 207: 785–799.

Yang, Z., Q. Sun, J. Guo, S. Wang, G. Song *et al.*, 2019 *GRSF1* -mediated *MIR-G-1* promotes malignant behavior and nuclear autophagy by directly upregulating *TMED5* and *LMNB1* in cervical cancer cells. Autophagy 15: 668–685.

Young, N. M., B. Hallgrímsson, and T. Garland, Jr., 2009 Epigenetic effects on integration of limb lengths in a mouse model: selective breeding for high voluntary locomotor activity. Evol. Biol. 36: 88–99.

Zheng, J.-J., W.-X. Li, J.-Q. Liu, Y.-C. Guo, Q. Wang *et al.*, 2018 Low expression of aging-related NRXN3 is associated with Alzheimer disease: A systematic review and meta-analysis. Medicine (Baltimore) 97: e11343.

Zhu, X., E.-S. Cho, Q. Sha, J. Peng, Y. Oksov *et al.*, 2014 Giant axon formation in mice lacking Kell, XK, or Kell and XK. Am. J. Pathol. 184: 800–807.

Zhu, X., A. Rivera, M. S. Golub, J. Peng, Q. Sha *et al.*, 2009 Changes in red cell ion transport, reduced intratumoral neovascularization, and some mild motor function

abnormalities accompany targeted disruption of the Mouse Kell gene (*Kel*). Am. J. Hematol. 84: 492–498.

Chapter 2

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

David A. Hillis and Theodore Garland, Jr.

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

ABSTRACT

Replicate lines under uniform selection often evolve in different ways. Previously, analyses using whole-genome sequence data for individual mice (*Mus musculus*) from four replicate High Runner (HR) lines and four non-selected control (C) lines demonstrated genomic regions that have responded consistently to selection for voluntary wheel-running behavior. Here, we ask whether the HR lines have evolved differently from each other, even though they reached selection limits at similar levels. We focus on one HR 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 four lines. Repeating analyses while dropping each of the other HR lines identified 12, 8, and 6 such regions. (Of these 45 regions, 37 were unique.) These results suggest that each HR line indeed responded to selection somewhat uniquely, but also that HR3 is the most distinct. We then applied two additional analytical approaches when dropping HR3 only (based on haplotypes and nonstatistical tests involving fixation patterns). All three approaches identified seven new regions (as compared with analyses using all four HR lines) that include genes associated with activity levels, dopamine signaling, hippocampus morphology, heart size, and body size, all of which differ between HR and C lines. Our results illustrate how multiple solutions and "private" alleles can obscure general signatures of selection involving "public" alleles.

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; b; Tenaillon *et al.* 2012; Wone *et al.* 2019), a phenomenon often termed multiple solutions (e.g., see Bock 1959; Bennett 2003; Garland, Jr. *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 one 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., multi-syndromic 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 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 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, Jr. and Rose 2009b). 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, Jr. *et al.* 2002: see below). Specifically, the High Runner (HR) mouse experiment includes four 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 four 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 four replicate HR lines vary in the extent to which daily running distance has evolved via increases in average speed versus 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, Jr. *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 Figure 3 in Garland, Jr. *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 don't involve the mini-muscle phenotype. Those non-mini-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 minutes 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 High Runner mouse selection experiment, based on systematic muscle dissections (Garland, Jr. *et al.* 2002). The *Myh4^{Minimsc}* allele was uncommon in the base population (frequency ~7%) and the phenotype has only been observed in two of the HR lines and in

one control line (Garland, Jr. *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 become fixed

for the *Myh4^Minimsc* allele by generation 36 (Garland, Jr. *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, Jr. *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, Jr. *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, Jr. *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 *et al.* 2021b; a). 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, Jr. *et al.* 2011).

Loci with such far-reaching pleiotropic effects as mini-muscle have great potential to result in non-additive 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 one underlying allele -- would differ from that of the other three 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 one or two lines) can obscure general signatures of selection involving "public" alleles (those present in all lines)(cf. Partridge and Gems 2002).

MATERIALS AND METHODS

High Runner 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 eight closed lines (generation 0). Four of these lines were randomly designated as High Runner (HR) lines (lab designated HR3, HR6, HR7, and HR8), which would undergo selection based on voluntary wheel running. The remaining four lines would serve as unselected control (C) lines (lab designated C1, C2, C4, and C5) (Figure 2.1). 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 12X 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 seven 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 one 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 SNPs variable 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). Additionally, p-values and the Aikake

Information Criterion corrected for small sample sizes (AICc) were calculated for four

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 three-step process.

First, we identified all SNPs differentiated with p-value $\leq$ 0.001. Second, we considered

that any two 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 1mbp as delineating

separate regions.

Power and Type I Error Simulations

All else being equal, dropping one of the eight 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 multi-model ANOVA method (Hillis *et al.* 2020), first with all eight lines and then dropping each of the HR lines one 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 eight lines at random. This process was repeated for all eight lines in sequence. 100,000 loci were thus created, and multi-model ANOVA was performed with all eight lines as well as dropping each of the HR lines.

Haplotype and Non-statistical Analyses Excluding HR3 (mini-muscle)

Following Hillis et al. (2020), we performed two additional analyses to gauge differentiation between the four C and three 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

89

"FixedHR/PolyC" (as in Hillis *et al.* 2020). Any loci or genomic region identified as differentiated in all three 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 eight 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 October 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 November 2021) (Bult *et al.* 2019).

Data Availability Statement

Original data were made available by Hillis et al. (2020) and can be found at https://doi.org/10.25386/genetics.12436649. Supplemental files can be found at https://academic.oup.com/genetics/advance-article/doi/10.1093/genetics/iyac165/6777268 (note that filenames on the website will exclude "2."). File S2.1 contains brief descriptions of supplemental tables. Table S2.1 contains all regions with at least 10 SNPs with $p \leq 0.001$ for any analyses where an HR line was dropped. Table S2.2 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). Table S2.3 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.

RESULTS

Principal Components Analysis

PCA across all lines (79 individuals) produced seven 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 (Figure 2.2A).

PCA of the 39 individuals in the HR lines included 4,679,533 variable SNP loci and produced three 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 three HR lines for scores on PC2 (Figure 2.2B,C,D).

SNP Analyses Excluding Individual HR Lines

Each of the analyses dropping one of the HR lines produced some new peaks, as compared with the original analyses (Figure 2.3). 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-fold greater than in the analyses including all 4 HR lines (Table 2.1). More specifically, this

difference is attributable mainly to a large increase in loci with p-values in this range in

two 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 $\leq$ 1e-03 after dropping any of the other

lines; however, dropping HR3 produces about 40,000 additional loci with low p-values in

these two regions (Table 2.2). Dropping HR8 resulted in a notable increase in loci with

p-values in the 1e-06 to 1e-08 range (Table 2.1), largely due to a single region containing

1,414 loci with uniquely low p-values (chr7:115,169,726-116,129,821) (Table 2.2). 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 2.1 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 one 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 2.2

and Table S2.1 concentrate on those regions with at least 10 SNPs with $p \leq 0.001$.

     Table S2.1 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 eight lines.

     The 45 regions listed in Table 2.2 are a subset of those shown in Table S2.1,

excluding regions where similar numbers of SNPs with $p \leq 0.001$ were produced when

dropping any HR line. The regions in Table 2.2 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 one specific HR line (e.g., chr1:155,052,375-157,767,127, see Figure 2.4 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 2.2, dropping the line fixed for $Myh4^{Minimsc}$ (HR3) produced more of these unique regions than dropping any other line (19 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 2.3). 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 2.3).

Power and Type I Error Simulations

Dropping any one of the eight 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 (Table S2.2). 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 0.045-fold when a line was dropped, whereas loci whose 8-line analyses produce a p-value < 1.00E-05 had an average increase of about 0.366-fold in p-values when a line was dropped.

Table 2.4 illustrates that Type I error rates for α = 0.05 are deflated in the 8-line analyses, as was noted previously (Hillis *et al.* 2020), and a similar deflation occurs for α = 0.01. Dropping an HR line from the analyses increases the Type I error rate for both α = 0.05 and α = 0.01 (Table 2.4). For α = 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 to 0.00286), and even more so for α = 0.0001 (range = 0.00060 to 0.00078). For some of the p thresholds (e.g., p ≤ 0.001), the increase was quite large relative to the Type I error rate for the 8-line analyses (Table 2.4).

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 2.1) were scaled to be out of 100,000 to match the simulation. When the estimated Type I error rate (Table 2.4) is subtracted from the frequency of calculated total positives (scaled from Table 2.1), many signatures of selection remain, particularly when dropping HR3 (Table 2.4).

Haplotype and Non-statistical 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 2.5, including one 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 three analytical methods (Table 2.5). These seven 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 2.5 also lists 15 additional regions identified by at least two of the three 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 three 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 two regions, chr3:46,438,071-52,624,971 and chr10:101,652,005-106,038,129 (Table 2.2), 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 four HR lines. However, the wider areas implicated by the other three 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 three 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) 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 (Figure 2.5). Therefore, each of the 37 new regions listed in Table 2.2 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.

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, Jr. *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 one C line and in two HR lines (e.g., see Figure 1 in Garland, Jr. *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 (non-additive 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 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 five 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 five of the loci will coincide with a selection limit, but these alleles may be different among replicate lines.

Signatures of Selection after Dropping One 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 one HR or C line from the SNP analyses usually produced loci fixed for opposite alleles between the different linetypes (Table 2.3). These SNPs unsurprisingly tend to be clustered 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 2.3 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, three 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; The International Mouse Phenotyping Consortium *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 2.3). 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 intriguing association when *Runx2* is found in a separate region identified when dropping HR3 (Table 2.6).

Variation in Olfactory Response to Selection

Olfaction is known to play an important role in some motivated behaviors (Nielsen 2017). 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 non-volatile organic compounds. Table 2.2 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 two

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 non-

vomeronasal 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 non-vomeronasal 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 2.2. To gauge the magnitude of this problem, we subtracted the

expected false positives (Table 2.4) from our total positives (scaled from Table 2.1). As

shown in Table 2.4, 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 three 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 seven genomic regions implicated by all three tests (SNPs, haplotype, and FixedHR/PolyC) that were not identified by any tests when analyzed with all 8 lines (Table 2.6). These regions contained nearly 61 genes; however, three 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. Additionally, human GWAS have implicated *Ncam1* in playing a role in heel bone mineral density (Kim 2018; 23andMe Research Team *et al.* 2019c) and addictive behaviors, specifically smoking (Kichaev *et al.* 2019; 23andMe Research Team *et al.* 2019a; b). 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 *et al.* 2021a). 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 versus 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 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). Additionally, *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). Additionally, 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, Jr. 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 (The International Mouse Phenotyping Consortium *et al.*

2016), impaired coordination and gait (Ho *et al.* 2020), decreased body mass and length

(The International Mouse Phenotyping Consortium *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, Jr. *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 2.2). 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 2.4). 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. Additionally, 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 seven new genomic regions of differentiation in 3 of the lines bred for high voluntary wheel running, as compared with the 4 non-selected Control lines. These regions contain genes that are both intuitive for voluntary-exercise behavior and correlate to known phenotypic differences between the High Runner 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, Jr. and Rose 2009a). 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 *E. 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 *Drosophila 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 two generations of random mating in our lab before being divided into eight 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 three of eight ever had mice with the mini-muscle phenotype due to the low allele frequency and recessive nature (Garland, Jr. *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 2.2 and 2.3), 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, Jr. and Rose 2009b).

ACKNOWLEDGEMENTS

FUNDING

CONFLICT OF INTEREST

The authors have no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, D.A.H., T.G.; investigation, D.A.H., T.G.; software, D.A.H.; formal analysis, D.A.H., T.G.; writing – original draft, D.A.H., T.G.; writing – review and editing, D.A.H., T.G.

TABLES
Table 2.1

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	2176	5
1.00E-07	2	1	215	6	7	4	138	5	1148	4
1.00E-08	2	1	207	5	2	1	126	5	608	4
1.00E-09	0	0	155*	4	0	0	63*	1	485*	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 Figure 2.3 for the Manhattan plot of these regions (Figure 2.3E for 8-line Manhattan plot).

* These SNPs have very low p-values (2E-39) due to fixation for opposite alleles in the C vs HR lines

** This is the p-value used for the red line in the Manhattan plots

Table 2.2

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≤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)

Table 2.3

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
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 three of these regions, at least some of these fixed regions are likely biologically significant across three, if not all four, HR lines.
* 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 2.4

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	5000	677	3512.9	2009	3314.4	2014	2769.9	2021	2619.9	1949	3056.7
0.01	1000	443	1033.7	654	1585.7	647	1053.0	640	848.7	624	1024.0
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

Type I error rates and estimated true positives

[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)

Table 2.5

Chromosome	Low BP	High BP	SNPs	Haplotype	FHRPC	Identified in 8-Line
1	156,378,655	156,798,297	x	x	x	LM
2	42,722,345	43,214,647	x	x	x	LM
3	35,754,848	36,551,567	x		x	
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
[a]4	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
[a]7	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
[a]9	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
[a]11	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
[a]13	48,002,204	48,380,363	x	x	x	
13	56,280,106	56,618,730	x	x		

116

Chr	Region Start	Region End						Hillis et al. (2020) 8-line
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	x		ALL[b]
16	87,080,074	88,667,433	x		x	x		
17	17,661,786	23,599,776	x		x	x		ALL
[a]17	44,063,088	44,685,529	x		x	x		
[a]18	56,147,892	60,494,440	x		x	x		

Regions Identified by Analyses Excluding HR3

[a] These were regions implicated by all three 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.

This table contains regions identified by at least two of the three 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).

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

Table 2.6

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

Genes Identified by all Analyses, after Excluding HR3. These regions were implicated by all three 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 Table S2.3.

FIGURES

Figure 2.1
Schematic illustration of the High Runner mouse artificial selection experiment, begun in 1993 with a base population of 224 outbred mice.
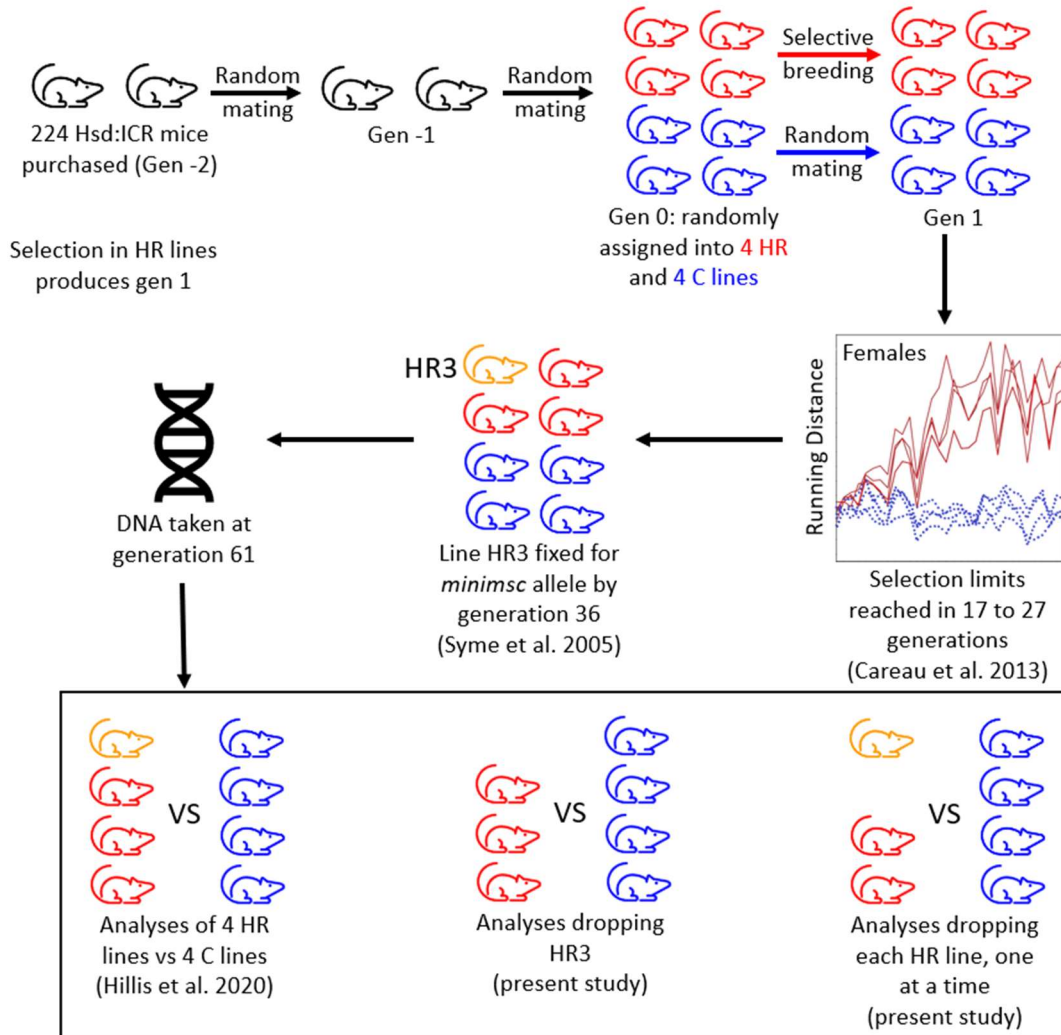
Figure 2.2

Principal components analysis of variable SNP loci. (A) Scatterplot of principle components (PCs) using all eight 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 four 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 three HR lines for scores on PC2.

A
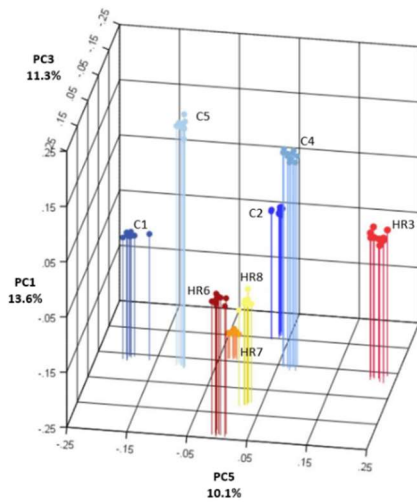
B
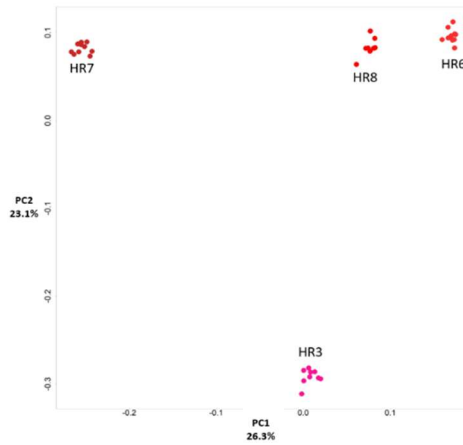
C
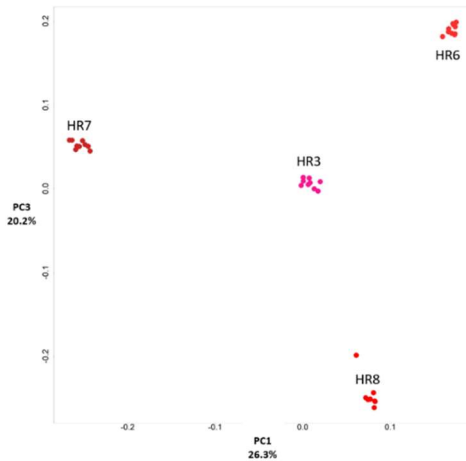
D
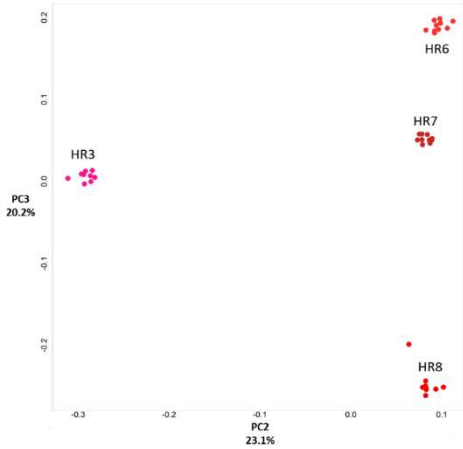
Figure 2.3
(A) Manhattan plot of the generation 61 individual mouse analyses, excluding HR3 (N=5,931,993). Vertical, dashed lines included represent loci fixed for opposite alleles (155 total loci) in the C and HR lines (e.g., p=2.43E-39). (B) Manhattan plot of the generation 61 individual mouse analyses, 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 Figure 2 of Hillis *et al.* 2020).

A

B

C

D

E

122
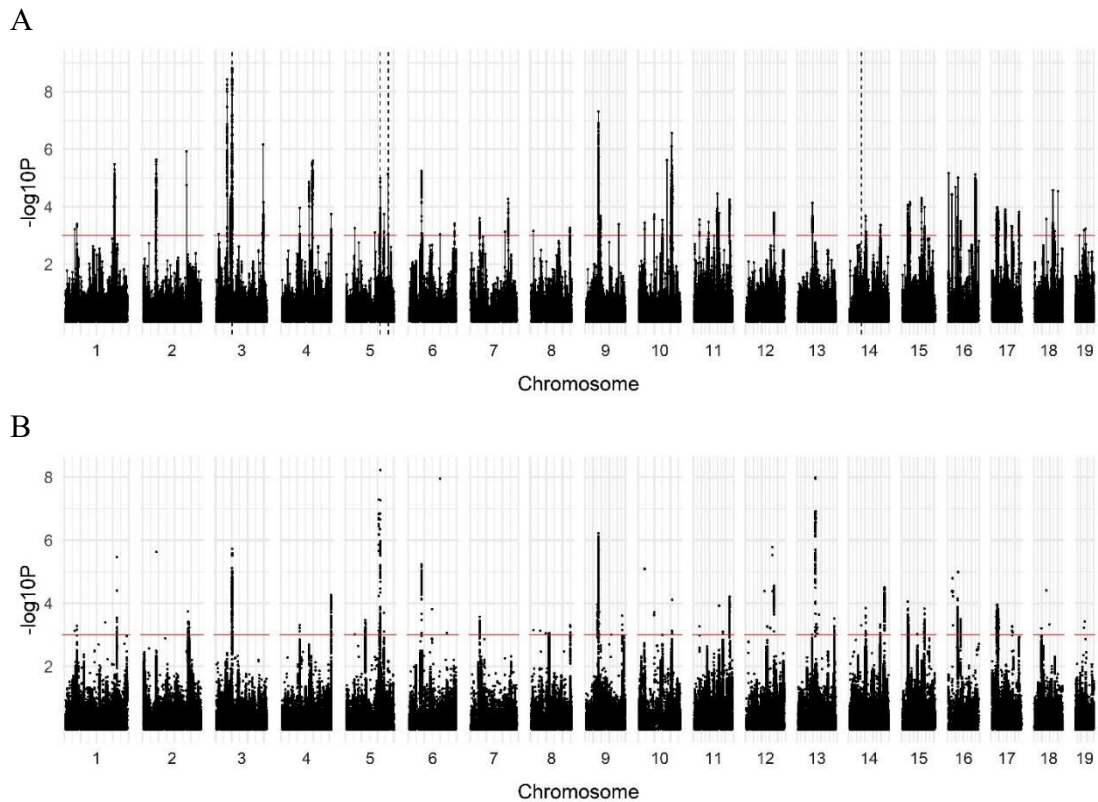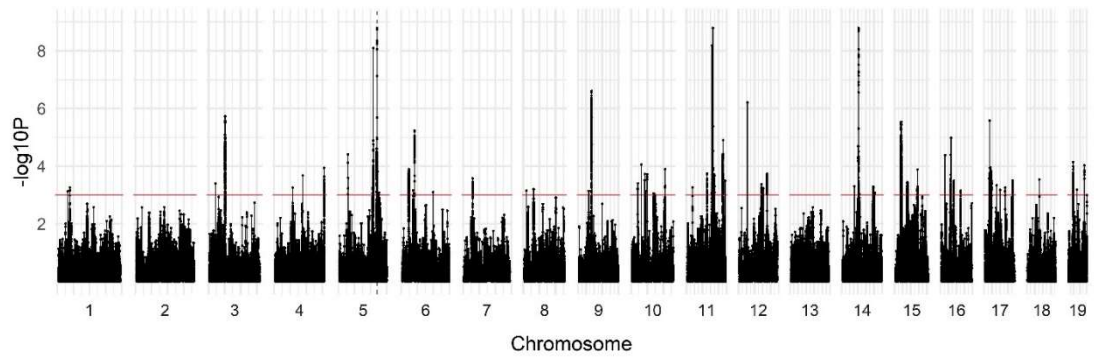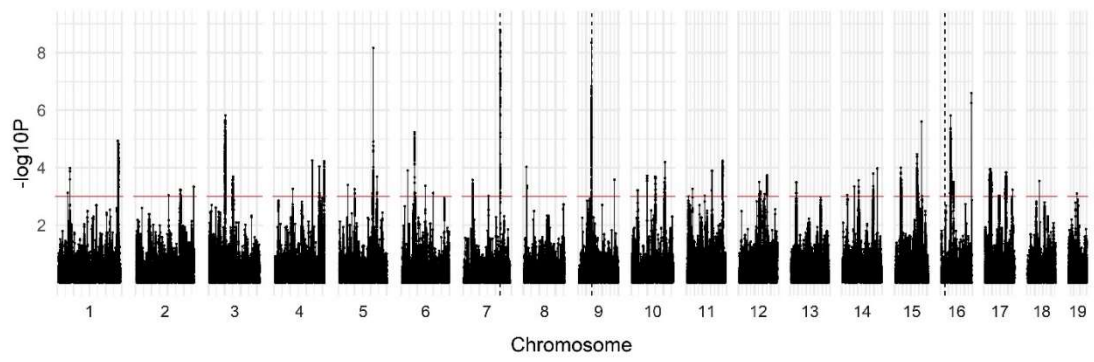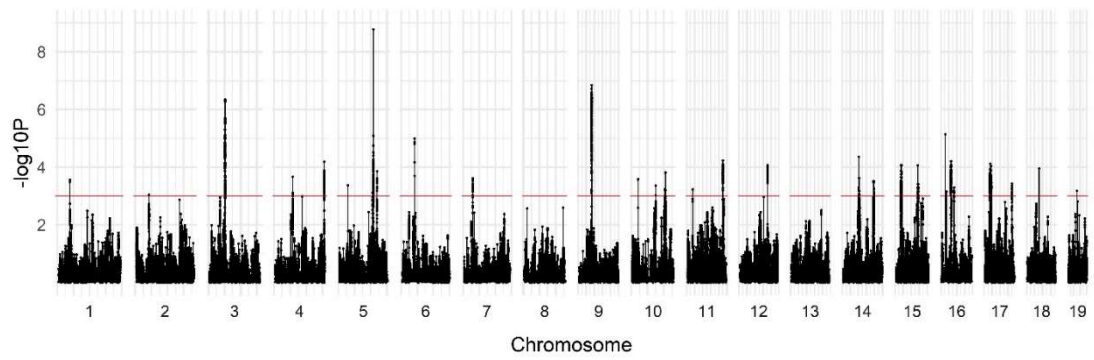
Figure 2.4

Allele frequencies for individuals SNPs within example regions from Table 2.2. (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 four 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.

Region 1, chr1 Position 155052375 to 157767127

Region 2, chr1 Position 163002979 to 163450173

123
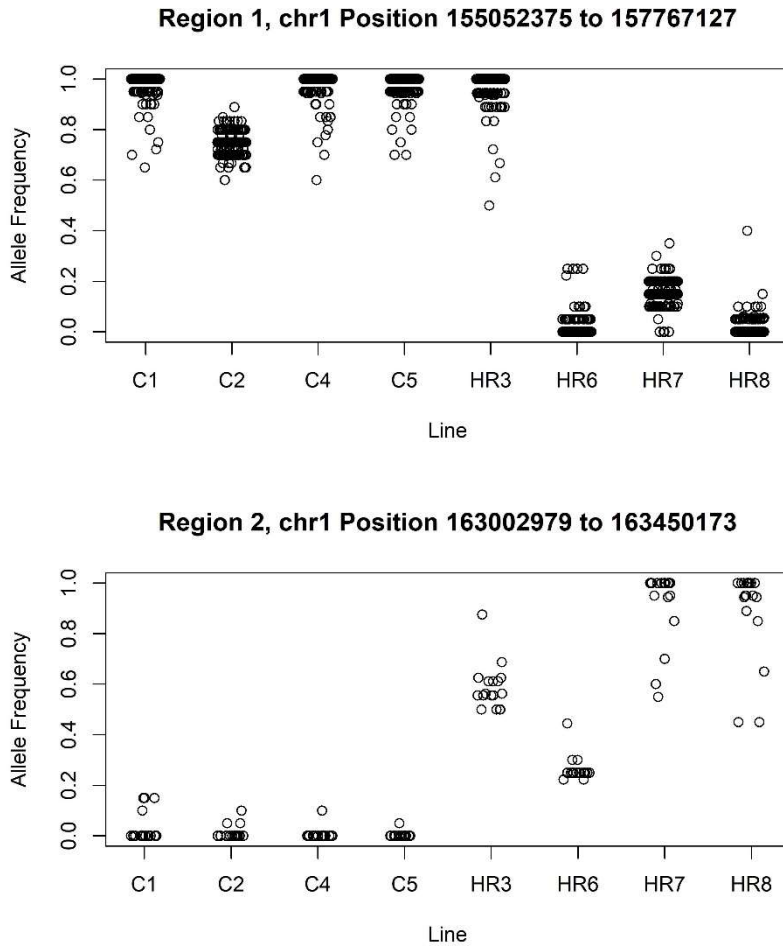
Region 3, chr1 Position 189994733 to 190372872
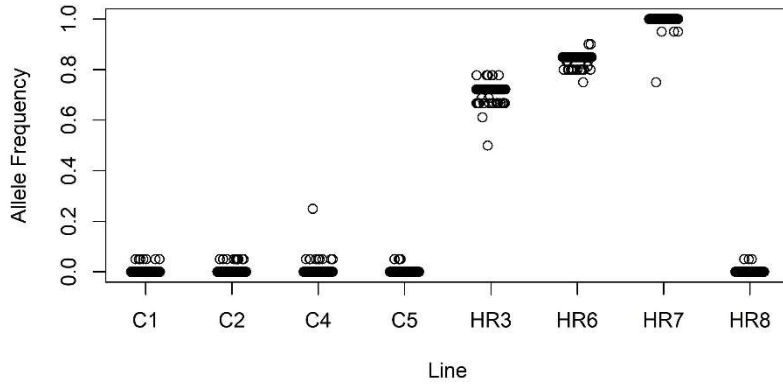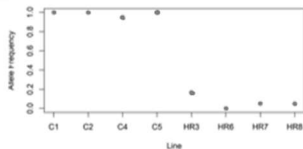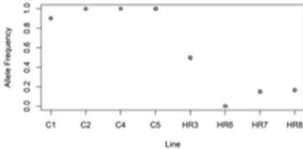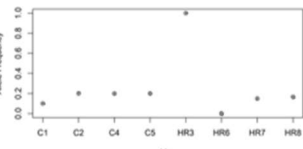
Figure 2.5

Illustration of different analysis strategies for detection of "private" alleles. The four 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 two analyses).

Analysis	Type I Error Rate	Description	Graph
4 HR vs 4 C	**VERY LOW**	All HR lines differentiated from controls (public alleles) Strongest evidence of (consistent) `response to selection.	G1
3 HR vs 4 C	**LOW**	3 HR lines differentiated from controls. Excluded HR line different enough from other HR lines to prevent 4v4 detection and potentially has private alleles.	G2
1 HR vs 4 C	**HIGH**	1 HR line differentiated from controls (potentially private alleles or drift).	G1: All HR lines G2: All HR lines G3: Line 3
1 HR vs 3 HR	**VERY HIGH**	1 HR line different from other 3 HR lines but locus is not necessarily different from controls.	G3

REFERENCES

23andMe Research Team, eQTLgen Consortium, International Cannabis Consortium, Social Science Genetic Association Consortium, R. Karlsson Linnér, *et al.*, 2019a 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. 51: 245–257. https://doi.org/10.1038/s41588-018-0309-3

23andMe Research Team, HUNT All-In Psychiatry, M. Liu, Y. Jiang, R. Wedow, *et al.*, 2019b Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat. Genet. 51: 237–244. https://doi.org/10.1038/s41588-018-0307-5

23andMe Research Team, J. A. Morris, J. P. Kemp, S. E. Youlten, L. Laurent, *et al.*, 2019c An atlas of genetic influences on osteoporosis in humans and mice. Nat. Genet. 51: 258–266. https://doi.org/10.1038/s41588-018-0302-x

Agrawal A. F., E. D. Brodie, and L. H. Rieseberg, 2001 Possible consequences of genes of major effect: Transient changes in the G-matrix, (A. P. Hendry, and M. T. Kinnison, Eds.). Genetica 112–113: 33–43. https://doi.org/10.1007/978-94-010-0585-2_3

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

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

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

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

Blum K., P. J. Sheridan, R. C. Wood, E. R. Braverman, T. J. Chen, *et al.*, 1995 Dopamine D2 receptor gene variants: association and linkage studies in impulsive-addictive-compulsive behaviour. Pharmacogenetics 5: 121–141. https://doi.org/10.1097/00008571-199506000-00001

Bock W. J., 1959 Preadaptation and multiple evolutionary pathways. Evolution 13: 194–211. https://doi.org/10.1111/j.1558-5646.1959.tb03005.x

Bronikowski A. M., J. S. Rhodes, T. Garland, Jr., T. A. Prolla, T. A. Awad, *et al.*, 2004 The evolution of gene expression in mouse hippocampus in response to selective breeding for increased locomotor activity. Evolution 58: 2079–2086. https://doi.org/10.1111/j.0014-3820.2004.tb00491.x

Bult C. J., J. A. Blake, C. L. Smith, J. A. Kadin, J. E. Richardson, *et al.*, 2019 Mouse Genome Database (MGD) 2019. Nucleic Acids Res. 47: D801–D806. https://doi.org/10.1093/nar/gky1056

Cadney M. D., L. Hiramatsu, Z. Thompson, M. Zhao, J. C. Kay, *et al.*, 2021 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. 234: 113389. https://doi.org/10.1016/j.physbeh.2021.113389

Careau V., M. E. Wolak, P. A. Carter, and T. Garland, Jr., 2013 Limits to behavioral evolution: the quantative genetics of a complex trait under directional selection. Evolution 67: 3102–3119. https://doi.org/10.1111/evo.12200

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. Biochem. Physiol. A. Mol. Integr. Physiol. 123: 155–162. https://doi.org/10.1016/S1095-6433(99)00044-6

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

Castro A. A., F. A. Karakostis, L. E. Copes, H. E. McClendon, A. P. Trivedi, *et al.*, 2021b Effects of selective breeding for voluntary exercise, chronic exercise, and their interaction on muscle attachment site morphology in house mice. J. Anat. joa.13547. https://doi.org/10.1111/joa.13547

Claghorn G. C., Z. Thompson, J. C. Kay, G. Ordonez, T. G. Hampton, *et al.*, 2017 Selective breeding and short-term access to a running wheel alter stride characteristics in house mice. Physiol. Biochem. Zool. 90: 533–545. https://doi.org/10.1086/692909

Cohan F. M., 1984a Genetic divergence under uniform selection. I. similarity among populations of *drosophila melanogaster* in their responses to artificial selection for modifiers of ci^d. Evolution 38: 55–71. https://doi.org/10.1111/j.1558-5646.1984.tb00260.x

Cohan F. M., 1984b Can uniform selection retard random genetic divergence between isolated conspecific populations? Evolution 38: 495–504. https://doi.org/10.1111/j.1558-5646.1984.tb00315.x

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

Cohan F. M., A. A. Hoffmann, and T. W. Gayley, 1989 A test of the role of epistasis in divergence under uniform selection. Evolution 43: 766–774. https://doi.org/10.1111/j.1558-5646.1989.tb05175.x

Cota C. D., P. Bagher, P. Pelc, C. O. Smith, C. R. Bodner, *et al.*, 2006 Mice with mutations in *Mahogunin ring finger-1* (*Mgrn1*) exhibit abnormal patterning of the left-right axis. Dev. Dyn. 235: 3438–3447. https://doi.org/10.1002/dvdy.20992

Dawes M., T. Moore-Harrison, A. T. Hamilton, T. Ceaser, K. J. Kochan, *et al.*, 2014 Differential gene expression in high- and low-active inbred mice. BioMed Res. Int. 2014: 1–9. https://doi.org/10.1155/2014/361048

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

Didion J. P., A. P. Morgan, L. Yadgary, T. A. Bell, R. C. McMullan, *et al.*, 2016 *R2d2* drives selfish sweeps in the house mouse. Mol. Biol. Evol. 33: 1381–1395. https://doi.org/10.1093/molbev/msw036

Dlugosz E. M., M. A. Chappell, D. G. McGillivray, D. A. Syme, and T. Garland, Jr., 2009 Locomotor trade-offs in mice selectively bred for high voluntary wheel running. J. Exp. Biol. 212: 2612–2618. https://doi.org/10.1242/jeb.029058

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

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

Garland T., 2003 Selection experiments: an under-utilized tool in biomechanics and organismal biology, pp. 23–56 in *Vertebrate biomechanics and evolution*, BIOS Scientific Publishers, Oxford, UK.

Garland, Jr. T., M. T. Morgan, J. G. Swallow, J. S. Rhodes, I. Girard, *et al.*, 2002 Evolution of a small-muscle polymorphism in lines of house mice selected for high activity levels. Evolution 56: 1267–1275. https://doi.org/10.1111/j.0014-3820.2002.tb01437.x

Garland, Jr. T., and M. R. Rose (Eds.), 2009a *Experimental evolution: concepts, methods, and applications of selection experiments*. University of California Press, Berkeley.

Garland, Jr. T., and M. R. Rose, 2009b *Experimental evolution: concepts, methods, and applications of selection experiments, Chapter 1*. University of California Press.

Garland, Jr. T., S. A. Kelly, J. L. Malisch, E. M. Kolb, R. M. Hannon, *et al.*, 2011 How to run far: multiple solutions and sex-specific responses to selective breeding for high voluntary activity levels. Proc. R. Soc. B Biol. Sci. 278: 574–581. https://doi.org/10.1098/rspb.2010.1584

Gerrits L., G. J. Overheul, R. C. Derks, B. Wieringa, W. J. A. J. Hendriks, *et al.*, 2012 Gene duplication and conversion events shaped three homologous, differentially expressed myosin regulatory light chain (MLC2) genes. Eur. J. Cell Biol. 91: 629–639. https://doi.org/10.1016/j.ejcb.2012.02.001

Girard I., J. G. Swallow, P. A. Carter, P. Koteja, J. S. Rhodes, *et al.*, 2002 Maternal-care behavior and life-history traits in house mice (Mus domesticus) artificially selected for high voluntary wheel-running activity. Behav. Processes 57: 37–50. https://doi.org/10.1016/S0376-6357(01)00206-6

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

Griffiths A. J. F., S. R. Wessler, S. B. Carroll, and J. Doebley, 2015 *An Introduction to Genetic Analysis*. W. H. Freeman & Company, New York, NY 10010.

Guderley H., P. Houle-Leroy, G. M. Diffee, D. M. Camp, and T. Garland, 2006 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. 144: 271–282. https://doi.org/10.1016/j.cbpb.2006.02.009

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

Hannon R. M., S. A. Kelly, K. M. Middleton, E. M. Kolb, D. Pomp, *et al.*, 2008 Phenotypic effects of the "Mini-Muscle" allele in a large HR x C57BL/6J mouse backcross. J. Hered. 99: 349–354. https://doi.org/10.1093/jhered/esn011

Hayashi M., K. Imanaka-Yoshida, T. Yoshida, M. Wood, C. Fearns, *et al.*, 2006 A crucial role of mitochondrial Hsp40 in preventing dilated cardiomyopathy. Nat. Med. 12: 128–132. https://doi.org/10.1038/nm1327

Hillis D. A., L. Yadgary, G. M. Weinstock, F. Pardo-Manuel de Villena, D. Pomp, *et al.*, 2020 Genetic basis of aerobically supported voluntary exercise: results from a selection experiment with house mice. Genetics 216: 781–804. https://doi.org/10.1534/genetics.120.303668

Hiramatsu L., J. C. Kay, Z. Thompson, J. M. Singleton, G. C. Claghorn, *et al.*, 2017 Maternal exposure to Western diet affects adult body composition and voluntary wheel running in a genotype-specific manner in mice. Physiol. Behav. 179: 235–245. https://doi.org/10.1016/j.physbeh.2017.06.008

Ho R. X.-Y., R. Amraei, K. O. C. De La Cena, E. G. Sutherland, F. Mortazavi, *et al.*, 2020 Loss of MINAR2 impairs motor function and causes Parkinson's disease-like symptoms in mice. Brain Commun. 2: fcaa047. https://doi.org/10.1093/braincomms/fcaa047

Holst B. D., P. W. Vanderklish, L. A. Krushel, W. Zhou, R. B. Langdon, *et al.*, 1998 Allosteric modulation of AMPA-type glutamate receptors increases activity of the promoter for the neural cell adhesion molecule, N-CAM. Proc. Natl. Acad. Sci. 95: 2597–2602. https://doi.org/10.1073/pnas.95.5.2597

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

Hung Choy Wong A., C. E. Buckle, and H. H. M. Van Tol, 2000 Polymorphisms in dopamine receptors: what do they tell us? Eur. J. Pharmacol. 410: 183–203. https://doi.org/10.1016/S0014-2999(00)00815-3

James J. W., 1970 The founder effect and response to artificial selection. Genet. Res. 16: 241–250. https://doi.org/10.1017/S0016672300002500

Johnson R. A., J. S. Rhodes, S. L. Jeffrey, T. Garland, and G. S. Mitchell, 2003 Hippocampal brain-derived neurotrophic factor but not neurotrophin-3 increases more in mice selected for increased voluntary wheel running. Neuroscience 121: 1–7. https://doi.org/10.1016/S0306-4522(03)00422-6

Jónás I., K. A. Schubert, A. C. Reijne, J. Scholte, T. Garland, *et al.*, 2010 Behavioral traits are affected by selective breeding for increased wheel-running behavior in mice. Behav. Genet. 40: 542–550. https://doi.org/10.1007/s10519-010-9359-8

Kane S. L., T. Garland Jr., and P. A. Carter, 2008 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. 81: 288–300. https://doi.org/10.1086/587093

Kaplan N. L., R. R. Hudson, and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics 123: 887–899. https://doi.org/10.1093/genetics/123.4.887

Karlsson E. K., D. P. Kwiatkowski, and P. C. Sabeti, 2014 Natural selection and infectious disease in human populations. Nat. Rev. Genet. 15: 379–393. https://doi.org/10.1038/nrg3734

Kelly S. A., P. P. Czech, J. T. Wight, K. M. Blank, and T. Garland, Jr., 2006 Experimental evolution and phenotypic plasticity of hindlimb bones in high-activity house mice. J. Morphol. 267: 360–374. https://doi.org/10.1002/jmor.10407

Kelly S. A., T. A. Bell, S. R. Selitsky, R. J. Buus, K. Hua, *et al.*, 2013 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 195: 1385–1395. https://doi.org/10.1534/genetics.113.154476

Kelly S. A., F. R. Gomes, E. M. Kolb, J. L. Malisch, and T. Garland, Jr., 2017 Effects of activity, genetic selection and their interaction on muscle metabolic capacities and organ masses in mice. J. Exp. Biol. 220: 1038–1047. https://doi.org/10.1242/jeb.148759

Kent W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, *et al.*, 2002 The Human Genome Browser at UCSC. Genome Res. 12: 996–1006. https://doi.org/10.1101/gr.229102

Kichaev G., G. Bhatia, P.-R. Loh, S. Gazal, K. Burch, *et al.*, 2019 Leveraging polygenic functional enrichment to improve GWAS power. Am. J. Hum. Genet. 104: 65–75. https://doi.org/10.1016/j.ajhg.2018.11.008

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

Kim S. K., 2018 Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture, (B. O. Williams, Ed.). PLOS ONE 13: e0200785. https://doi.org/10.1371/journal.pone.0200785

Klomberg K. F., T. Garland, J. G. Swallow, and P. A. Carter, 2002 Dominance, plasma testosterone levels, and testis size in house mice artificially selected for high activity levels. Physiol. Behav. 77: 27–38. https://doi.org/10.1016/S0031-9384(02)00767-9

Kolb E. M., E. L. Rezende, L. Holness, A. Radtke, S. K. Lee, *et al.*, 2013 Mice selectively bred for high voluntary wheel running have larger midbrains: support for the mosaic model of brain evolution. J. Exp. Biol. 216: 515–523. https://doi.org/10.1242/jeb.076000

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

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

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

Malisch J. L., W. Saltzman, F. R. Gomes, E. L. Rezende, D. R. Jeske, *et al.*, 2007 Baseline and stress-induced plasma corticosterone concentrations of mice selectively bred for high voluntary wheel running. Physiol. Biochem. Zool. 80: 146–156. https://doi.org/10.1086/508828

Malisch J. L., S. A. Kelly, A. Bhanvadia, K. M. Blank, R. L. Marsik, *et al.*, 2009 Lines of mice with chronically elevated baseline corticosterone levels are more susceptible to a parasitic nematode infection. Zoology 112: 316–324. https://doi.org/10.1016/j.zool.2008.09.004

Mathes W. F., D. L. Nehrenberg, R. Gordon, K. Hua, T. Garland, Jr., *et al.*, 2010 Dopaminergic dysregulation in mice selectively bred for excessive exercise or obesity. Behav. Brain Res. 210: 155–163. https://doi.org/10.1016/j.bbr.2010.02.016

Mayr E., 1942 *Systematics and the Origin of Species*. Columbia University Press, New York.

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

Meek T. H., B. P. Lonquich, R. M. Hannon, and T. Garland, Jr., 2009 Endurance capacity
of mice selectively bred for high voluntary wheel running. J. Exp. Biol. 212:
2908–2917. https://doi.org/10.1242/jeb.028886

Middleton K. M., C. E. Shubin, D. C. Moore, P. A. Carter, T. Garland, Jr., *et al.*, 2008
The relative importance of genetics and phenotypic plasticity in dictating bone
morphology and mechanics in aged mice: Evidence from an artificial selection
experiment. Zoology 111: 135–147. https://doi.org/10.1016/j.zool.2007.06.003

Middleton K. M., B. D. Goldstein, P. R. Guduru, J. F. Waters, S. A. Kelly, *et al.*, 2010
Variation in within-bone stiffness measured by nanoindentation in mice bred for
high levels of voluntary wheel running. J. Anat. 216: 121–131.
https://doi.org/10.1111/j.1469-7580.2009.01175.x

Munafò M., T. Clark, E. Johnstone, M. Murphy, and R. Walton, 2004 The genetic basis
for smoking behavior: A systematic review and meta-analysis. Nicotine Tob. Res.
6: 583–598. https://doi.org/10.1080/14622200410001734030

Natarajan C., F. G. Hoffmann, H. C. Lanier, C. J. Wolf, Z. A. Cheviron, *et al.*, 2015
Intraspecific polymorphism, interspecific divergence, and the origins of function-
altering mutations in deer mouse hemoglobin. Mol. Biol. Evol. 32: 978–997.
https://doi.org/10.1093/molbev/msu403

Nguyen Q. A. T., D. Hillis, S. Katada, T. Harris, C. Pontrello, *et al.*, 2020 Coadaptation
of the chemosensory system with voluntary exercise behavior in mice, (H.
Matsunami, Ed.). PLOS ONE 15: e0241758.
https://doi.org/10.1371/journal.pone.0241758

Nielsen B. L. (Ed.), 2017 *Olfaction in animal behaviour and welfare*. CABI, UK.

Noble E. P., 2003 D2 dopamine receptor gene in psychiatric and neurologic disorders and
its phenotypes. Am. J. Med. Genet. 116B: 103–125.
https://doi.org/10.1002/ajmg.b.10005

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

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

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

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

Rezende E. L., F. R. Gomes, M. A. Chappell, and T. Garland Jr., 2009 Running behavior and its energy cost in mice selectively bred for high voluntary locomotor activity. Physiol. Biochem. Zool. 82: 662–679. https://doi.org/10.1086/605917

Rhodes J. S., G. R. Hosack, I. Girard, A. E. Kelley, G. S. Mitchell, *et al.*, 2001 Differential sensitivity to acute administration of cocaine, GBR 12909, and fluoxetine in mice selectively bred for hyperactive wheel-running behavior. Psychopharmacology (Berl.) 158: 120–131. https://doi.org/10.1007/s002130100857

Rhodes J. S., H. van Praag, S. Jeffrey, I. Girard, G. S. Mitchell, *et al.*, 2003 Exercise increases hippocampal neurogenesis to high levels but does not improve spatial learning in mice bred for increased voluntary wheel running. Behav. Neurosci. 117: 1006–1016. https://doi.org/10.1037/0735-7044.117.5.1006

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

Roberts M. D., G. N. Ruegsegger, J. D. Brown, and F. W. Booth, 2017 Mechanisms associated with physical activity behavior: insights from rodent experiments. Exerc. Sport Sci. Rev. 45: 217–222. https://doi.org/10.1249/JES.0000000000000124

Rooij E. van, D. Quiat, B. A. Johnson, L. B. Sutherland, X. Qi, *et al.*, 2009 A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance. Dev. Cell 17: 662–673. https://doi.org/10.1016/j.devcel.2009.10.013

Schmill M. P., 2021 Neuroanatomical, behavioral, and physiological correlates of high voluntary wheel running

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

Simões P., J. Santos, I. Fragata, L. D. Mueller, M. R. Rose, *et al.*, 2008 How repeatable is adaptive evolution? The role of geographical origin and founder effects in

laboratory adaptation. Evolution 62: 1817–1829. https://doi.org/10.1111/j.1558-5646.2008.00423.x

Smith J. M., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23. https://doi.org/10.1017/S0016672300014634

Smith J. R., G. T. Hayman, S.-J. Wang, S. J. F. Laulederkind, M. J. Hoffman, *et al.*, 2019 The Year of the Rat: The Rat Genome Database at 20: a multi-species knowledgebase and analysis platform. Nucleic Acids Res. gkz1041. https://doi.org/10.1093/nar/gkz1041

Smits P., P. Li, J. Mandel, Z. Zhang, J. M. Deng, *et al.*, 2001 The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. Dev. Cell 277–290. https://doi.org/10.1016/s1534-5807(01)00003-x

Stinchcombe J. R., C. Weinig, K. D. Heath, M. T. Brock, and J. Schmitt, 2009 Polymorphic genes of major effect: consequences for variation, selection and evolution in *Arabidopsis thaliana*. Genetics 182: 911–922. https://doi.org/10.1534/genetics.108.097030

Storz J. F., 2016 Gene duplication and evolutionary innovations in hemoglobin-oxygen transport. Physiology 31: 223–232. https://doi.org/10.1152/physiol.00060.2015

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

Swallow J. G., P. Koteja, P. A. Carter, and 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. https://doi.org/10.1242/jeb.202.18.2513

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

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

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

Tenaillon O., A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, *et al.*, 2012 The molecular diversity of adaptive convergence. Science 335: 457–461. https://doi.org/10.1126/science.1212986

The International Mouse Phenotyping Consortium, M. E. Dickinson, A. M. Flenniken, X. Ji, L. Teboul, *et al.*, 2016 High-throughput discovery of novel developmental phenotypes. Nature 537: 508–514. https://doi.org/10.1038/nature19356

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

Wallace I. J., K. M. Middleton, S. Lublinsky, S. A. Kelly, S. Judex, *et al.*, 2010 Functional significance of genetic variation underlying limb bone diaphyseal structure. Am. J. Phys. Anthropol. 143: 21–30. https://doi.org/10.1002/ajpa.21286

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

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

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

Yoshida M., S. Minamisawa, M. Shimura, S. Komazaki, H. Kume, *et al.*, 2005 Impaired Ca2+ Store Functions in Skeletal and Cardiac Muscle Cells from Sarcalumenin-deficient Mice. J. Biol. Chem. 280: 3500–3506. https://doi.org/10.1074/jbc.M406618200

Zheng X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, *et al.*, 2012 A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28: 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Chapter 3

Detected selection signatures change beyond a selection limit

David A. Hillis, Lei Yu, Zhenyu Jia, Liran Yadgary, George M. Weinstock, Fernando Pardo-Manuel de Villena, Daniel Pomp, Frank Chan, and Theodore Garland, Jr.

ABSTRACT

Following a change in selective regime, selection is expected to increase the quality of adaptation across generations, which should lead to changing genetic selection signatures when one compares an adapted population with one that has not faced similar selection. In laboratory selection experiments with relatively small population sizes, random genetic drift may have increasingly adverse effects on the statistical power to detect selection signatures. The purpose of the present study was to compare selection signatures in the "High Runner" selection experiment, which involves voluntary exercise of laboratory mice with 4 replicate lines bred for wheel-running behavior (HR) and an additional 4 replicate non-selected control (C) lines. Previously, we reported multiple regions of differentiation between the HR and C lines, based on whole-genome sequence data for 10 mice from each line sampled at generation 61, which was nearly ~30-35 generations after selection limits had been reached in all of the HR lines. Thirteen of these regions were consistently identified in three separate types of analysis.

Here, we analyzed pooled sequencing data from approximately 20 mice for each of the 8 lines at generation 22, around the time the HR lines were reaching selection limits. Differentiation analyses of the allele frequencies at ~4.4 million SNP loci used the regularized T-test, and results were then compared to those from generation 61, also treated as pooled.

Analyses of generation 22 detected a total of 436 unique differentiated regions with FDR = 0.01. Comparable analyses involving pooling generation 61 individual mouse genotypes into allele frequencies by line only produced 21 such differentiated

regions. Furthermore, although regions identified with the pooled generation 61 analyses did commonly overlap with the 436 regions at generation 22, the strongest selection signatures for regions identified in either generation (based on p-value, number of differentiated SNP loci, and number of statistical tests confirming differentiation) were, at best, only weakly identified in the other. "Strict" culling methods were applied to the 436 regions detected at generation 22, reducing these to 21 regions (this $N = 21$ is coincidental with the number detected at generation 61 noted above).

Given these results, we conducted simulations to test the hypothesis that constraints on a trait under positive selection could increase differences in detected selection signatures before vs after a selection limit. As expected, including such a constraint reduced the ability to detect differentiated loci and increased the chance that detectable selection signatures would change over the course of several generations. However, the magnitude of this effect was not large enough effect to produce the ~20-fold increase in selection signatures detected at generation 22 vs 61. The simulations did show that the ability to detect loci with large effect size was much greater at generation 22 than 61, but the increase in detection rate for low-effect size loci between generations 22 and 61 resulted in more total selection signatures expected at generation 61.

The 21 differentiated regions identified at generation 22 with strict culling measures include 355 genes related to a wide variety of functions. Gene ontology using ToppGene identified pathways related to energy homeostasis and dopamine as being uniquely overrepresented among these top differentiated regions. Genes included in the differentiated regions that could not be analyzed by ToppGene include those related to

olfactory and vomeronasal systems, consistent with previous identification of such

systems at generation 61.

INTRODUCTION

Although evolution can result in organisms with spectacular capabilities or able to survive in exceptionally inhospitable environments, all adaptations are bound within certain limits. These limits are commonly observed in laboratory and agricultural selection experiments (Dobzhansky and Spassky 1969; Al-Murrani and Roberts 1974; Careau *et al.* 2013; Schlötterer *et al.* 2015; Lillie *et al.* 2019). Among various possible causes of selection limits (Al-Murrani and Roberts 1974; Falconer 1989; Douhard *et al.* 2021), the simplest explanation is the loss of genetic variation, such that narrow-sense heritability declines to zero (e.g., Brown and Bell 1961). However, selection experiments have frequently found that genetic variation remains after reaching a selection limit (e.g., Lerner and Dempster 1951; Roberts 1966; Dobzhansky and Spassky 1969; Bult and Lynch 2000; Burke *et al.* 2010; Careau *et al.* 2013; Lillie *et al.* 2019; Hillis *et al.* 2020). Even for alleles favored by selection, fixation is far from guaranteed (Burke *et al.* 2010; Schlötterer *et al.* 2015; Stephan 2016; Hillis *et al.* 2020).

One selection experiment that has continued selection long after reaching a limit is the High Runner (HR) mouse experiment, which started in 1993 with the purchase of 224 outbred ICR mice from Harlan Sprague Dawley (Swallow *et al.* 1998a). These were randomly bred for two generations, then split into ten breeding pairs to found each of eight closed lines. Four of these lines were designated to serve as non-selected control lines, while the other four were selected based on voluntary wheel running. In selected lines, all mice are given access to wheels for 6 days and the male and female of each family with the highest running on days 5 and 6 would be used as breeders (no sib-

mating). After about 22 generations of selection, three of the four HR lines (with the fourth line following suit a few generations later) had plateaued in their running at approximately 2.5 to 3 times as many revolutions as the controls (Careau *et al.* 2013). Recently, the experiment has reached its 100th generation since selection began and, with exception of some generations when the experiment moved from Wisconsin to California (generations 32 to 35) and during Covid-19 lockdowns (generations 91 to 98), selection has continued nearly uninterrupted in the interim.

Numerous physiological and morphological differences between the HR and control lines have been documented (Rhodes *et al.* 2005; Swallow *et al.* 2009; Garland, Jr. *et al.* 2011a; Wallace and Garland, Jr. 2016). These include traits associated with motivation to run, such as changes in dopamine (Rhodes *et al.* 2001; Mathes *et al.* 2010), serotonin (Waters *et al.* 2013), and endocannabinoid signaling (Thompson *et al.* 2017), as well as changes in brain size and structure (Kolb *et al.* 2013a). Additionally, changes associated with ability to run have been found, including endurance capacity (Meek *et al.* 2009), maximal aerobic capacity (VO_{2Max}) (e.g., Swallow *et al.* 1998b; Kolb *et al.* 2010; Dlugosz *et al.* 2013; Hiramatsu *et al.* 2017; Cadney *et al.* 2021; Castro et al. in revision), heart size (Kolb *et al.* 2010, 2013b; Kelly *et al.* 2017), skeletal muscle physiology (Dumke *et al.* 2001; Syme *et al.* 2005; Guderley *et al.* 2008), and bone morphology (Garland, Jr. and Freeman 2005; Kelly *et al.* 2006; Middleton *et al.* 2008, 2010; Wallace *et al.* 2010, 2012; Castro and Garland, Jr. 2018; Copes *et al.* 2018; Schwartz *et al.* 2018).

Previously, whole-genome differentiation analyses using individual mouse data from 10 males from each of the eight lines at generation 61 identified at least 13 genomic

regions differentiated between the control and HR lines (Hillis *et al.* 2020; Hillis and

Garland Jr 2022). Within these regions were genes associated with development of the

brain, heart, bones, and limbs, in addition to reward pathways, and even the vomeronasal

system (see also Nguyen *et al.* 2020). Dropping individual lines from analyses revealed

new potential signatures of selection and demonstrated that the HR lines have evolved in

different ways at the genomic level ("multiple solutions" Garland, Jr. *et al.* 2011b) that

increase wheel-running behavior (Hillis and Garland Jr 2022). Despite being ~30-35

generations past the selection limit, a great deal of genetic diversity remained in all 8

lines including many regions identified as differentiated between the HR lines and

controls.

With the selection limit achieved near generation 22, one might expect many if

not most biologically relevant SNPs to already be differentiated by that generation. Thus,

with respect to the ability to detect selection signatures, little advantage would be gained

from allowing ~30-35 generations to pass before testing for allelic differentiation

between the HR and control lines. Furthermore, simulations performed by Baldwin-

Brown et al. (2014) demonstrate that increasing the number of generations could reduce

power to detect some loci under selection, which they attributed to noise created by

random genetic drift. Reasonably, one might expect that drift over enough generations

may cause control lines to diverge from each other in allele frequencies, such that

selection signatures are obscured in statistical tests that compare replicate sets of selected

and control lines. For example, if some control lines become fixed for one allele and the

remaining control lines become fixed for another, then, even if all HR lines were fixed

for the same allele favored by selection, statistically significant differentiation would be difficult to detect. Therefore, analyses of a generation close to when a selection limit is first reached would be optimal for tests of genetic differentiation.

In the present study, we analyze pooled sequence data from each of the four HR lines and four control lines at generation 22. Although these analyses identify many regions containing genes associated with systems known to be phenotypically differentiated between the HR and control linetypes, they largely differ from those previously identified with the generation 61 individual mouse sequence data (Hillis *et al.* 2020). Furthermore, the number of differentiated regions detected at generation 22 are more than 20-fold greater than those detected with generation 61 data (treated as pooled data).

We first discuss possible methodological causes of these differences (e.g., pooled vs individual mouse data) and find them lacking. We therefore develop a simple simulation model, with leptokurtic distribution of locus effect sizes, to test the possibility that a hypothetical physiological constraint on wheel running could contribute to the differences between generations 22 and 61 selection signatures. Ignoring locus effect size, results demonstrate that such constraints can contribute to a reduction in power and increased variability in the detected response to selection in generations after the selection limit. However, the magnitude of these effects appears insufficient to explain the differences observed between generations 22 and 61 in the real data. In addition, effect size was an important determinant of the ability to detect selection signatures in the simulations, including a more than 2-fold increase in power to detect loci with large

effect size at generation 22 as compared to generation 61. Thus, with strict culling

procedures, we suspect that many of the selection signatures detected at both generations

are likely to represent loci with relatively large effects on wheel running. The regions

detected at generation 22 include genes related to reward pathways (GABAergic),

metabolism, as well as posttranscriptional and translational processing. Genes related to

olfactory/vomeronasal systems are also identified, as they were at generation 61 (Hillis *et*

al. 2020; Nguyen *et al.* 2020; Hillis and Garland Jr 2022).

METHODS

High Runner Mouse Model

As described previously (Careau *et al.* 2013; Swallow *et al.* 1998a), 112 males and 112 females of the outbred Hsd:ICR strain were purchased from Harlan Sprague Dawley in 1993 and designated as generation -2. Mice would be randomly bred for 2 generations (-2 and -1) with 2-3 generation -1 mice from each family randomly chosen to contribute to 1 of 8 different closed lines. Four of these lines were randomly picked to be "High Runner" (HR) lines, in which mice would be selected for breeding based on voluntary wheel running. The remaining 4 lines were used as Control (C) lines, without any selection. Generation 0 was the first generation where HR lines were paired based on running levels (10 males and 10 females for each line) with generation 1 the first product of selection.

Wheel running measurements were collected by giving mice at approximately 6-8 weeks of age, access to wheels for six days. The amount of running (total revolutions) on days 5 and 6 was used as the selection criterion. For the HR lines, the highest-running male and female from within each of 10 families were chosen as breeders (within-family selection). For the non-selected C lines, one male and one female from each of 10 families were chosen as breeders, independent of wheel running measurements. Sib-mating was disallowed in all lines (Swallow *et al.* 1998a).

Genome Sequencing and Allele Frequency Determination

Roughly 10 male and 10 female mice were taken from each line at generation 22 and their DNA was pooled for determination of allele frequency for each line. This pooled DNA was sequenced with paired end pooled sequencing with Illumina HiSeq 2500 sequences were trimmed and aligned to the GRCm38/mm10 mouse genome assembly. Generation 22 used trimmomatic v0.39 for trimming, BWA v0.7.17 for alignment, Samtools v1.14 for sorting and indexing, picard v2.26.11 for marking duplicates, and GATK v4.1.8.1 for calling SNPs. SNPs were filtered to keep those with read quality ("RQ") $\geq$ 20, DP $\geq$ 10, were missing either quality score, or missing the allele frequency all together, or had MAF > 0.0126. Allele frequencies ("AF") were determined for generation 22 by taking the read depth of the alternate nucleotide allele (i.e., allele differing from the GRCm38/mm10 alignment) and dividing by the read depth for the locus. After all quality control methods were implemented, 4,446,523 loci remained for generation 22.

The generation 61 data were taken from Hillis et al. (2020). 80 male mice (10 from each line) were subject to whole genome sequencing 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 12X per mouse. SNPs were filtered to keep those with genotype quality ("GQ") > 5, read depth ("DP") > 3, minimum allele frequency ("MAF") > 0.0126 for all samples, and Mapping Quality ("MQ") > 30. One of the 80 mice 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 two of the

79 mice were also removed from analysis. After all quality control methods were implemented, 5,932,148 loci remained for analyses. To allow comparison with the pooled sequencing data from generation 22, we calculated allele frequencies as the number of alternative alleles divided by 2 times the number of mice (i.e., 20 or 18 for HR3).

Statistical Analyses

For generations 22 and 61 we used an arcsine-squared transformation (Ahrens *et al.* 1990) of the AF. Analyses were conducted on both generations using a traditional T-test, regularized T-test (RegT)(Baldwin-Brown *et al.* 2014, see also Baldi and Long 2001), and a variant of the regularized T-test which uses a sliding window to calculate $\bar{v}$ (WRT test) (Supplemental File S1). The regularized T-test was based on a Bayesian method meant to minimize the type-I errors caused by sampling error with small sample sizes (Baldi and Long 2001; Baldwin-Brown *et al.* 2014), such as the 8 total lines in the HR mouse selection experiment. We performed these tests and determined the permutation-based false discovery rate (FDR) for each method (see below). For comparison, we also performed the RegT and WRT tests on loci found in both generation 22 and 61 (from pooling individual mouse genotypes) data sets along with the FDR. Since standard T-tests do not require whole genome or region variances of other loci, the p-values of loci shared between the two generations could simply be extracted from the complete original analyses.

Permutation-Based False Discovery Rate

To determine relative power generation 22 allele frequencies with arcsine-square transformation using T-test, regularized T-test, and WRT test, we attempted to calculate a critical threshold by estimating the FDR of 10% (Benjamini and Hochberg 1995; Xie *et al.* 2005). However, after calculating p-values for complete permutations of the different lines within linetype to better understand the null distribution, we concluded that this estimated FDR was underestimating the true false discovery rate. Therefore, using these same permutations, we calculated the FDR directly.

Direct calculations of FDR were performed by calculating FDR for each locus of the unpermuted data whose p-value was below 0.01 in accordance with the equation:

$$FDR = \frac{n\ False\ Positives}{n\ rejected\ Null\ Hypotheses}$$

This was implemented for each locus with:

$$FDR = \frac{\frac{n\ permuted\ loci\ significant\ at\ p}{35}}{n\ unpermuted\ loci\ significant\ at\ p}$$

Loci with nominal p-value < 0.05 were ordered by FDR score, the p-value was identified for the locus with the largest FDR below 0.01, and any p-values less than or equal to the p-value for this locus was treated as significant. The SNPs with FDR = 0.01 were then further grouped into "significant regions" by grouping any loci within 1mbp of another and separating groups whose closest SNPs are further than 1mbp.

"Strict" Culling for Biological and AF Change Analyses

Rather than attempt to focus on the genes of more than 100 regions for each of the different statistical tests, analyses of biological significance and comparisons of change in allele frequencies between generations 22 and 61 were done using a subset of the regions identified by FDR. WRT and regularized T-test first culled by removing regions containing only one significant locus, then culled such that only regions containing at least 20 significant loci or the lowest p-value among loci in the region was below 1.00E-04. Regions associated with the T-tests were culled in a similar manner as the WRT and RegT test, except the p-value cutoff used was 1.00E-06 due to naturally lower p-values. These culling methods should also serve to reduce the influence of sampling error, as it would be increasingly unlikely for sampling error to simultaneously underestimate among-line variance across multiple linked SNPs and lines. We will refer to these additional culling methods below as "strict" culling.

Comparison of Selection Signatures in Generations 22 and 61

Changes in allele frequencies from generation 22 to 61 were analyzed for each region identified by strict culling for generations 22 and 61. For regions significant at generation 22, each region and its included SNPs with nominal p<0.05 at generation 22 were matched with SNPs at generation 61. The allele frequencies of these SNPs were averaged for each line and generation and line graphs created (one for each line) with generation 22 AF on the left and generation 61 AF on the right. This was then repeated

for regions significant at generation 61, except each region and its included SNPs with nominal p<0.05 at generation 61 were matched with SNPs at generation 22.

Simulations to Compare Presumptive Statistical Power Across Generations

The available data from the two generations differ in multiple ways that might affect cross-generation comparisons of selection signatures. Each generation, each line is reduced to ~20 individuals when ~10 breeding pairs are formed. An ideal "sample" from a given generation would include all 20 of those breeding individuals. Instead, our sample from generation 22 was of ~10 males and 10 females per line that were sampled at random at the time of weaning (i.e., they were not the 20 breeding parents). In contrast, the mice from generation 61 were a semi-random sample of 10 males from each line (except nine from HR3 and one female that was unintentionally used from another line) (Hillis *et al.* 2020).

For a pooled DNA sample, as for generation 22, a further ideal condition is for the sample of DNA from each mouse to be of equal volume and concentration through the extraction and pipetting steps prior to pooling. This would then result in each mouse's alleles being represented in equal quantities in the pooled sequencing sample.

The next source of error is read depth, which is effectively a random sampling of alleles from the pooled sample. Our generation 22 samples were read at an average depth of 24X. Thus, the frequency of alternative nucleotide alleles for a given SNP locus was calculated by counting the number of alternative alleles, which was taken as anything other than the reference. Thus, not all of the 40 alleles (as one of two possibilities)

contributed by the 20 mice could have been identified with a read depth of 24X, which acts as 24 samples taken with replacement.

The generation 61 data are from individual sequencing of 10 mice per line at an average read depth of 12X, with those results then used to predict the genotype for each SNP and mouse (Hillis *et al.* 2020). This should allow for the representation of nearly all alleles (N = 2 alleles x 10 mice). Originally, those data were analyzed as such via mixed models to detect selection signatures (Hillis *et al.* 2020). Here, to allow comparison with the pooled sequencing data from generation 22, we calculated allele frequencies as the number of alternative alleles divided by 2 times the number of mice (i.e., 20 or 18 for HR3), which should incorporate 19-20 unique alleles in equal proportion. Given that the data available from the two generations differ in multiple ways, we used simulations in an attempt to assess how this might affect our results.

For generation 22, simulations to elucidate possible sampling errors were performed such that alleles for 20 mice were sampled using a random binomial distribution assuming population allele frequencies of (0.05, 0.10, 0.15, …,0.90, and 0.95). Then an allele depth was randomly sampled from the actual quality data for the SNPs used in the generation 22 analyses and alleles were sampled from these simulated 20 mice (with replace) equal to this read depth. The allele frequency was then calculated as the number of alternative alleles (1) divided by the total read depth. This generated a distribution of allele frequencies given a particular starting AF for the population and was repeated 100,000 times for each starting population AF.

For generation 61, simulations were performed such that alleles for 10 mice were sampled using a random binomial distribution assuming population allele frequencies of (0.05, 0.10, 0.15, …,0.90, and 0.95). Then for each simulated mouse's genotype, a genotype quality was randomly sampled from the actual quality data for the SNPs used in the generation 61 analyses. If the simulated genotype for the mouse was heterozygous, then the genotype quality would be used to generate a 0 or 1 with the probability of a 1 equaling that of the probability of a genotyping error. If a 1 was generated (thus an error occurred) the second allele for the mouse was replace with a copy of the first allele of the mouse. The allele frequency was then calculated as the number of alternative alleles (i.e., 1) for all ten mice divided by the total alleles (i.e., 20). This generated a distribution of allele frequencies given a particular starting AF for the population and was repeated 100,000 times for each starting population AF.

Power analyses were then done by sampling four AF values from the simulated AF values from an actual population AF of 0.4 for one linetype. Likewise, four AF values were sampled from the simulated AF values from an actual population AF of 0.6 for the other linetype. Sampled allele frequencies were transformed using an arcsine-squared transformation. A T-test (assuming unequal variance) was then conducted comparing these 8 sampled AF values. Note that this could not be done for RegT and WRT tests because it would require simulations of regional or genome-wide variance structure. These sampling and T-tests were repeated 10,000 times.

Simulations Comparing Power With and Without a Biological Constraint

We used simple simulations to begin to address whether a biological constraint on a trait under selection (e.g., wheel running) might affect (1) the ability to detect selection signatures at generations before (e.g., generation 22) versus long after (e.g., generation 61) selection limits were reached, (2) the consistency of those signatures between generations, and (3) the rate at which loci with different allelic effect sizes respond to selection. Our rationale for using a constraint model is explained in the Discussion. As a heuristic, some of the parameters in these simulations were chosen to approximate values observed in the selection experiment.

Running levels were calculated based on the general equation:

$$y = \mu + v_g + v_e$$

where y is equal to the phenotype (wheel revolutions/day) of an individual mouse; μ is the "base" mean number of revolutions (held constant at the starting value set at generation 0); v_g is the variance contributed by genetic variation; and v_e is the variance contributed by environmental effects.

As a regression model, this equation is:

$$y = \mu + \beta_1 X_1 + \beta_2 X_2$$

where the genetic variance is represented by $\beta_1 X_1$ and the environmental variance is represented by $\beta_2 X_2$ (see below).

The equation we applied for these simulations is:

$$y = 4{,}400 + 1.35X_1 + 1{,}650X_2$$

The mean of 4,400 (revolutions/day) was picked to approximate the empirically determined starting running levels at generation 0 (Swallow *et al.* 1998a). X_1 represents the summed effect on wheel running of all alleles carried by the individual, where, to simulate a leptokurtic distribution (Barton and Turelli 1989; Reeve 2000; Reeve and Fairbairn 2001), these alleles are coded as having variable allelic affects (specifically, ±0.4, ±0.8, ±1.6, ±3.2, ... ±204.8) at frequencies inversely proportional to their effect size (specifically, 720 loci with effect ±0.4, 480 loci with effect ±0.8, … 8 loci with effect ±204.8) for a total of 2,096 loci. X_2 is determined by randomly sampling from a normal distribution with mean = 0 and SD = 1. The values for β_1 (1.35), β_2 (1,650), and the number of loci (N = 2,096), were determined in conjunction with one another to approximate realistic (in no particular order) (1) heritability of wheel running at the base generation being about 0.45 (Careau *et al.* 2013), (2) within-line coefficients of variation as being about 0.55 (Swallow *et al.* 1998a), and (3) realistic response to selection in the HR lines (i.e., achieving ~16,000 revolutions around generation 22)(Careau *et al.* 2013). In addition, 2,096 approximates the number of haplotype blocks observed across all eight

lines (Hillis *et al.* 2020). Any running level calculated as below 100 was set to 100. The maximum wheel-running for unconstrained simulations was 50,000 revolutions, which is nearly twice as high as has ever been observed in actual measurements from the selection experiment (Rhodes *et al.* 2003; Careau *et al.* 2013) and the highest running levels produced by the simulations was 30,413.

For the starting population of any given line, two alleles were first assigned to each of the 2,096 independently segregating starting loci for 20 mice (based on the actual selection procedures: Swallow *et al.* 1998a) using a random binomial distribution with p = 0.5. For control lines, mice were paired, and alleles sampled from each of the pair to produce two male and two female offspring (to match the number of mice that are typically wheel-tested in the selection experiment). The first of each sex for each family was then chosen to contribute to the next generation, which is functionally equivalent to the selection experiment, where breeders are chosen a random within family and sex for control lines. For HR lines, alleles were sampled from the parents for each of five males and five females. Running distances were then calculated for all offspring and the male and female with the highest running levels within each family were selected to breed for the subsequent generation. For both linetypes, siblings were barred from pairing (again, following the selection experiment).

Simulations were run for 61 generations and alleles for all breeding pairs were saved at generation 0 and every 5 generations through 60, as well as generations 22 and 61. This was then repeated for 4 control lines and 4 HR lines. For the constrained simulation, all HR mice that ran more than 16,000 revolutions were treated as equal and

if multiple mice within a sex and family reached this threshold, then the breeder for the next generation was picked at random from those or above 16,000 revolutions. These simulations were repeated 100 times assuming no constraint and 100 times with the constraint. T-tests assuming unequal variance between the 4 control lines and the 4 HR lines were performed at each of these "saved" generations (0, 5, 10, etc.) for the allele frequencies at each locus, with an arcsine-squared transform (Ahrens *et al.* 1990). Power was then calculated for each simulation at each saved generation by dividing the number of loci with $p \leq 0.05$ by the total number of loci (N = 2,096).

Standardized selection differentials are calculated as was done by Careau et al. (2013), by subtracting from the mean running for each sex and family the running level of the bred individual from that litter and dividing the difference by the standard deviation of the sex for that litter. Relative power under the constrained and unconstrained models was calculated using unpaired T-tests (unequal variance) for each saved generation. Relative power across generations was also calculated using unpaired T-tests (unequal variance), separately for constrained and unconstrained simulations. Relative consistency in detected selection signatures was calculated by first identifying the specific significant loci (at a nominal a = 0.05) at generations 22 and 61 in each simulation. Then the percentage of loci found significant at generation 22 that remained significant at generation 61 was calculated. Unpaired T-tests (unequal variance) were performed comparing these percentages for the constrained simulations versus the unconstrained simulations. Lastly, ability to detect loci with different effect sizes was compared using a T-test (unequal variance) of generation 22 constrained vs

unconstrained models, generation 61 constrained vs unconstrained models, constrained generation 22 vs generation 61, and unconstrained generation 22 vs generation 61. All graphs and estimates which require the calculation of a mean value, missing values are excluded from the calculations, which can result in high variation (most visible in heritability estimate, see Results).

Analyses were performed again implementing possible sampling error calculated by the previous simulations. This was implemented by taking the actual allele frequency for each line at generations 22 and 61 in the simulations using the constraint model. These allele frequencies were then replaced with an allele frequency sampled from the results of nearest population allele frequency of the sampling error simulations (i.e., 0.05, 0.10, 0.15... 0.95). For example, if the allele frequency for a given line at generation 22 (constraint model) was 0.25, then this 0.25 would be replaced by a randomly sampled estimated allele frequency from the sampling error simulations (generation 22) where 0.25 was the actual population allele frequency. Generation 61 allele frequencies were similarly replaced using the results of the generation 61 sampling error simulations.

Possible Biological Function of Generation 22 "Strict" Differentiated Regions

To identify genes and associated knockout (KO) phenotypes, gene annotation information was collected from NCBI (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/GENE_INFO/Mammalia) on April 22, 2022. A list of genes was extracted from the "gene info" file using R libraries AnnotationDbi,

VariantAnnotation (Obenchain *et al.* 2014), and GenomeInfoDb (Arora *et al.* 2020).

Knockout phenotypes were then determined using MGI Batch Query

(http://www.informatics.jax.org/batch/) on June 13, 2022 (Smith and Eppig 2009).

The KO phenotypes associated with genes in our "strict" culling regions were

categorized based various biologically relevant terms (e.g., cardiac, skeletal muscle,

lung). The relative proportion of genes with such relevant terms (as compared to the total

number of genes) was then compared to the proportion of genes with these same terms

when considering all genes in the MGI Batch Query database. A chi-squared test was

then used to identify which of these terms are overrepresented among the generation 22

regions and/or the generation 61 regions.

Gene ontology was performed by submitting the genes from "strict" culling

regions (same as those used in KO phenotypes) to ToppGene

(https://toppgene.cchmc.org/enrichment.jsp) on October 26, 2022 (Chen *et al.* 2007).

Data Availability Statement

Generation 61 data were made available by Hillis et al. (2020) and can be found at

https://doi.org/10.25386/genetics.12436649. Generation 22 fastq files are available on

the SRA database, accession = PRJNA758905

(https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA758905). Supplemental files were

submitted with the current dissertation and are available via ProQuest ETD.

Supplemental File S3.1 contains brief explanation of regularized and windowed

regularized T-test. Supplemental Table S3.1 contains Type I error rates for loci with

different effect sizes. Supplemental Table S3.2 contains power for loci with different effect sizes under different constraint models. Supplemental Table S3.3 contains all regions identified as differentiated at generation 22 (FDR = 0.01). Supplemental Table S3.4 contains all genes found in the "strict" culled generation 22 regions. Supplemental Table S3.5 contains proportions of genes found in "strict" culled regions with key knockout phenotypes. Supplemental Table S3.6 contains complete ToppGene ontology results for genes found in the "strict" culled generation 22 regions. Supplemental Figure S3.1 shows the change in power for loci with different effect sizes (constrained model simulation) over multiple generations. Gen22_analyses.Rmd contains all relevant R code used for data processing, analyses, and visualization of generation 22 data. Constraint_sim3.Rmd contains all relevant R code used for the constraint simulations and their analyses and visualization.

RESULTS

Genetic Variation = Basic Facts and Figures

The number of variable loci used in the present study includes 4,446,523 for generation 22 and 5,932,148 for generation 61. Generation 61 data had an average read depth of 12X per mouse for 10 mice in each of the 8 lines, producing an average read depth of over 100 per line for detection of many more variable SNPs in each line. The overlap of base positions between generations 22 and 61 was 2,045,546 SNPs.

Differentiated SNPs and Chromosomal Regions

For analyses containing all generation 22 loci (N = 4,446,523), all three methods indicated a substantial number of differentiated loci, ranging from 917 to 1,367 (Table 3.1) based on 0.01 FDR. These loci fall into 436 unique regions (separated by at least 1 million base pairs) across all three tests, with many of the most substantial peaks overlapping among the three analysis methods (Table 3.2). The standard T-tests almost always produced lower p-values than the RegT tests and WRT tests (Figures 3.1A and 3.2C). Results for the WRT-tests were similar to those for the regularized T-tests (Pearson's r = 0.9997 with arcsine-square transform) (Figure 3.1B), though several of the most differentiated loci appear to have slightly lower p-values for the former.

At generation 61, the three analysis methods indicated from 634 to 2,251 loci (Table 3.1) based on 0.01 FDR. Although identifying similar numbers of loci as the generation 22 analyses, P-values for individual SNPs for generations 22 and 61 show little similarity (Figure 3.1D-F), with arcsine-square transform Pearson's r = 0.110, r =

0.115, and r = 0.116 for the T-test, regularized T-test, and windowed regularized T-test, respectively. For comparison to the Manhattan plots (see below), correlations with the -$\log_{10}$ transform are also included (Figure 3.1G-L). Given the normality of the p-value distributions with the arcsine-square transform (Figure 3.1N-P), we will focus on the correlation estimates for this transform. Ultimately, the SNPs identified at generations 22 and 61 were largely different. Moreover, the SNPs identified at generation 61 clustered into only 21 unique regions, as compared with the 436 regions for generation 22 (Figure 3.2).

Given such notable differences between the SNPs and regions implicated by generation 22 and 61 analyses (Table 3.2), analyses were repeated focusing only on the loci found in both data sets (N = 2,045,546). With fewer loci being analyzed, fewer significant SNPs were identified at FDR = 0.01, as well as fewer regions for all analyses except for WRT with generation 61. The total peaks identified when using only the shared SNPs includes 322 and 14 regions for generations 22 and 61, respectively.

Although X chromosome data were not available for the generation 61 analyses, data for the X chromosome were available in generation 22. One SNP in particular (chrX:100,735,252) was identified by all 3 tests and is located in the Gdpd2 gene associated with osteoblast differentiation and growth (Yanaka *et al.* 2003; Corda *et al.* 2009).

Regions After "Strict" Culling

Using all available SNPs for generation 22, after applying "strict" culling (see Methods), the remaining regions were reduced to 14, 11, and 13 for the T-tests, RegT tests, and WRT tests, respectively. Seven of these regions were shared by all three tests and 3 were shared only between the regularized test and its windowed variant. Despite the differences among the three analyses, all of the regions implicated by individual analyses included or were near genes with intuitive implications for running behavior (see Discussion). Therefore, we chose to include all of the regions implicated by these analyses (N = 22) for both comparisons to generation 61 and biological interpretations of the response to selection. For generation 61, strict culling reduced the total peaks to only 6 unique regions across all three analyses.

Despite the HR lines reaching selection limits around generation 22 or shortly thereafter (Careau *et al.* 2013), the most differentiated 22 regions based on the three tests (Table 3.2) have little fixation. Of the SNPs in these 22 regions (N = 166,946), only about 11.95% are fixed in the HR lines, which is not significantly different from the 12.31% fixed in the control lines (unequal variance t-test comparing % fixed in the 4 HR versus 4 C lines: $t = -0.389$, $df = 4.348$, p-value = 0.7156). If we repeat this fixation comparison for the loci shared between generations 22 and 61 (N = 82,019), 3.00% are fixed in the HR lines, which is still not significantly different from the 2.47% fixed in the control lines ($t = -0.432$, $df = 3.267$, p-value = 0.6928).

Comparison of Selection Signatures at Generation 61 for Individual vs. Pooled Sequencing Data

Originally, the generation 61 individual mouse data were analyzed using mixed models (Hillis *et al.* 2020). We compared the previously published p-values from those analyses with the p-values produced after pooling data by line and analyzing by T-test, RegT, and WRT tests (Figure 3.3). The mixed model analyses produced lower p-values in general, as would be expected due to loss of power with pooling (Xu and Garland 2017), with the difference being greater for lower p-values. As a result, fewer SNP loci and hence fewer chromosomal regions were identified as significantly differentiated between the HR and C lines with pooled data. Of the total regions detected with FDR = 0.01, 8 were found with each of the analyses (T-test, RegT, and WRT) that matched the 13 "consistent" regions identified with the mixed model analyses (Hillis *et al.* 2020). The 5 consistent regions that were not identified by analyses of the pooled data tended to have relatively large p-values for individual SNP loci or cover a narrower area of the genome, as compared with the other 8 consistent regions.

Simulations to Compare Presumptive Statistical Power Across Generations

Simulations were conducted to gauge how much the allele frequencies determined through sequencing reflect allele frequencies of the actual populations at generations 22 and 61. Generation 22 allele frequencies have greater variance from the actual population AF than generation 61 (see Figures 3.4A-B for an example of the 0.5 population AF distribution). The greater error variance in generation 22 is associated

with reduced statistical power of 0.3864 versus 0.5031 for generation 61 when comparing

simulated allele frequencies of 0.4 and 0.6 (Figures 3.4C-D).

Simulations Comparing Power With and Without a Biological Constraint

Simulations were performed modeling response to selection assuming either a constraint

at 16,000 revolutions per day or no such constraint (see Methods). For both constrained

and unconstrained simulations, wheel running for HR and control lines diverge

recognizably at least by generation 6 (Figure 3.5A). The replicate HR lines for

unconstrained and constrained models appear fairly similar for earlier generations

(Figures 3.5B and C, respectively), presumably because mice are not widely achieving

constrained running levels. As expected, the among-line variation for control lines

increases gradually across generations. For the HR lines under the unconstrained model,

among-line variance does not increase to a noticeable extent, whereas the constrained

model shows a large reduction in among-line variance after the constraint is reached.

The calculated heritability (slope of the regression of offspring [generation 1] on

midparent [generation 0]) for all 200 simulations for control lines indicate that our

parameters resulted in a narrow-sense heritability of about 0.4883 (N=8,000 families).

For individual lines, the estimated heritability for successive generations was highly

variable, as would be expected with such small sample sizes (10 families/line). However,

the means clearly indicate a slow loss of heritability in the control lines and a more rapid

loss in the HR lines for the unconstrained model (Figure 3.6). For the constrained model,

given how the constraint is implemented, by later generations we had parents, or parents

165

and offspring, with identical running levels of 16,000 revolutions, which results in an

inability to calculate heritability for some lines in many later generations (Figure 3.6C).

As noted in the Methods, when means are calculated, missing values are excluded from

the calculation. In panel C, starting at around generation 16, having all parents being

identical or offspring identical to both parents results in missing values for multiple lines

and families within each line, leading to highly variable heritability estimates.

The standardized selection differentials (calculated within family and sex) for the

unconstrained model remained very consistently around 0 for the control lines (Figure

3.7A) and 1.24 for the HR lines (Figure 3.7A). Selection differentials for the constrained

simulations are similar to those for the unconstrained model until about generation 14,

then drop rapidly, bottoming out near 0.16 by generation 61 (Figure 3.7B) (and note that

the selection differential rarely becomes zero within a given HR line).

Under both models, Type I error rate for a = 0.05 when comparing allele

frequencies of HR with C lines was deflated at generation 0, regardless of the effect size

for the locus. Type I error ranged from 0.0390 to 0.0434 with no preference for any

effect size (Supplemental Table S3.1).

As expected, power to detect differentiation between the HR and C lines

increased across generations (more rapidly across generations 5 through 15 than later),

but never exceeded 0.059 for any generation or either model, and reached a maximum by

about generation 50 under both models. Comparing models at each generation indicates

that power is significantly higher under the unconstrained model by generation 20,

although the difference is trivial (0.0015 with P=0.0218) (Table 3.3). This differential in

power increased through generation 61, when it reached 0.0061 (P=7.21E-15). Although the information in Table 3.3 does not tell us about power to detect loci based on effect size, it does establish that we expect more total selection signatures at generation 61 than 22 (see Discussion).

The correlation between generation 22 and 61 p-values for the constrained model was 0.3948 versus 0.4006 for the unconstrained model (total N = 200, unpaired-T = -2.2808, P = 0.0236). In the unconstrained model, 34.5% of the loci significantly differentiated at generation 22 (α = 0.05) were still differentiated at generation 61, versus 32.2% under the constrained model (unpaired-T = -3.8872, P = 1.39E-04). This consistency is more than 3 times as much as the real data (9.12% for T-tests), which may simply reflect the drop in correlation between generations 22 and 61 p-values (Figure 3.1). Incorporation of sampling error into the constrained model lowered the correlation between p-values to 0.2801 and the proportion of loci significant at generation 22 still significant at generation 61 to 25.2%.

Comparisons of power to detect differentiation between the HR and C lines in relation to effect size of locus and generation under two simulation models (Supplemental Table S3.2) indicates:

1. power increased with effect size, as expected;

2. at generation 22, power was similar between the two models, with the exception of loci with effect size 102.4, where power was greater under the unconstrained model;

3. at generation 61, power was consistently greater under the unconstrained model, especially for loci with effect sizes in the range of 12.8 to 102.4;

4. under both models, power was greater at generation 61, except for the loci with the two largest effect sizes, where the power difference was reversed.

Comparison of Selection Signatures in Generations 22 and 61 for Pooled Data

When the average allele frequencies of SNPs within regions identified by strict culling at generation 22 (this study) are compared to the average AF of those loci at generation 61, an increase in among-line variance is apparent for generation 61, within both the HR and C linetypes (Figure 3.8). All else being equal, this increase in among-replicate variance should lower the statistical power to detect differentiation between the HR and C linetypes. In agreement with this expectation, most of these strict regions at generation 22 (Table 3.2) are no longer significantly differentiated at generation 61 (Table 3.2 and Figure 3.8). However, several of the 6 strictly culled regions at generation 61 also show some evidence of differentiation at generation 22 (Figure 3.9). Although strict culling methods exclude regions identified with generation 61 AF analyses, regions implicated in generation 22 with FDR = 0.01 culling alone do have considerable overlap with some of these 6 regions identified at generation 61. Generation 61 regions 3, 5, and 6 (Table 3.2) were significant at FDR = 0.01 for all three analyses at generation 22 (Supplemental Table S3.3).

For some of the regions identified as significant at generation 22, differentiation may have been lost by generation 61 as result of a single line diverging from the others (for example, line 3 in region 2 or line 7 in region 4 [Figure 3.8]). In general, mean

differences at generation 22 are much smaller than at 61, but also with much less among-line variance. A particular example of this includes region 16 (Figure 3.8), which is the only region identified at generation 22 (after strict culling) to continue to be detected as differentiated at generation 61 (see Discussion).

Possible Biological Function of Generation 22 "Strict" Differentiated Regions

A total of 355 genes (including predicted genes and miRNA) were identified in the 21 "strict" culling generation 22 regions (Supplemental Table S3.4). Of these, 144 genes had associated knockout phenotypes. Knockout phenotypes associated with these identified genes include changes to neurological, cardiac, skeletal, and lung development and more. However, after comparing the knockout phenotypes of the differentiated genes to those of the whole genome, we found that only the genes associated with lung to be overrepresented (chi-squared, $p = 0.0212$, df = 1) (see Supplemental Table S3.5). Therefore, we focused on the results provided by ToppGene.

Of the 355 differentiated genes, 224 were recognized by ToppGene and used for identifying potential biological function. Those not recognized were generally olfactory, vomeronasal, miRNA, or predicted genes. The full listing of significant ToppGene results can be found in Supplemental Table S3.6.

GO Molecular Functions implicate post-transcriptional and translation regulation largely due to the miRNA sequences included in the strict regions. Volume-sensitive anion channel activity is also included due to the *Lrrc8* components in region 5

(chr5:102,846,390-106,315,986). GO Biological Processes implicate energy homeostasis and response to phenylalanine (dopamine and epinephrine precursor), also citing predominantly miRNA sequences as matched genes for the process. GO Cellular Component implicates the RISC complex, again citing miRNA sequences for gene silencing. Implicated mouse phenotypes include abnormal periodontal ligament and cementum morphology both phenotypes citing a nearly identical list of genes found in region 5. Parkinson's disease was the one pathway implicated by the included genes due to the association of 7 miRNA sequences with the disease, some of which are also implicated for response to phenylalanine. The most significant coexpression term is "Human Brain Bray09 298genes."

ToppCell Atlas results implicate numerous systems: e.g., blood and bone marrow-primary blood derived cancer, frontal cortex-neuronal GABAnergic neuron, ventricular cardiomyocyte, frontal cortex-neuronal glutamatergic neuron (Supplemental Table S3.6).

Diseases associated with these genes include various cancers, particularly those associated with the thyroid and lung. Additionally, argentaffinoma (tumor which secretes large amounts of serotonin) was associated with these genes. Genes associated with cocaine abuse were also implicated.

DISCUSSION

Overview

Previously, whole-genome sequence data for individual mice at generation 61 of the High-Runner mouse selection experiment were used to identify 13 genomic regions differentiated between HR and control lines (termed "consistent" regions because they appeared with three different analytical methods) (Hillis *et al.* 2020). These regions contained genes associated with known phenotypic differences between the HR and control lines and intuitive associations with running ability and/or motivation. However, given that the HR lines had begun to reach selection limits around generation 22 (Careau *et al.* 2013), 39 additional generations, with ongoing random genetic drift, could have obscured many selection signatures. Therefore, in the present study, we analyzed allele frequencies for the lines sampled at generation 22, based on DNA pooled by line. These analyses of generation 22 identify hundreds of genomic regions differentiated between the HR and C lines (FDR = 0.01), despite using pooled sequence data rather than sequences for individual mice (Xu and Garland 2017). We then reanalyzed the data from generation 61 as allele frequencies by line, to mimic the data available for generation 22, and found that the regions identified as differentiated at generation 61 are, at best, weakly differentiated at generation 22. Nevertheless, both generations' differentiated regions contain genes that make biological sense for wheel-running behavior. Below, we discuss (1) implications of the differences in data type between generations 22 and 61, (2) possible statistical and biological explanations for the differences in identified regions,

and (3) genes and biological systems highlighted by the genomic regions identified by generation 22 analyses (after strict culling).

Differences in Selection Signatures at Generations 22 and 61

We expected estimates of selection signatures to be similar at generations 22 and 61, based on the fact that the HR lines had mostly reached selection limits by generation 22 (Careau *et al.* 2013), such that the most biologically important loci would have gone to fixation or at least reached equilibria across most or all of the HR lines. In agreement with this expectation, of the 13 "consistent" regions identified by Hillis et al. (2020) for generation 61 (using individual mouse data), 8 were still identified by at least one of the tests (FDR = 0.01) using the generation 61 genotypes pooled into allele frequencies per line. Generation 22 analyses of pooled sequence data identified 11 of the 13 consistent regions with at least one test (although several of these regions were only detected by a few SNPs: Supplemental Table S3.3). Specifically, T-tests identified 10, regularized T-tests identified 8, and windowed regularized T-tests identified 9 of the 13 consistent regions listed by Hillis et al. (2020). Interestingly, the consistent region on chromosome 14 was more strongly detected with all generation 22 analyses than with the generation 61 pooled sequence analyses (Table 3.2).

On the other hand, the strongest selection signatures observed at generation 61 with the data treated as pooled sequences are not among the strongest ones observed at generation 22 (based on number of SNPs detected and their p-values), despite continued selection on the HR lines. For example, region 1 of the generation 61 "strict" culling

pooled analyses (chr3:51,190,735-51,623,627) included 948 SNPs (RegT test, FDR = 0.01) (Table 3.2), and although the generation 22 analyses did identify this region, it was only by the T-tests and only with a single SNP at FDR = 0.01. As another example, region 3 of the generation 61 pooled analyses (chr9:41,321,625-42,478,817) included 1,277 SNPs (WRT test, FDR = 0.01), but the same region was only identified by 3 SNPs by T-tests and RegT tests and 2 SNPs by WRT test in the generation 22 analyses (Supplemental Table S3.3). When directly comparing SNPs differentiated at FDR = 0.01 for the different tests within a generation, we see considerable overlap among the results of each statistical method for significant SNPs (Figures 3.10A and B). However, despite some overlap in regions (as noted above), we see no overlap in SNPs between generations 22 and 61 for any test, with exception of a single SNP for the WRT test (Figures 3.10C-E).

In addition to the differences in individual SNP results, a 20-fold greater number of regions was identified by generation 22 analyses than generation 61 pooled analyses at FDR = 0.01 (Table 3.1). This ratio applies to all statistical tests and the complete SNP analyses for each generation, as well as the analyses of SNPs shared by the two generations. Moreover, the SNPs identified at generation 61 were clustered into far fewer regions (Table 3.1). Broadly, these difference in numbers of selection signatures have at least two possible explanations, which are not mutually exclusive: (1) differences in data type, quality, and quantity; (2) biological differences between generations 22 and 61.

Differences in Data Type, Quality, Quantity, and Sampling Error

Our power to detect differentiation in allele frequencies should have been lower for generation 22 than for generation 61 (Figure 3.4C and D). As also noted in the Methods, the estimates for SNP allele frequencies per line at generation 61 were based on ~10 mice/line sampled and an 12X average read depth per mouse, yielding a total of 5,932,148 variable SNP loci (Hillis *et al.* 2020). For generation 22, pooled sequencing was done with ~20 mice/line and an average read depth of 24X, yielding 4,446,523 variable SNPs (Table 3.1). Generally, with an average read depth of 12X per mouse, both alleles will be represented for each mouse (i.e., 20 alleles per line) for generation 61 allele frequencies. However, with 24X average read depth for generation 22, simulations involving sampling alleles with replacement show that generation 22 is prone to vary more from the actual population allele frequency (Figure 3.4A and B). Thus, the much greater number of differentiated SNPs and chromosomal regions detected at generation 22 would not appear to be simply a function of greater statistical power versus generation 61. Thus, we now consider possible biological explanations.

Biological Differences

One way to highlight the differences in selection signatures detected at generations 22 and 61 is to note that of the differentiated regions detected for generation 61, two of them contain hundreds of statistically significant SNPs (FDR = 0.01) shared between the generation 22 and 61 data sets. Despite this, those two regions are not among the more differentiated regions in the Manhattan plots (Figure 3.2).

What biological explanations might account for such discrepancies? One possibility is a physiological constraint that eliminates the need for all loci favorable to wheel running to be maintained at high frequencies once a selection limit is reached (see verbal model in Hillis and Garland Jr 2022). We consider this from the perspective that many complex traits are influenced by hundreds or thousands of loci (Wood *et al.* 2014; Long *et al.* 2015). Voluntary exercise behaviors would likely be among them, given that they incorporate numerous physiological and morphological traits related to ability (e.g., cardiac muscle, skeletal muscle, bone, metabolism, water and temperature homeostasis) as well as aspects of motivation and reward (e.g., dopamine signaling, chemosensory systems) (Lightfoot *et al.* 2018; Wang *et al.* 2022).

Although biological constraints can be defined in various ways (Garland *et al.* 2022), in the present context, a constraint would be anything that limits the maximum revolutions that an individual mouse can run during the testing period. Previously, we discussed how different unique responses to identical selection criteria (i.e., "multiple solutions") could occur and referenced constraints as a potential explanation (Hillis and Garland Jr 2022). To utilize and expand on their example, 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" or single limiting factor in the biological systems required for high 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 five such alleles are needed to achieve the maximum amount of wheel running

175

permitted by joint pain. Under this model, if selection acts on a population to increase

running, then (1) fixation of the favorable allele at any five of the 10 loci will coincide

with a selection limit determined by pain tolerance, (2) none of the alleles at any of the 10

loci must be fixed to reach the pain-determined limit, (3) more than 5 favorable alleles

could be maintained at intermediate allele frequencies, and (4) as long as enough

favorable alleles are maintained for the selection limit, some favorable alleles can be lost

without detriment to wheel running. These factors allow for substantial variation among

the replicate lines and considerable flexibility for change within a given line, even for

favorable wheel-running alleles at the selection limit. This possibility of "genetic churn"

beyond a selection limit that is caused by a physiological constraint also implies that

genotype-to-phenotype maps (Travisano and Shaw 2013; Zamer and Scheiner 2014;

Porto *et al.* 2016; Zinski *et al.* 2021) may be moving targets and hence difficult to

identify. Therefore, we used simulations to compare power to detect and consistency in

detected selection signatures, both with and without a physiological constraint.

Simulations Comparing Power under Constrained versus Unconstrained Models

Similarities Between the Constrained Model and Real Data

The constrained model appears to better reflect what we observe in the actual response to

selection, in two main ways. First, we do not see a clear selection limit achieved under

the unconstrained model (Figure 3.5). Although the response to selection diminishes

over time (likely due to the reduction in heritability: Figure 3.6B), a clear plateau is not

apparent. In contrast -- as must be the case -- a clear plateau occurs under the constrained

model, at or near the value of the constraint. Second, the standardized selection

differential for the unconstrained model remains around 1.24 for all generations (Figure

3.7B), unlike the selection differential for the constrained model, which decreases (Figure

3.7C), as was observed in the actual HR lines (Careau *et al.* 2013).

Correlation Between Generations 22 and 61 P-values

For the tests comparing allele frequencies at each of 2,096 loci between the HR

and C lines, the correlation of p-values between generations 22 and 61 was statistically

lower in the constrained model (r=0.3948) as compared to the unconstrained model

(r=0.4006), though still 4x higher than for the real genomic data (r=0.0909) (Figure 3.1).

Even with the inclusion of sampling error, the correlation (r=0.2801) is more than 3x

greater than for the real genomic data, which indicates that other factors (e.g., gene

interactions) must be contributing to the differences between the generations.

In spite of the similarity in the correlation of p-values between generations, the

between-generation matching of detected selection signatures was significantly greater

under the unconstrained (34.5%) than under the constrained model (32.2%). Presumably,

this (small) difference occurs because, as the constraint is approached, the selection

intensity on wheel running (Figure 3.7) and at the level of individual loci that remain

polymorphic, is reduced. This reduction leads to less fixation of favored alleles.

However, the relatively small difference between models in matching of selection

signatures is not enough to explain the large differences in the real data between

generations 22 and 61 (Table 3.2). The inclusion of sampling error into the estimates decreased the 32.2% matching between generations 22 and 61 differentiated loci to 25.2%. This level of matching with simulated data remains more than 2.5-fold higher than for the real data (9.12%), thus implicating the presence of additional factors that reduce matching in the real data (e.g., epistatic effects).

Overall, our simulations fail to demonstrate why we observe a 20X drop in significant regions from generation 22 to 61 (Table 3.1), implying instead that we should detect more at generation 61 than at 22 (Table 3.3).

Effect Sizes of Loci

Under both models, more loci were detected at later generations (Table 3.3). However, the power to detect loci with the largest effect size was much higher at earlier than later generations (Supplemental Table S3.2 and Figure S3.1). This pattern makes sense in consideration of the factors that affect the average difference in allele frequency between the HR and control lines and the variance among replicate lines within linetypes. Drift will generally increase the variance among lines with each generation. The allele frequencies in the simulated control lines will be affected only by this drift. Allele frequencies in the HR lines will be affected both by drift and selection, where selection will have stronger effects at loci with larger effect sizes. This results in something of a race between selection increasing the difference in allele frequencies between HR and control lines, while drift increases among-line variance for both HR and control lines. For loci with small effect sizes, drift will have a relatively greater influence over allele

frequencies than selection at any generation, and thus detection rates never vary far from the Type I error rate, i.e., power is virtually zero (Supplemental Table S3.2). Loci with large effect sizes, however, are able to differentiate rapidly, often leading to fixation of the favored allele in our simulations. Even after fixation in the HR lines, drift is still able to increase allele-frequency variance among the control lines (potentially to the point of fixing loci for opposite alleles), thus further reducing the power to detect any differentiation. Thus, the power to detect signatures of selection should increase the most rapidly across generations for loci with the largest effect sizes, but power is also expected to decline after fixation of the favored alleles in the HR lines and with continuing increase in variance among the control lines (Supplemental Figure S1).

That the power to detect a locus as differentiated is correlated with its effect size is unsurprising. For example, under both models the power to detect selection signatures for loci with 0.4 effect size is about 20.6-fold less than the power to detect loci with 204.8 effect size (generation 22). This gap diminishes to about 7.9-fold difference by generation 61 (Supplemental Table S3.2), presumably due to the reasons described in the previous paragraph. However, the 0.4 effect size loci are far more numerous than the 204.8 effect size loci (N = 720 and 8, respectively). Consequently, the number of 0.4 effect size loci detected as significant is more than 4-fold greater than the number of 204.8 effect size loci detected.

For identifying possible biological functions, we would ideally focus on loci with relatively large effect size, as these will have the most direct influence on the phenotype and may serve as potential targets for future functional studies. We have no information

on effect sizes of SNPs or regions detected as differentiated for our real data. The relative proportions of low- and high-effect size loci among the detected selection signatures in the real data will likely vary from our simulations, depending on the actual distribution of those effect sizes and other factors. However, the simulations do implicate that we may have numerous small-effect size loci among our detected selection signatures. The inclusion of the "strict" culling method meant to prioritize regions which would have a large effect size. Having more linked loci differentiated loci would be expected from those regions under strong selection because recombination would have fewer generations to break up linked base pairs before the region becoming fixed in the HR lines. Since the simulations have so many more loci with small effect sizes, at generation 0, when we compare the lowest p-value produced for each simulation for the 0.4 effect size we tend to see lower p-values than loci with 204.8 effect size simply because of more opportunities to produce low p-value. However, by generation 22, the lowest p-values produced by loci with effect size 0.4 (mean p = 0.00274) are about twice that of 204.8 (mean p = 0.00153), for the constrained model. This advantage for loci with higher effect size may vary with the effect size distribution in the real data.

Possible Biological Functions of Generation 22 "Strict" Differentiated Regions

A total of 355 genes (including predicted genes and miRNA) were identified in the 21 "strict" culling generation 22 regions (Supplemental Table S3.4). These include genes associated with a wide variety of knockout phenotypes (including those related to the

development/maintenance of cardiac, neurological, bone, etc.) but only those related to lung were overrepresented.

ToppGene

Of the 355 genes mentioned above, 224 were recognized by ToppGene. 75 of the recognized genes are miRNAs, and many of the top suggested biological functions appears to cite these miRNAs rather than protein coding genes.

Significant molecular functions associated with the 21 strict culling regions are predominantly related to post-transcriptional and translational regulatory activity. This could mean that most of the phenotypic changes we see in the HR lines are a result of changes in protein expression. This is likely correlated with the identification of the RISC (RNA-induced silencing complex) as the most implicated cellular component among the genes, given that it is involved in regulating transcriptional and translational pathways (Pratt and MacRae 2009).

Additionally, from the ToppGene analyses, we identified various genes associated with phenylalanine (a precursor for dopamine synthesis), cocaine abuse, or GABAergic neurons (e.g., *Adamts19*, *Dnah11*, various MiRNAs), which may be tied to the increased motivation to run in the HR mice. Particularly, the HR mice have been to respond more intensely to dopamine drugs (particularly those that block dopamine transporter proteins) that the control mice (Rhodes *et al.* 2005). However, the strongest implicated process is energy homeostasis. This association with energy homeostasis appears to result from a region we identified at generation 22 with the traditional T-test, specifically

(Chr12:109,367,487-110,687,541). Labialle et al. (2014) performed a knockout of the miR-379/miR-410 miRNA cluster found here (which is the largest known cluster of miRNA specific to placental mammals) and found that KO neonates had dysregulated glucose and glycogen metabolism, as well as reduced ketone levels in severely hypoglycemic mice, and generally altered gene expression in the liver. Differentiation of these same miR genes in the HR lines may then implicate some substantial link to neonatal metabolism and/or running behavior, which can be energetically costly (Rezende *et al.* 2009; Garland, Jr. *et al.* 2011a; Copes *et al.* 2015).

Other Genes Not Included in the ToppGene Data Base

A number of genes of note were not included in the ToppGene analyses due to lack of data available in the database. Among these genes are olfactory, vomeronasal, miRNA, or predicted genes.

Importantly, among the genes identified are 12 vomeronasal sensor proteins and 8 (non-vomeronasal) olfactory proteins identified across two separate regions (Table 3.2, G22 regions 6 and 16), along with one additional gene (*Nmnat2*) associated with olfactory bulb morphology (Gilley *et al.* 2013). Generation 61 analyses also identified genes related to olfaction, and particularly vomeronasal systems (Hillis *et al.* 2020; Nguyen *et al.* 2020). Additionally, analyses dropping individual lines from generation 61 differentiation tests demonstrated that the replicate HR lines have responded to selection via different non-vomeronasal olfactory genes (Hillis and Garland Jr 2022). Therefore, differentiation in these systems at generation 22 does establish some consistency with the

generation 61 analyses. For logistical reasons, roughly 2/3 of the mice are tested on wheels that had a different mouse running on it the previous week without the wheel being washed in the interim. Analyses of 11,420 control and 26,575 HR mice across 78 generations have shown wheel running is affected by the presence of a previous mouse and that the effect is greater in HR lines (Dewan *et al.* 2019).

The region that was identified only by the regularized T-tests (chr1:77,650,941-78,065,681) does not contain any annotated genes of interest (therefore, did not contribute to ToppGene analyses); however, it is about 136kb upstream of *Epha4*. *Epha4* knockouts have fewer abducens motor neurons (Nugent *et al.* 2017), along with altered gait (Dottori *et al.* 1998; Borgius *et al.* 2014) and limb and motor coordination (Dottori *et al.* 1998; Mohd-Zin *et al.* 2016), as well as heart hypertrophy when knocked out in rats (Li *et al.* 2021). Outside of KO studies, *Epha4* has also been implicated in pain control (Kim *et al.* 2020; Wang *et al.* 2021), and osteoclast activation (Lau and Sheng 2018).

Surprisingly, two adjacent SNPs (Chr11:44,593,289-44,593,290, Supplemental Table S3.3) were found to be fixed for opposite alleles in the HR versus C lines at generation 22 (quality measures excluded these loci at generation 61). These loci appear to be about 24.8kb upstream of *Ebf1*, a gene associated with olfactory receptor gene expression (Wang and Reed 1993; Wang *et al.* 2004) and whose knockouts are associated with heart and bone development (Hesslein *et al.* 2009; The International Mouse Phenotyping Consortium *et al.* 2016). As noted in the Introduction, the linetypes differ in both cardiac and skeletal phenotypes.

Comparisons with Other Rodent Selection Experiments

A few other selection experiments with traits related to exercise behavior or physiology in rodents provide relevant genetic data. In wild-derived voles selectively bred for high aerobic capacity during a swimming test, Konczal et al. (2015) used RNA-seq to compare gene expression profiles in liver and heart. More than 300 genes were differentially expressed in the four selected lines as compared with four non-selected control lines at generation 13. However, inferred SNP frequencies did not distinguish the selected and control lines. They identified a differentially expressed gene with associated with glycogen metabolism (*PYGL*), along with some related to lipid metabolism. Glycogen metabolism is relevant for aerobic exercise and may show some congruence with our metabolism findings at generation 22 (see above). Another comparison between control lines and vole lines bred for increased predatory aggression with crickets identified genes associated with GABAergic activity (Konczal *et al.* 2016). Beyond the GABAergic genes identified in the present study, the HR mice have also been shown to have higher predatory aggression as compared to control lines (Gammie *et al.* 2003).

Gene ontology of differentially expressed genes comparing rats bred for high vs low treadmill endurance found regulation of transcription, cellular response, and cellular metabolic processes to be among the most frequently differentiated terms (Ren *et al.* 2016). However, comparisons of high- with low-endurance selected rat lines (no control line was maintained) will confound genes that improve performance with those that reduce performance in ways that may reflect various disease states (e.g., see Wisløff *et al.* 2005; Thyfault *et al.* 2009; Palpant *et al.* 2009).

A selection experiment involving high and low voluntary wheel running lines of rats was developed to study the dopamine reward pathway (Roberts *et al.* 2012). After 8 generations of selection, gene expression in the nucleus accumbens, a brain region associated with reward pathways and voluntary running, indicated that "nervous system development and function" was among the differentiated terms between the high and low running rats when housed without wheel access (Roberts *et al.* 2013). Further expression analyses of nucleus accumbens at generation 8 identified several differentially express dopamine-related genes when comparing the high-running and non-selected control line, both with and without wheel access (Roberts *et al.* 2014). These findings are also congruent with the GABAergic related genes identified in the present study and altered response to dopamine related drugs (Rhodes *et al.* 2005).

LIMITATIONS AND CONCLUSIONS

Some of the limitations of the present study include trying to compare results of pooled genome sequencing (generation 22) to individual mouse sequencing (generation 61: Hillis *et al.* 2020). Though the alleles of the individual mice can be combined to imitate pooled genome sequences, the differences in number of mice sampled and sampling error make comparisons difficult (see Methods). This is illustrated by the 3x decrease in p-value correlations (between generations 22 and 61) as compared to the both the unconstrained and constrained simulations. Nevertheless, as argued above, neither the increase in number of regions detected as differentiated at generation 22 nor the lack of

correspondence between detected regions at generations 22 and 61 can be explained solely by methodological differences.

The constraint simulations have their own limitations in that they do not account for male vs female running differences (females run more than males) or the observed strong seasonal variation (Careau *et al.* 2013). In addition, dominance, epistasis, and gene-environment interactions were left out of the models. Exclusion of these features may be why achieving realistic response to selection required a heritability estimate that pushes the upper limit of calculated heritability for the selection experiment. This model also does not include linkage disequilibrium or realistic rates of recombination. Lastly, we did not explore the potential effects of relaxing selection for four generations, as when the mice were moved from Wisconsin to California (see Introduction). A cluster of generations of no selection in the HR lines could allow for some drift of the favored alleles.

Attempts to determine biological function of all 436 regions identified at generation 22 by the different statistical tests at FDR = 0.01 did not produce realistic ontologies (e.g., "fin development" was identified). This may be attributable to several regions functioning as trans- regulators for distal genes, large abundance of hitchhiking genes irrelevant to running, and/or incomplete understanding of the biological function of some of the genes identified. Thus, the biological functions discussed above are based on only a subset of the likely response to selection.

Although, we are unsure as to why we see so many regions at FDR = 0.01 that do not correspond to the generation 61 findings by Hillis et al. (2020), our simulations

suggest that regions with the strongest effect sizes on wheel running are likely to be among the generation 22 regions. Given the statistical significance and number of SNPs identified in our "strictly" culled differentiated regions, these regions are most likely to have had the greatest impact on wheel running at the start of the selection experiment. Among these regions are genes related to olfactory/vomeronasal function, reward pathways, and a miRNA cluster that has been associated with energy homeostasis in neonatal development. All of these associations make sense based on known phenotypic differences between the HR and control lines (see Introduction).

Future directions might include more complex simulations (e.g., see Baldwin-Brown *et al.* 2014; Stephan 2016; Castro *et al.* 2019), which may better help to explain the 20X increase in regions detected at generation 22. Including genomic data from more generations (especially from the base population, generations near to but before the seleciton limit, and current generations [i.e., around 100]) may provide more clarity regarding how the response to selection changes across phases of the selection response (cf. Rose *et al.* 2005; Castro *et al.* 2021). Analyses using all loci and a kinship matrix would enable determination of some interactions between genes. Functional analyses, such as knockouts of some of the genes whose alleles are favored by selection, may provide direct evidence of influence on wheel-running behavior (e.g., Schmidt *et al.* 2008; Chaouloff *et al.* 2011; MacKay *et al.* 2019). Furthermore, analyses of other physiological aspects of these KO mice may help to better understand the mechanisms by which the gene influences wheel running.

ACKNOWLEDGEMENTS

FUNDING

CONFLICT OF INTEREST

The authors have no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualization, D.A.H., Lir.Y., F.P.M.dEV., D.P., T.G.; investigation, D.A.H., Lir.Y., G.M.W., F.P.M.dEV., D.P., T.G.; software, D.A.H., Lei.Y.; formal analysis, D.A.H., Lir.Y., T.G.; writing – original draft, D.A.H., T.G.; writing review and editing, D.A.H., F.C, T.G.

TABLES

Table 3.1

Data	Test	Total Loci	FDR 0.01 - logP	Significant SNPs	All Regions	Regions after strict culling
Gen22AF	T-test (arcsine)	2,045,546	3.20	746	304[a]	4
Gen22AF	Reg T-test (arcsine)	2,045,546	2.67	473	166[a]	3
Gen22AF	WRT Test (arcsine)	2,045,546	2.66	630	187[a]	6
Gen61AF	T-test (arcsine)	2,045,546	3.58	619	12[b]	3
Gen61AF	Reg T-test (arcsine)	2,045,546	2.90	1,140	7[b]	2
Gen61AF	WRT Test (arcsine)	2,045,546	3.06	1,285	11[b]	4
Gen22AF	T-test (arcsine)	4,446,523	3.28	1,367	403[c]	14
Gen22AF	Reg T-test (arcsine)	4,446,523	2.64	917	239[c]	11
Gen22AF	WRT Test (arcsine)	4,446,523	2.62	1,184	258[c]	13
Gen61AF	T-test (arcsine)	5,932,148	3.64	634	19[d]	4
Gen61AF	Reg T-test (arcsine)	5,932,148	2.80	2,251	11[d]	5
Gen61AF	WRT Test (arcsine)	5,932,148	3.23	1,449	11[d]	5

Comparison of SNPs with FDR = 0.01 between HR and C lines of mice using different statistical tests at generations 22 and 61.

Number of SNPs listed represents those that are statistically significant based on a False Discovery Rate of 1% using permutations. Analyses with 2,045,546 loci incorporate only loci which are shared between generations 22 and 61 (analyses were repeated for RegT and WRT tests for accurate variance structure with less loci). Regions distinguished by being separated from the next closest significant locus by more than 1 million bp. Additional regions remaining after additional culling methods have either 20 significant loci or at least 2 significant loci with one having a p-value <1.00E-06 (T-test) or <1.00E-04 (Regularized T-test and Windowed Reg. T-test).

[a] Number of unique regions = 322.
[b] Number of unique regions = 14.
[c] Number of unique regions = 436.
[d] Number of unique regions = 21.

Table 3.2

G22 Region	G61 Region	G22 Ttest	G22 RegT	G22 WRT	G61 Ttest	G61 RegT	G61 WRT	Chr	minPOS	maxPOS	Size	Loci	Shared Loci
1				x				1	77,650,941	78,065,681	414,741	5	0
2		x	x	x				1	152,318,219	153,239,876	921,658	40	25
3			x	x				2	78,021,909	78,974,325	952,417	3	0
	1				x	x	x	3	51,190,735	51,623,627	432,893	948	594
4				x				5	32,384,612	32,975,871	591,260	32	4
5		x	x	x				5	102,846,390	106,315,986	3,469,597	63	37
	2					x	x	6	40,933,658	41,751,702	818,045	27	3
6		x	x	x				6	122,815,876	124,446,843	1,630,968	43	20
7		x						7	51,146,343	53,400,869	2,254,527	6	5
8		x						9	15,895,185	15,895,206	22	22	1
	3				x	x	x	9	41,321,625	42,478,817	1,157,193	1,277	647
9		x						9	52,502,359	53,282,125	779,767	3	2
10				x				9	80,349,989	82,894,555	2,544,567	25	18
11				x				10	14,067,617	18,376,599	4,308,983	25	7
12		x	x	x				10	20,890,526	22,373,922	1,483,397	34	3
	4						x	10	104,966,751	105,529,701	562,951	2	0
13		x	x	x				12	109,367,487	110,687,541	1,320,055	21	5
14		x	x	x				12	111,821,592	118,456,814	6,635,223	39	19
15		x	x	x				13	46,088,694	47,278,395	1,189,702	33	13
16		x	x	x				14	51,198,229	53,776,455	2,578,227	42	7
17			x	x				14	76,834,210	78,080,942	1,246,733	5	3
	5				x	x	x	15	19,009,437	20,632,943	1,623,507	71	39
18		x						15	52,960,842	52,960,853	12	8	0
	6				x	x		16	31,604,977	32,394,022	789,046	28	11
19		x						17	53,762,011	54,723,155	961,145	22	0
20		x	x	x				18	56,480,454	60,343,089	3,862,636	128	81
21		x	x	x				18	78,018,740	78,504,680	485,941	4	4

Genomic regions identified as differentiated between 4 selectively bred HR lines and 4 non-selected control lines under "strict culling". A test is deemed to have produced a differentiated region if the region contains at least 20 SNPs significant at FDR = 0.01 or at least 2 SNPs significant at FDR = 0.01 and at least one SNP with p-value < 1.0E-04 (for RegT and WRT tests) or p-value < 1.0E-06 (for T-test) (See Methods: Determination of Selection Signatures). "Loci" listed are those significant at FDR = 0.01 (Table 3.1), the counts themselves match the number of differentiated loci identified by the statistical test with the most such loci. "Shared Loci" are the number of loci listed in the "Loci" column that are also shared between both generations.

Loci highlighted in red match "consistent" regions identified by Hillis et al. (2020)

Table 3.3

Generation	Unconstrained power	P-value for Unc. Comparing present generation to previous generation[1]	Constrained power	P-value for C. Comparing present generation to previous generation[1]	P-value for Constrained vs Unconstrained[2]
0	0.0418[3]	NA	0.0404[3]	NA	0.0273
5	0.0427	0.1317	0.0419	0.0147	0.1668
10	0.0444	0.0037	0.0443	0.0001	0.9017
15	0.0471	2.00E-05	0.0463	0.0019	0.2055
20	0.0485	0.0267	0.0470	0.3057	0.0218
22	0.0491	0.3376	0.0474	0.4325	0.0083
25	0.0501	0.1347	0.0487	0.0370	0.0409
30	0.0523	0.0014	0.0493	0.4272	1.09E-05
35	0.0541	0.0056	0.0502	0.1695	7.85E-08
40	0.0565	0.0009	0.0513	0.1464	4.23E-11
45	0.0580	0.0403	0.0521	0.2896	5.70E-14
50	0.0592	0.0949	0.0526	0.4489	3.53E-17
55	0.0591	0.9633	0.0533	0.3857	5.83E-13
60	0.0593	0.8042	0.0533	0.9948	6.71E-14
61	0.0590	0.6987	0.0529	0.6052	7.21E-15

Observed statistical power for comparing allele frequencies in simulations with constrained vs unconstrained levels of wheel-running behavior (averages for 100 simulations and ignoring variation in effect size across loci),

[1]These p-values are calculated using a T-test assuming unequal variance comparing the power of the generation for that line to the previously listed generation (e.g., unconstrained power at generation 5 compared to generation 0 has a p-value of 0.1317).

[2]P-value from a T-test (unequal variance) comparing the power of the 100 unconstrained simulations to the 100 constrained simulations

[3]Type I error rate.

192

FIGURES

Figure 3.1

Scatterplot comparisons of the generations 22 and 61 p-values with Pearson's correlation (with arcsine-square transformation for A-F and -$\log_{10}$ transform for G-L):

(A) generation 22 regularized T-test vs generation 22 standard T-test (cor = 0.9959),

(B) generation 22 regularized T-test vs generation 22 sliding window regularized T-test (cor = 0.9997),

(C) generation 22 standard T-test vs generation 22 sliding window regularized T-test (cor = 0.9964),

(D) generation 22 standard T-test vs generation 61 standard T-test (cor = 0.0898),

(E) generation 22 regularized T-test vs generation 61 regularized T-test (cor = 0.0900),

(F) generation 22 windowed regularized T-test vs generation 61 windowed regularized T-test (cor = 0.0909)

(G) generation 22 regularized T-test vs generation 22 standard T-test (cor = 0.9853),

(H) generation 22 regularized T-test vs generation 22 sliding window regularized T-test (cor = 0.9992),

(I) generation 22 standard T-test vs generation 22 sliding window regularized T-test (cor = 0.9869),

(J) generation 22 standard T-test vs generation 61 standard T-test (cor = 0.1103),

(K) generation 22 regularized T-test vs generation 61 regularized T-test (cor = 0.1151),

(L) generation 22 windowed regularized T-test vs generation 61 windowed regularized T-test (cor = 0.1156),

(M) ranked p-values, generation 22 windowed regularized T-test vs generation 61 windowed regularized T-test (cor = 0.0928),

(N) distribution of raw p-values (generation 22, WRT test),

(O) distribution of arcsine-square transformed p-values (generation 22, WRT test), and

(P) distribution of -$\log_{10}$ transformed p-values (generation 22, WRT test)

M

N

O

P

195
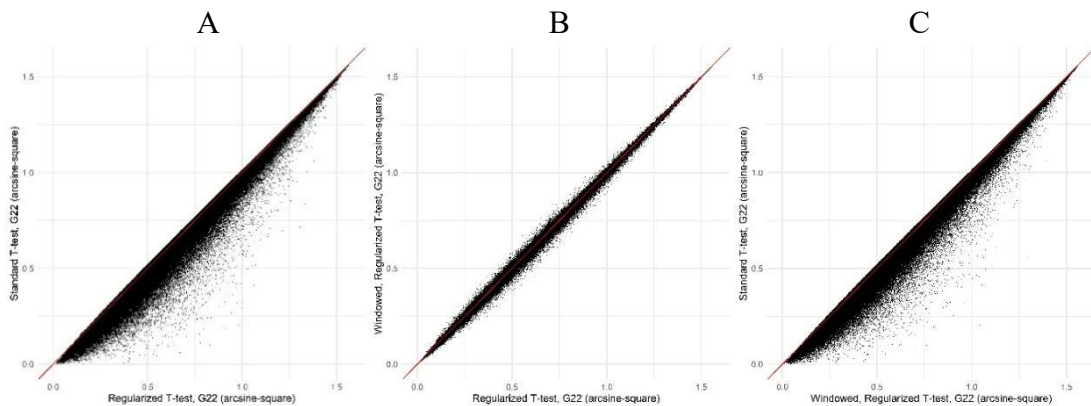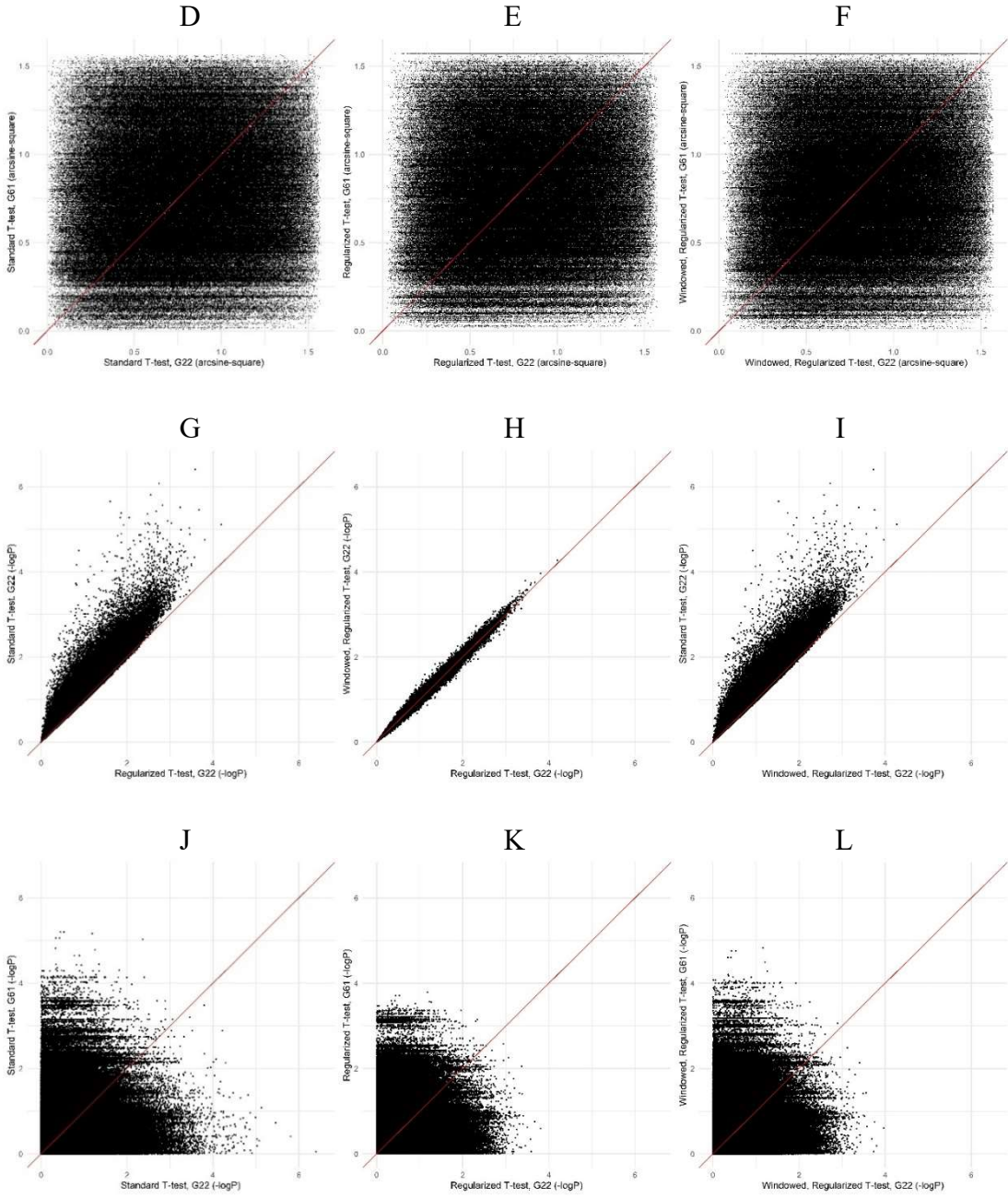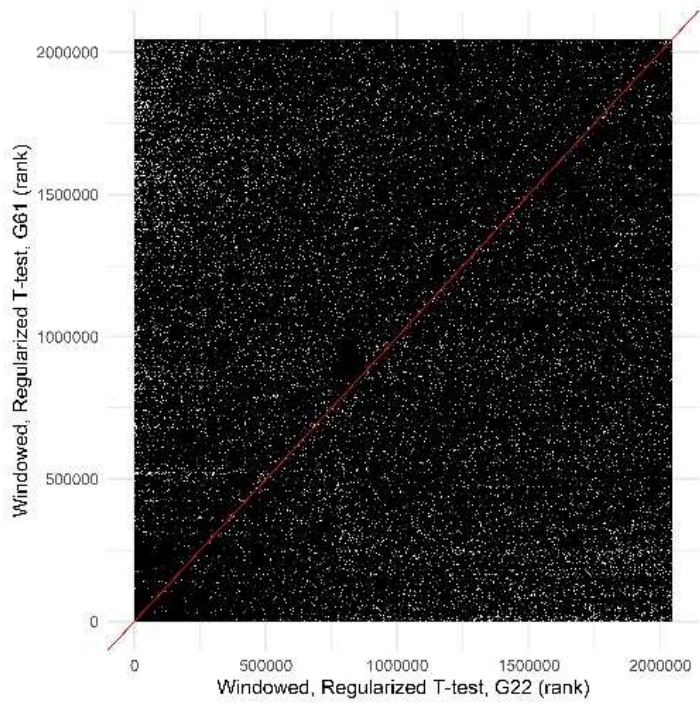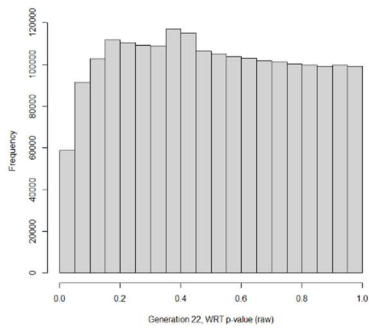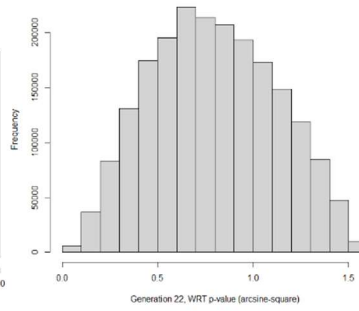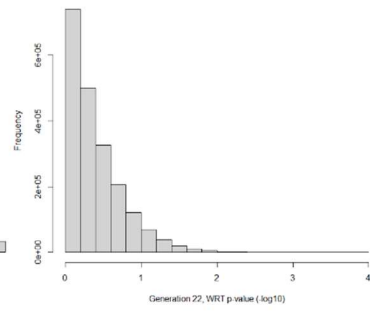
Figure 3.2
Manhattan plots for results from (A) generation 22 T-tests (shared loci), (B) generation
22 regularized T-test (shared loci), (C) the generation 22 sliding window regularized T-
test (shared loci), (D) the generation 61 (pooled) T-test (shared loci), (E) the generation
61 (pooled) sliding window regularized T-test (shared loci), (F) the generation 61
(pooled) regularized T-test (shared loci), (G) generation 22 T-tests (all loci), (H)
generation 22 regularized T-test (all loci), (I) the generation 22 sliding window
regularized T-test (all loci), (J) the generation 61 (pooled) T-test (all loci), and (K) the
generation 61 (pooled) sliding window regularized T-test (all loci), and (L) the generation
61 (pooled) regularized T-test (all loci). W=0.1 was used so that these results would be
comparable to analyses performed by Xu and Garland (2017). The red line indicates the
critical threshold (FDR = 0.01) for that individual test.

A

B

C

D

E

F

G

H

I

J

198
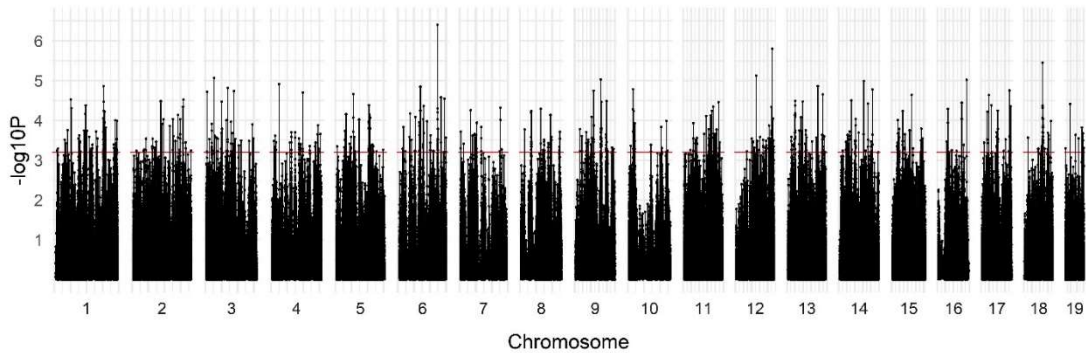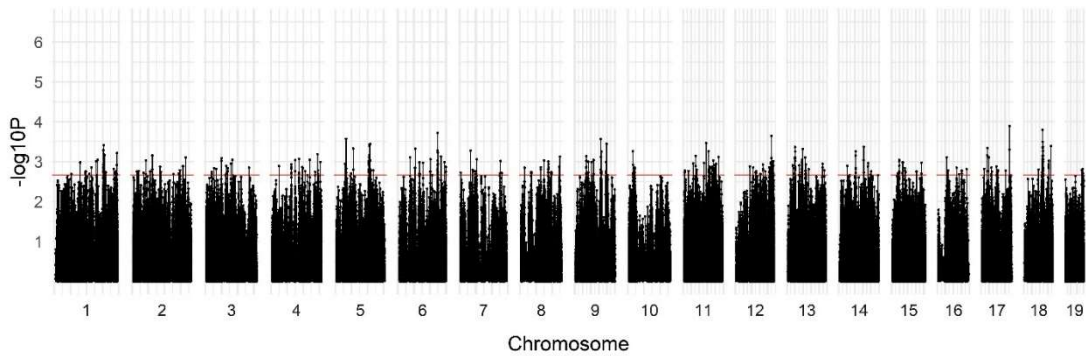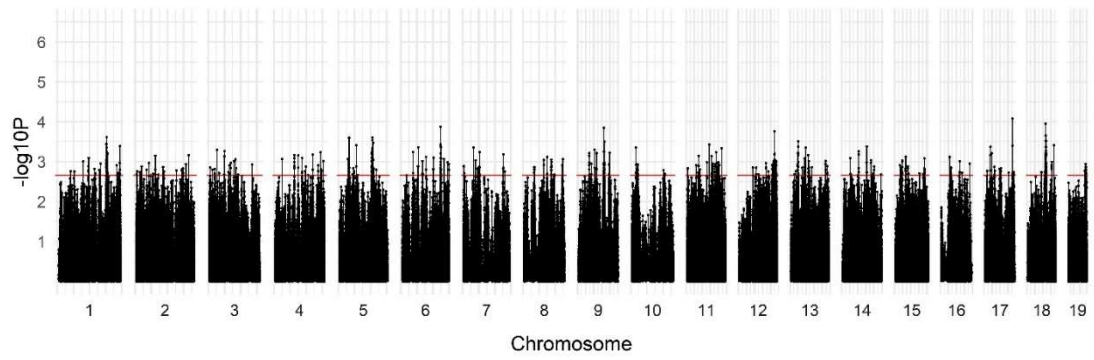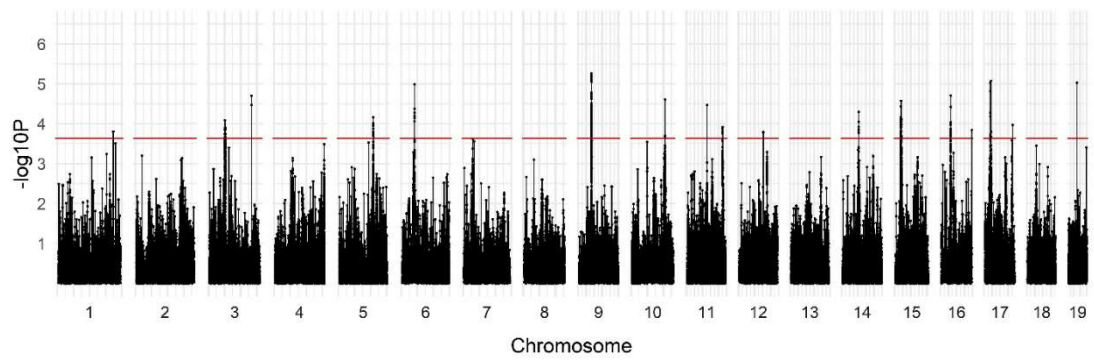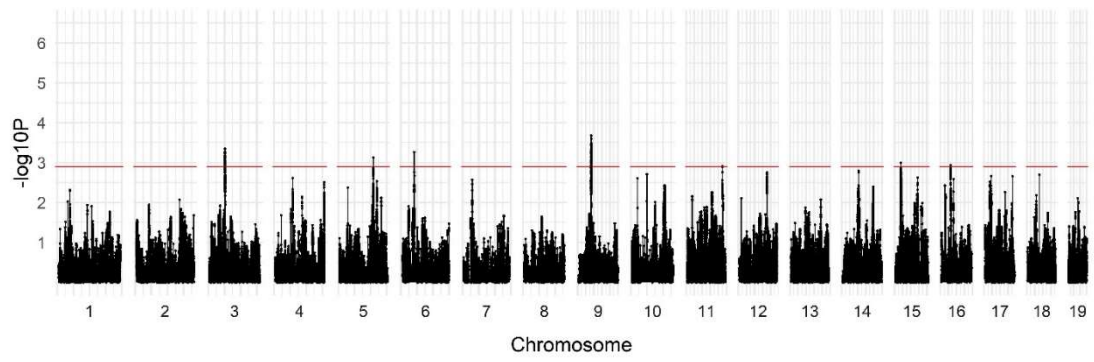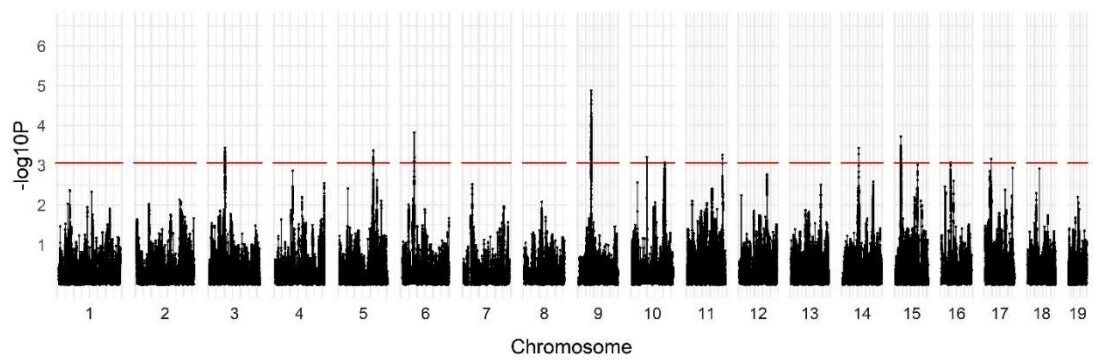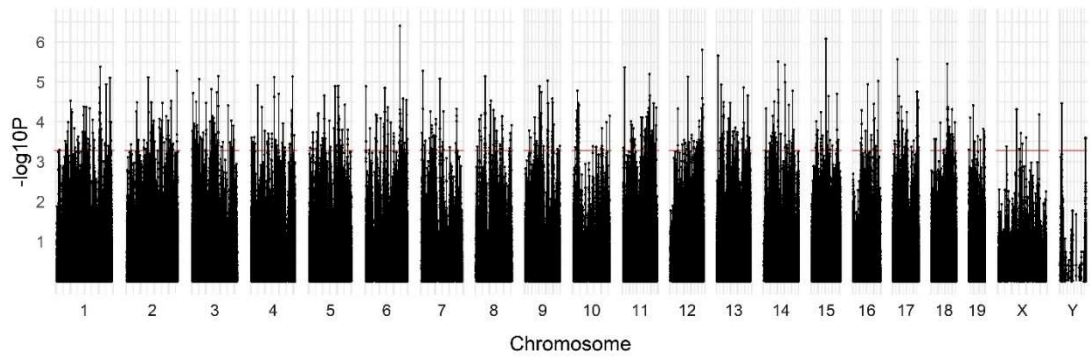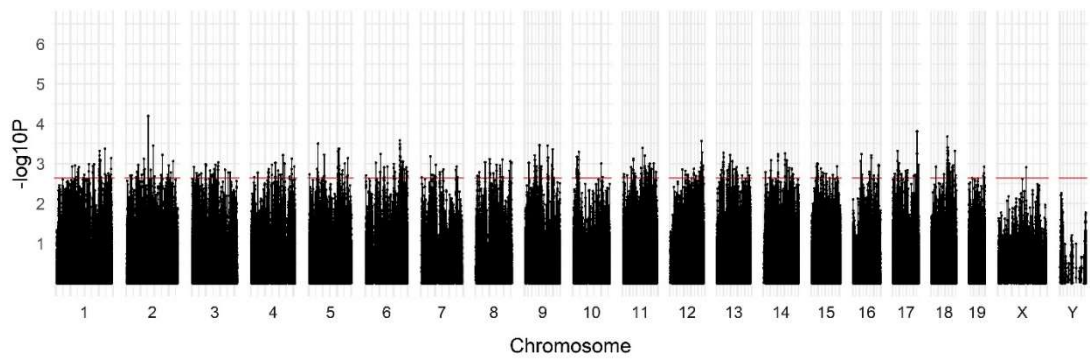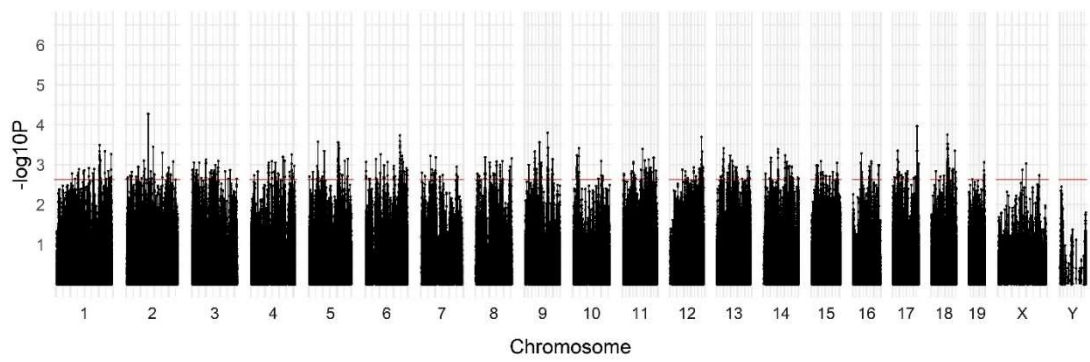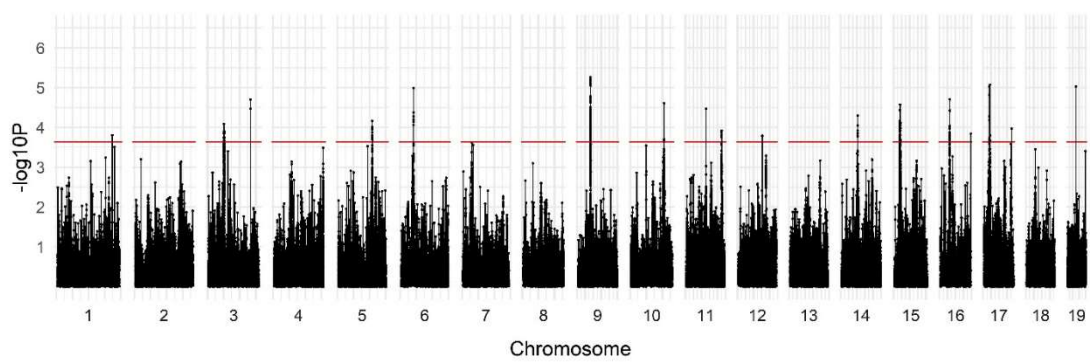
K
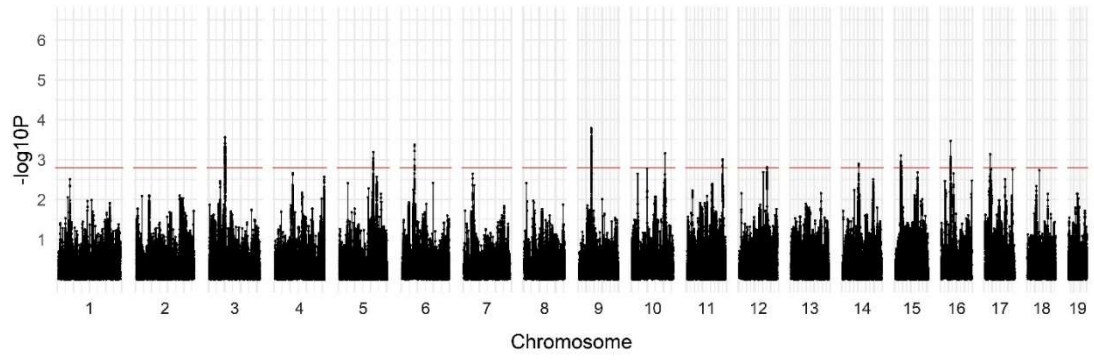
L
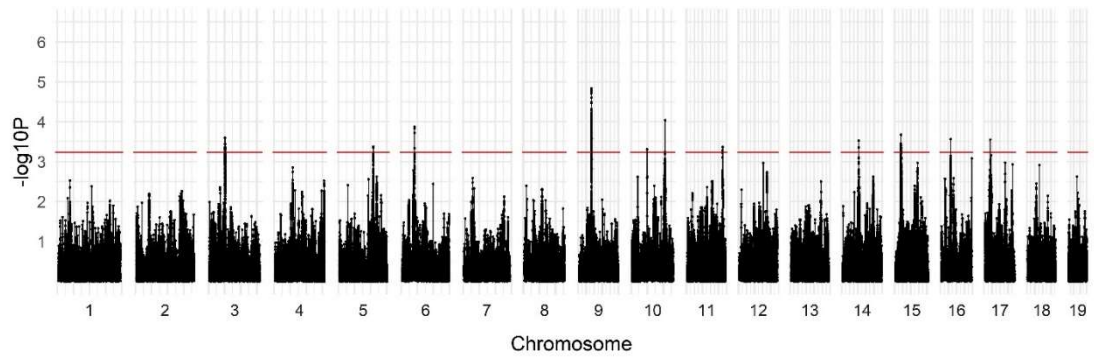
Figure 3.3

Scatterplot comparing the -log p-values of the generation 61 mixed model analyses (individual mouse) with -log p-values produced when these same data are treated as pooled sequencing allele frequencies and analyzed with (A) T-test, (B) regularized T-test, and (C) windowed, regularized T-test. Red line has intercept = 0 and slope = 1. Green line represents the least squares regression line.
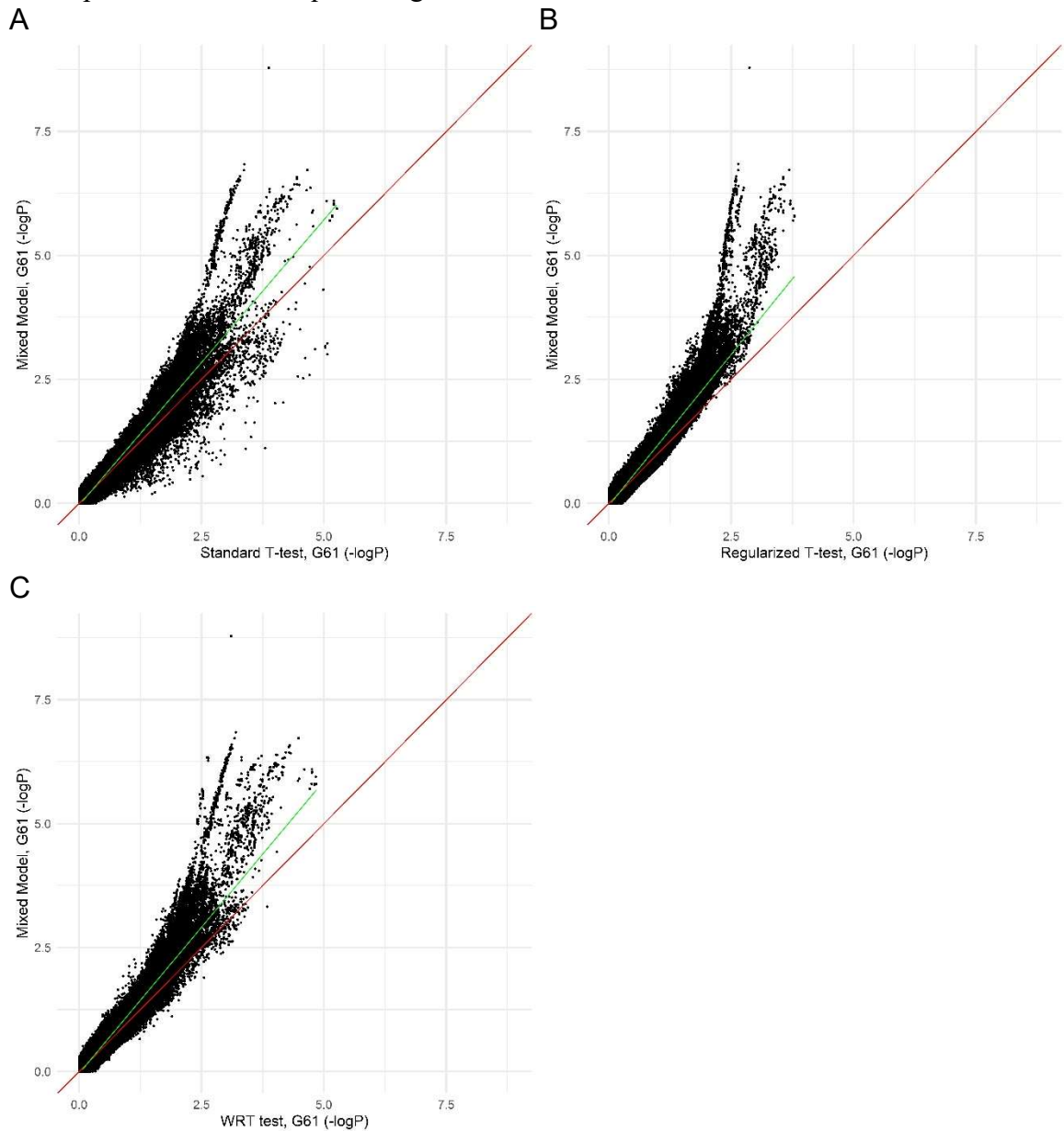
Figure 3.4
Simulations for a population allele frequency of 0.5 for (A) generation 22 and (B)
generation 61. See Methods for details. Values shown are allele frequencies for each of
100,000 simulated data sets for a single line. Methodological differences in the sampling
of mice and sequencing procedures for the two generations result in greater sampling
error for generation 22 (i.e., larger SD). Note that binning is done such that loci that fall
on a break (e.g., 0.05) are grouped into the lower bin (e.g., 0 to 0.05).
Similar simulations were then conducted to create data sets for use in estimating
statistical power for detecting selection signatures for generations 22 and 61. (C)
Distribution of p-values for simulated allele frequencies of 0.4 versus 0.6, for generation
22. Power for a = 0.05 is 0.3864. (D) Distribution of p-values for simulated allele
frequencies of 0.4 versus 0.6, for generation 61. Power for a = 0.05 is 0.5031.

A
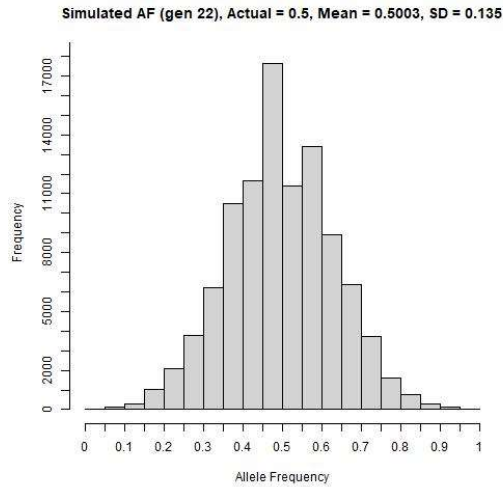
B
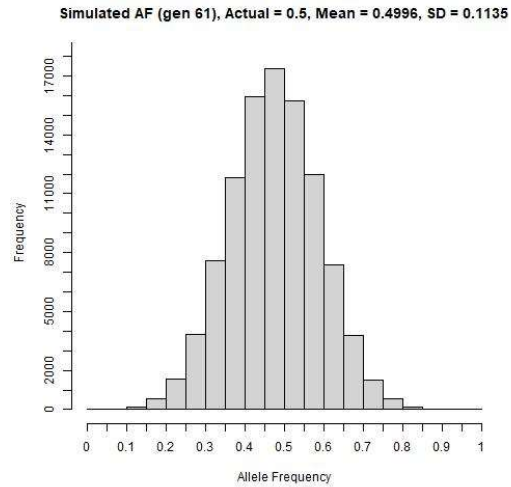
C
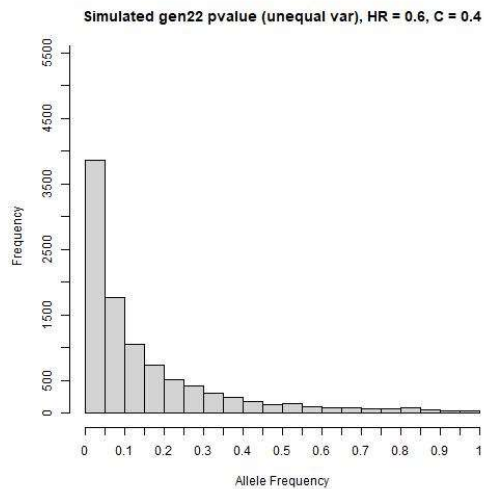
D
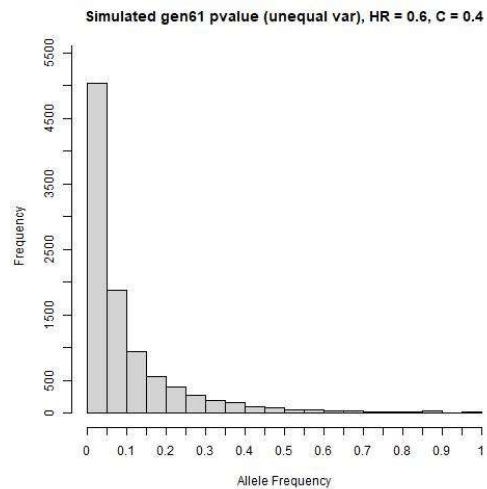
Figure 3.5
Running levels for (A) mean running of 4 non-selected control lines (blue) for 200
simulations compared with mean running of 4 HR lines under 100 unconstrained
simulations (dark red) and 100 constrained simulations (light red). (B) Individual HR
and control lines single unconstrained simulation. (C) Individual HR and control lines
single constrained simulation. X-axis is generations 1 through 61. Y-axis is revolutions.

A

B

C
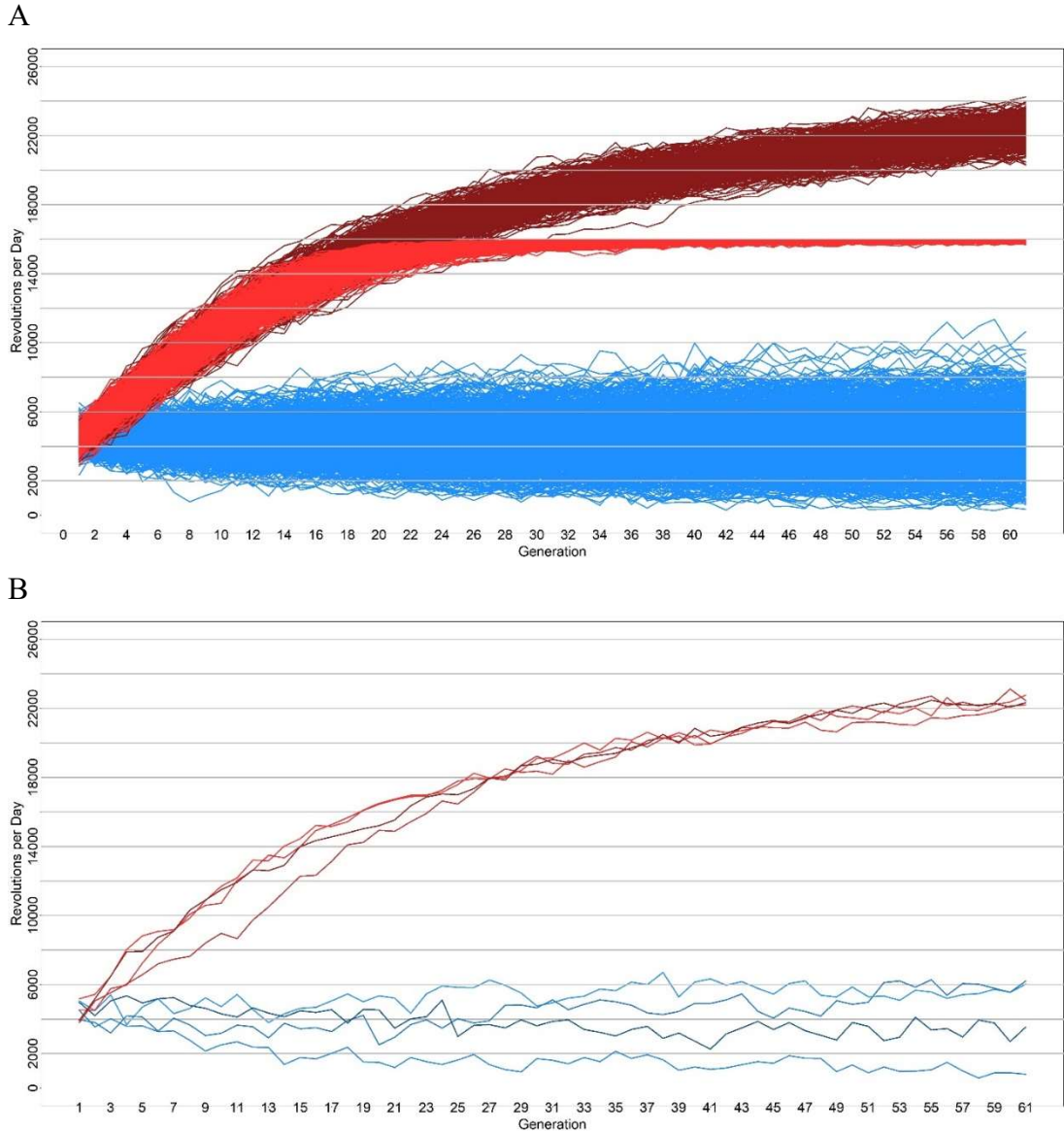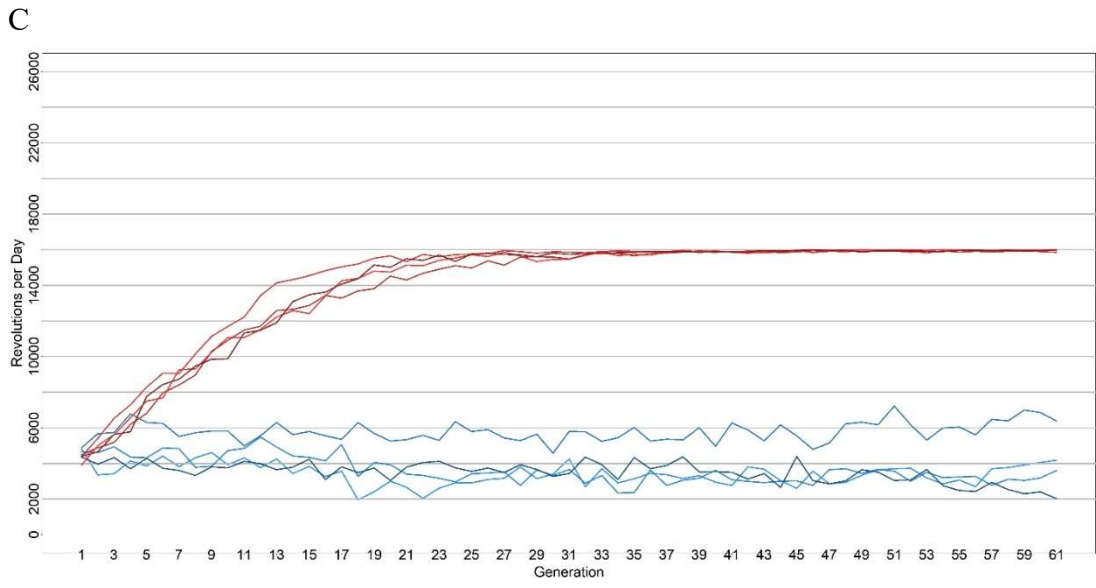
Figure 3.6
Black lines show the mean heritability for four lines within each linetype for (A) 100
control line simulations (arbitrarily the first 50 from each model), (B) 100 simulations for
HR lines with unconstrained running, and (C) 100 simulations for HR lines with
constrained running (16,000 revolutions). The red lines indicate overall average (mean of
the 100 black lines). Note that due to all parents sometimes having identical values (or
offspring perfectly matching parents) in the constrained simulations (C), heritability
could not be calculated as the constraint became frequently reached. Due to dropping
missing values from mean heritability calculations under the constrained model, large
fluctuations in heritability can be seen as early as generation 16, and past generation 40
the grand mean (red line) perfectly matches the one line for which heritability can still be
calculated (see Results for further explanation).
A

B

C

205
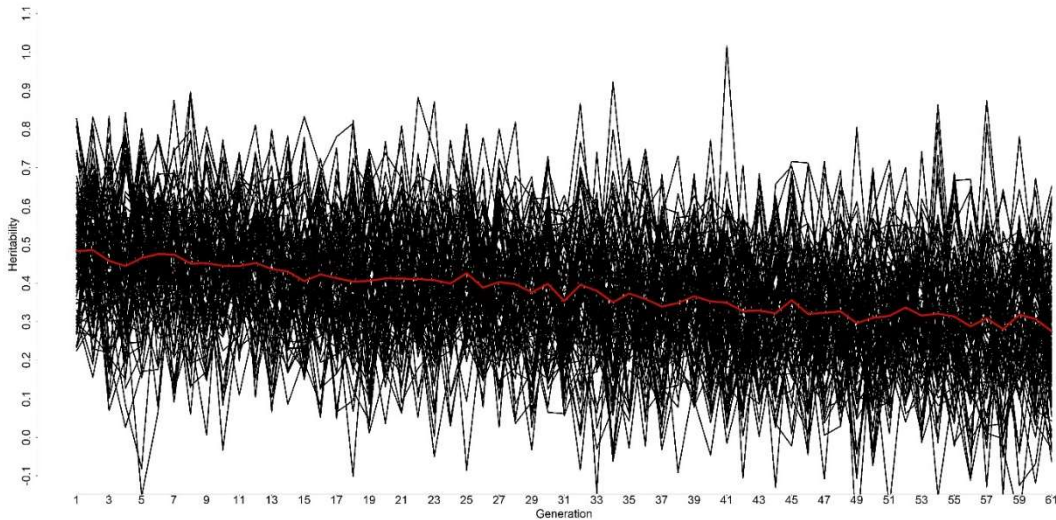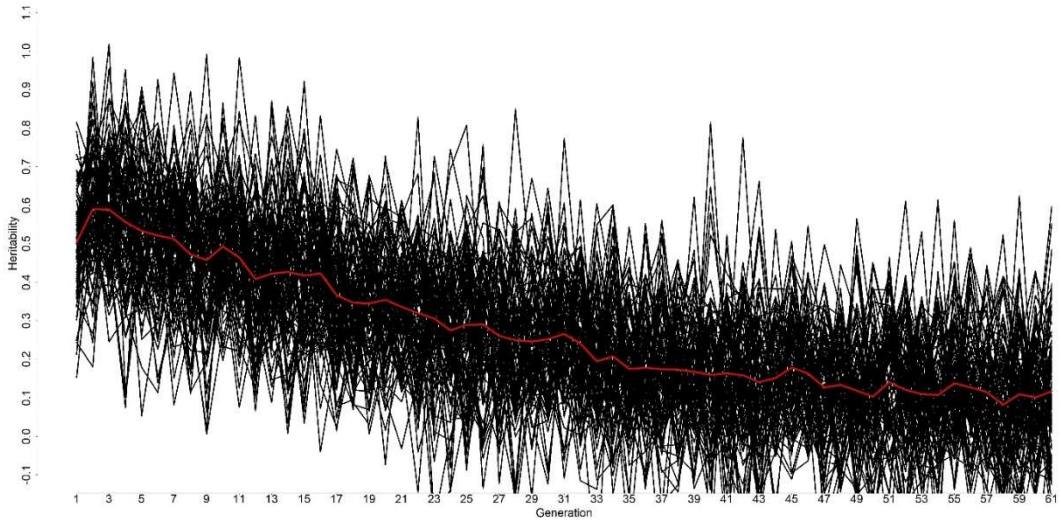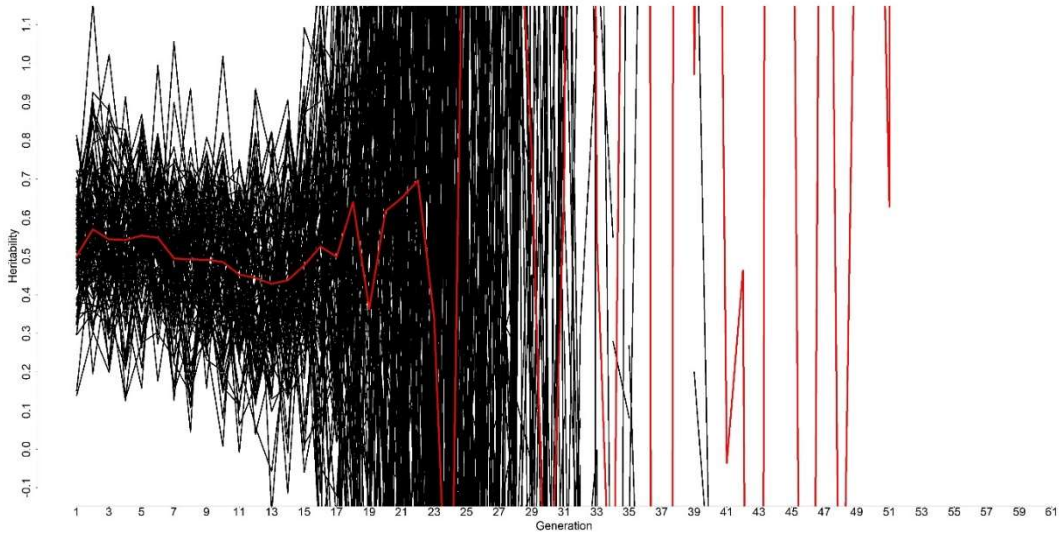
Figure 3.7
Standardized selection differentials (calculated within family and sex) from simulations for (A) control lines, (B) HR lines without a constraint, and (C) HR lines with a constraint. The red line represents the mean of all selection differentials (N = 100). Selection differential is set to 0 when running levels are 16,000 for all mice in the family/sex. Note different axes for panel A versus B and C.

A

B
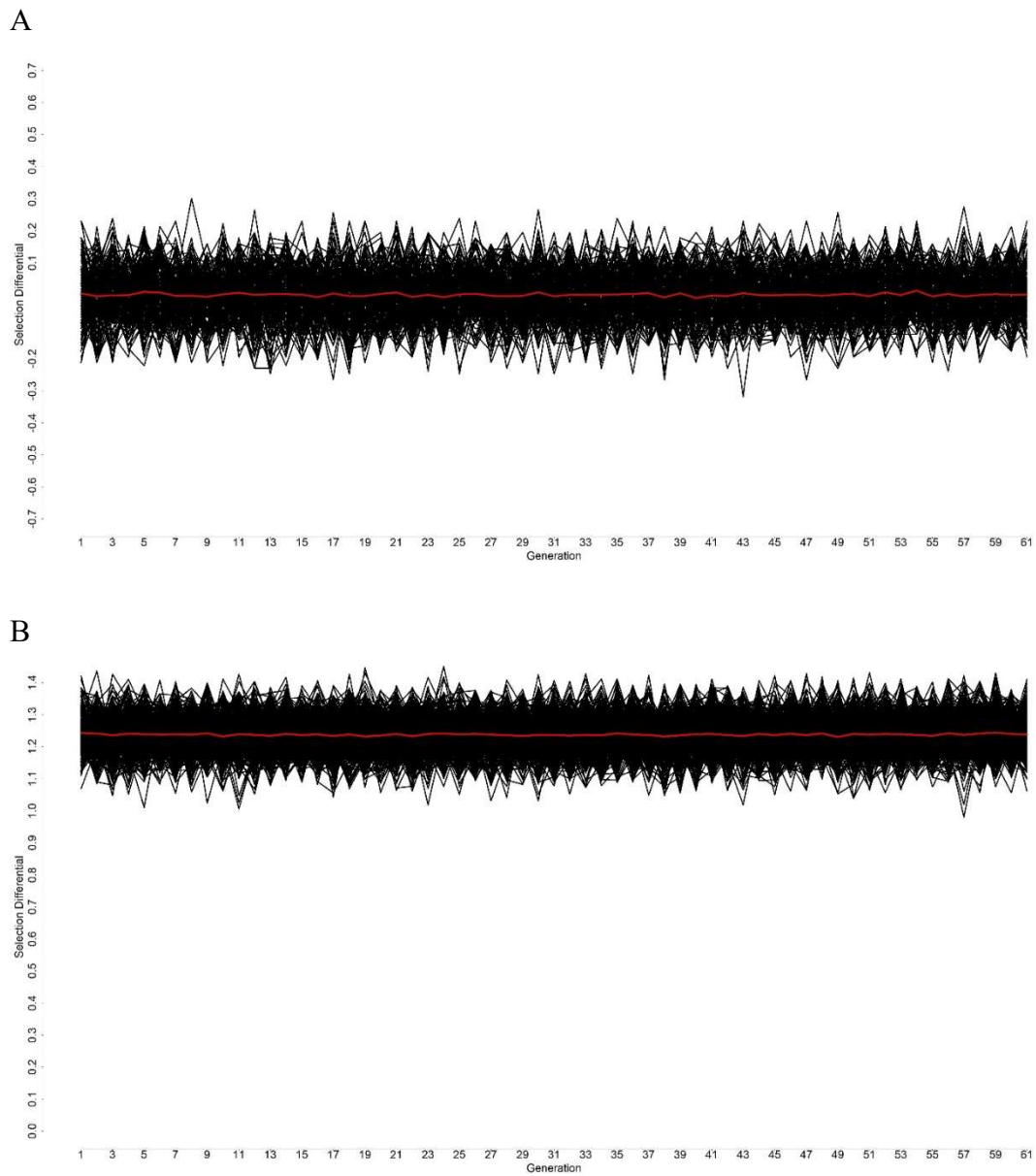
C
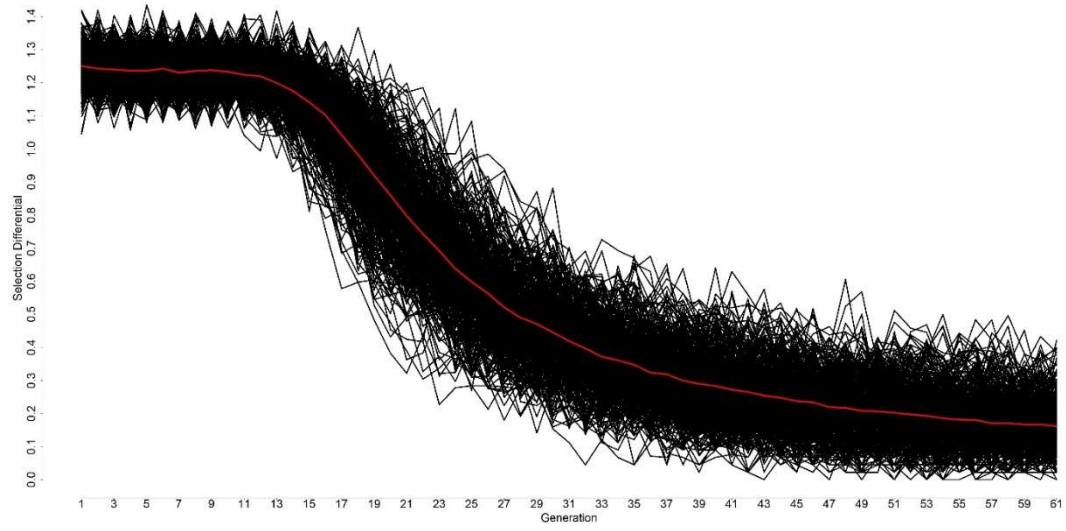
Figure 3.8

Allele frequencies of regions identified as significant (via strict culling) at generation 22 (Table 3.2) (excepting region 18, for which no loci were available in the generation 61 data). Each panel represents a separate region and its included SNPs with nominal p<0.05 at generation 22 and any SNPs at generation 61 which matched the generation 22 SNPs (shared loci, see Table 3.1). The allele frequencies of these SNPs were averaged for each line and generation and line graphs created (one for each line) with generation 22 AF on the left and generation 61 AF on the right. Control lines are represented in a blue-like color and HR lines represented in a red-like color. Only region 16 was found differentiated by Hillis et al. (2020).
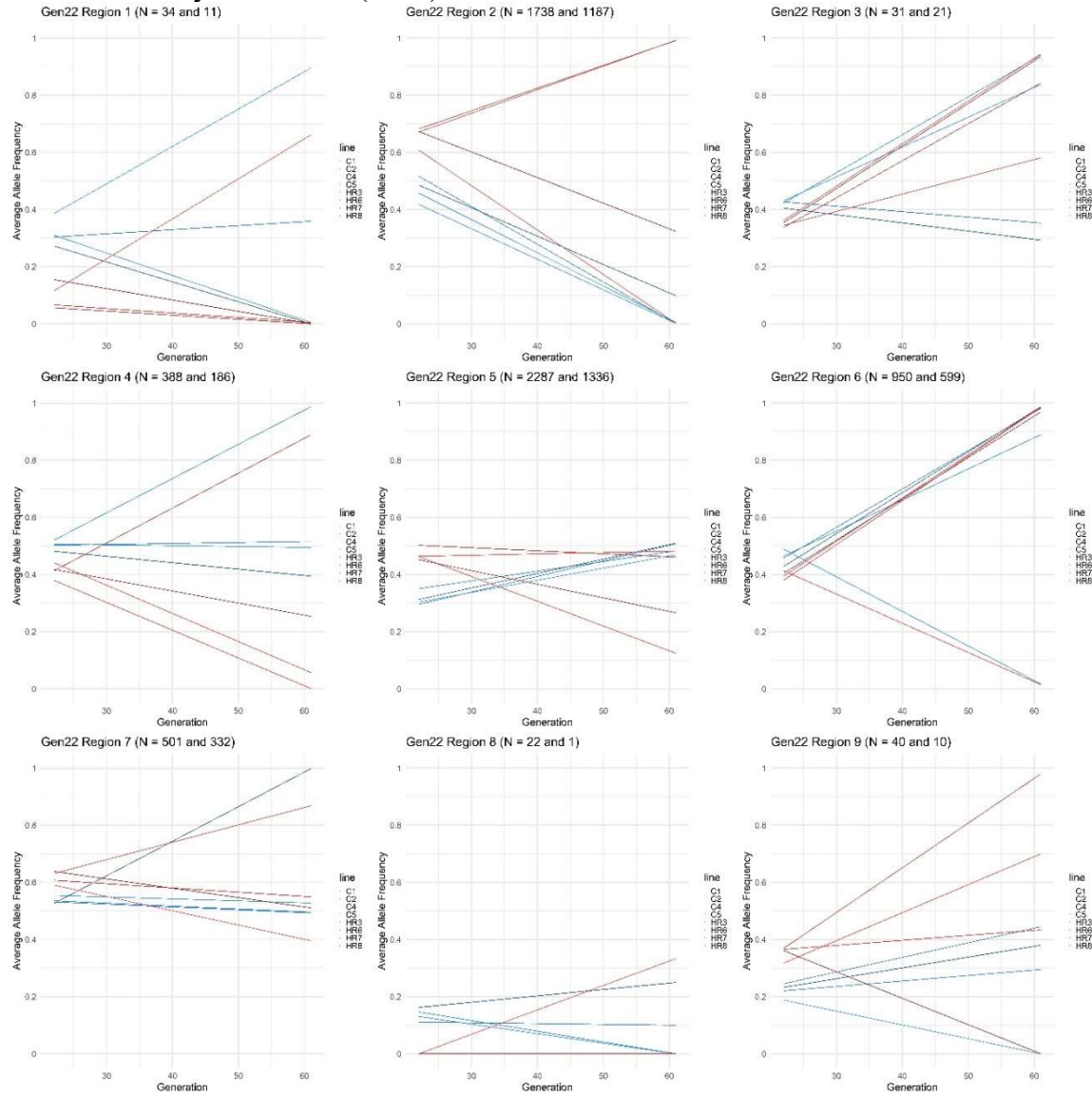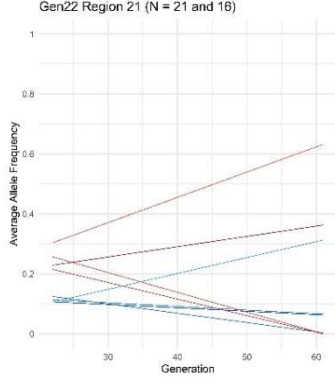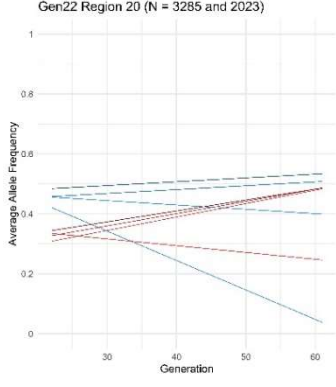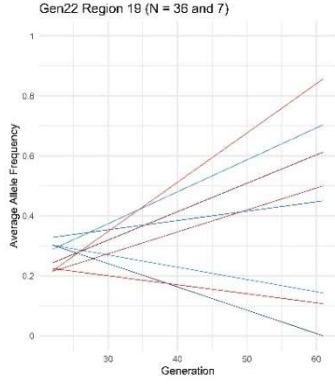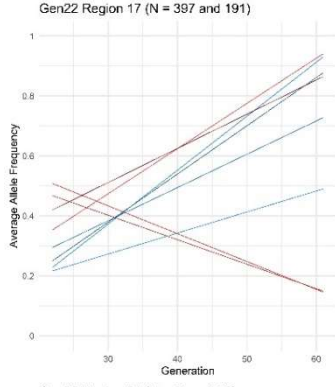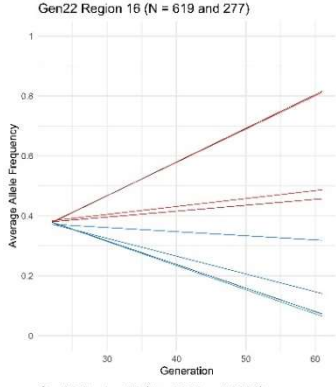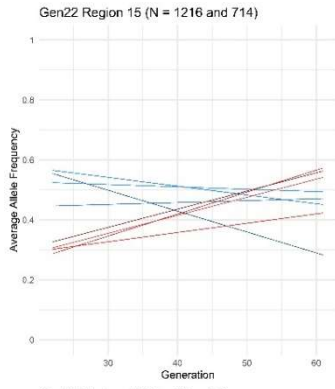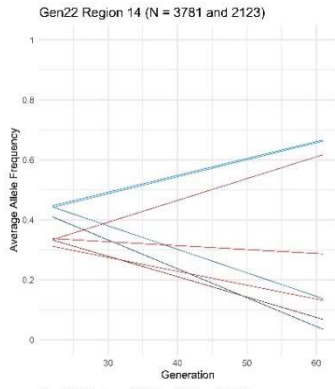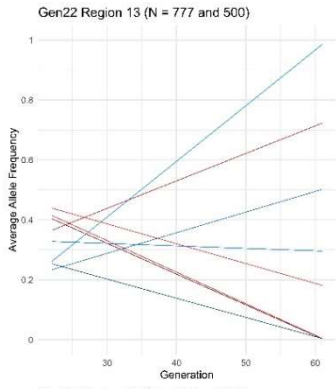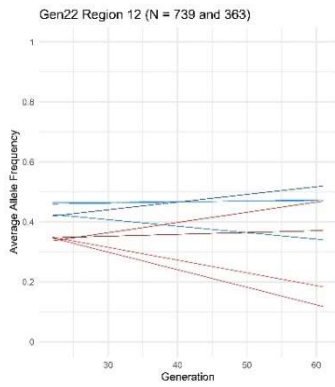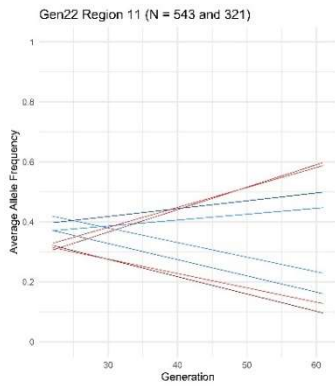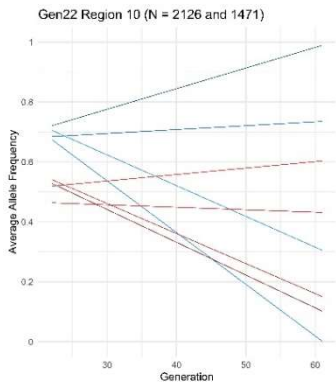
Figure 3.9

Allele frequencies of regions identified as significant (via strict culling) at generation 61 (Table 3.2). Each panel represents a separate region and its included SNPs with nominal p<0.05 at generation 61 and any SNPs at generation 22 which matched the generation 61 SNPs (shared loci, see Table 3.1). The allele frequencies of these SNPs were averaged for each line and generation and line graphs created (one for each line) with generation 22 AF on the left and generation 61 AF on the right. Control lines are represented in a blue-like color and HR lines represented in a red-like color. In order to identify how many other regions in the genome look like region 6 (gen61 results), I scanned the genome for loci where the maximum allele frequency for the HR lines was lower in generation 22 than the minimum allele frequency for the C lines, but the minimum AF for the HR was higher than the maximum AF for C lines at generation 61. I also scanned for the inverse scenario as well. This method produced 47 loci falling into 18 regions separated by at least 1 million bp. None of which correspond to region 6 since region 6 is the average AF per line over a region of the genome containing 441 SNPs. More creative scanning methods will be needed to identify truly similar regions. However, this minimum vs maximum comparison producing 47 loci would seem to indicate that such swaps in relative allele frequencies should be uncommon (though gen22 region 15, almost does the same thing and is also not amount the 18 regions identified by the scan through SNPs).
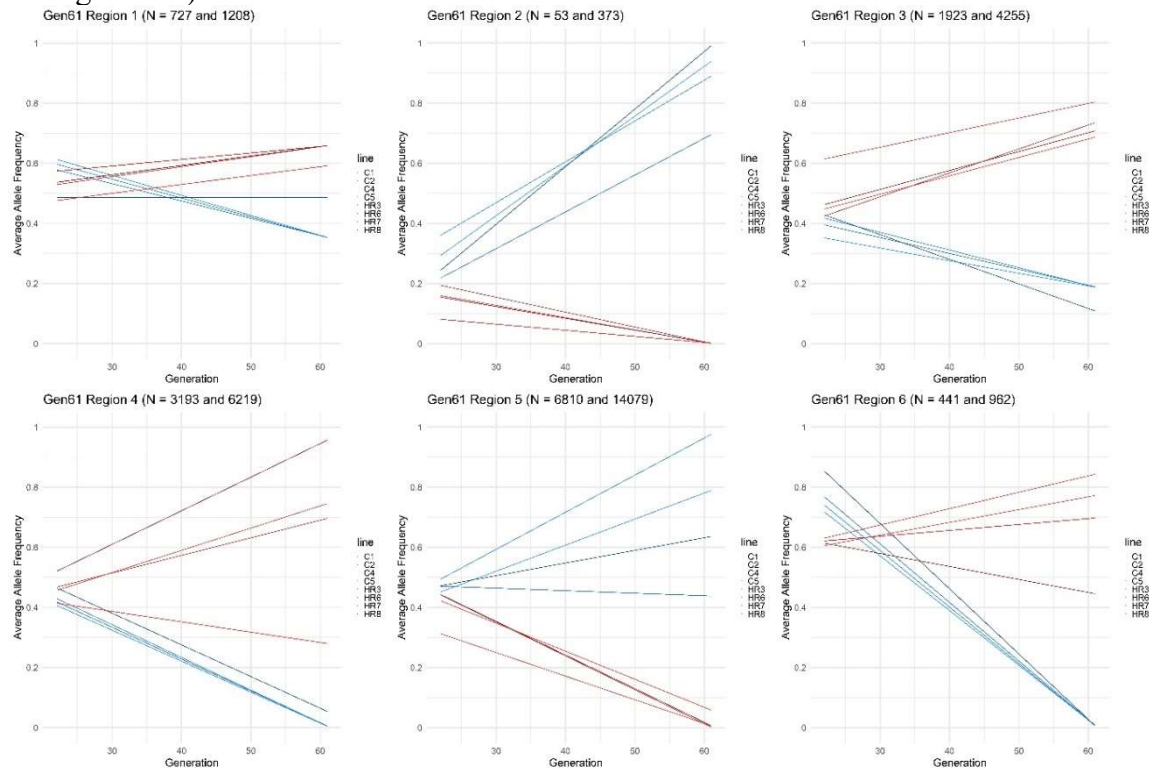
Figure 3.10

Venn Diagrams showing the overlap of SNPs "shared" by generations 22 and 61 and significant at FDR = 0.01 (pooled analyses) for:
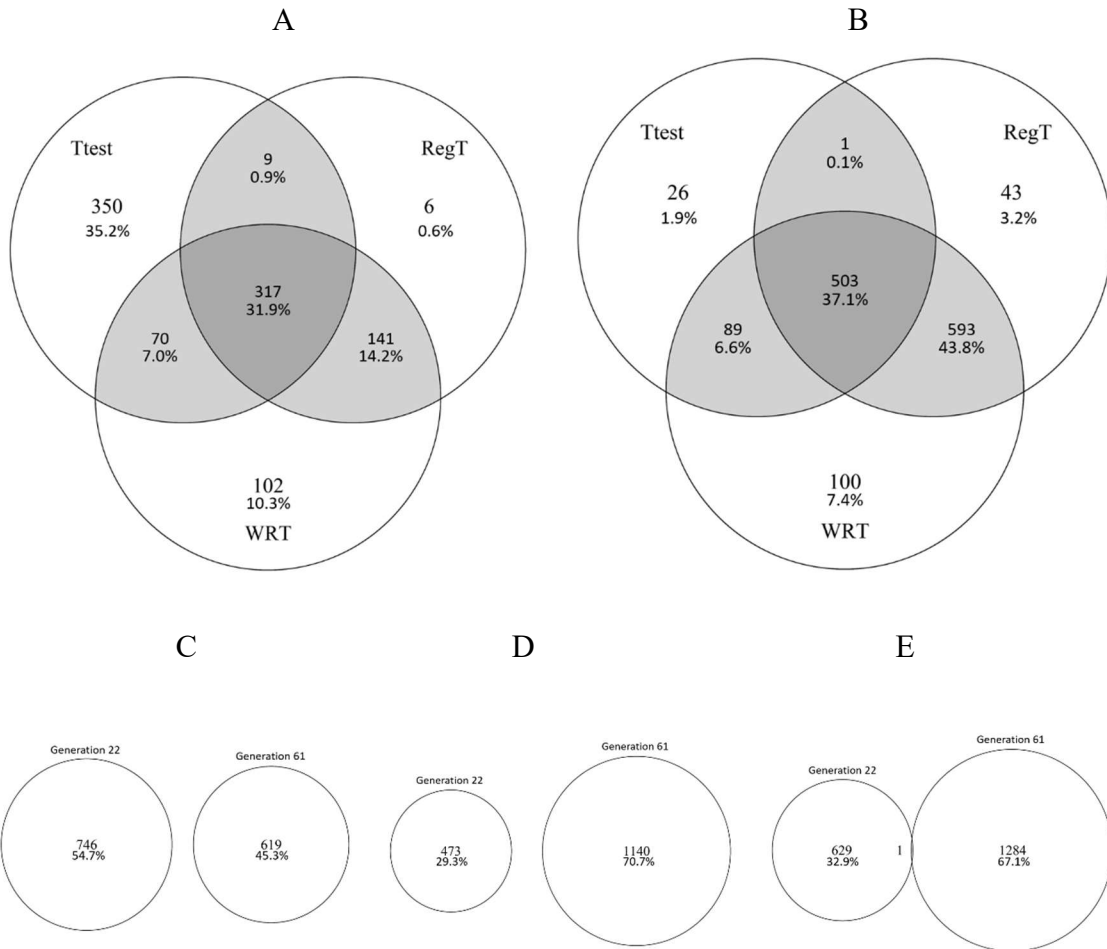
(A) generation 22 Ttest, RegT, and WRT analyses (N = 995)
(B) generation 61 Ttest, RegT, and WRT analyses (N = 1,355)
(C) generations 22 and 61 Ttest analyses (circle area proportionate to SNPs)
(D) generations 22 and 61 RegT analyses (circle area proportionate to SNPs)
(E) generations 22 and 61 WRT analyses (circle area proportionate to SNPs)

REFERENCES

Ahrens W. H., D. J. Cox, and G. Budhwar, 1990 Use of the arcsine and square root transformations for subjectively determined percentage data. Weed Sci. 38: 452–458. https://doi.org/10.1017/S0043174500056824

Al-Murrani W. K., and R. C. Roberts, 1974 Genetic variation in a line of mice selected to its limit for high body weight. Anim. Sci. 19: 273–289. https://doi.org/10.1017/S0003356100022856

Arora S., M. Morgan, M. Carlson, and H. Pages, 2020 GenomeInfoDb: Utilities for manipulating chromosome names, including modifying them to follow a particular naming style. R Package Version 1242.

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

Baldwin-Brown J. G., A. D. Long, and K. R. Thornton, 2014 The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. Mol. Biol. Evol. 31: 1040–1055. https://doi.org/10.1093/molbev/msu048

Barton N. H., and M. Turelli, 1989 Evolutionary quantitative genetics: how little do we know? Annu. Rev. Genet. 23: 337–370. https://doi.org/10.1146/annurev.ge.23.120189.002005

Benjamini Y., and Y. Hochberg, 1995 Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57: 289–300.

Borgius L., H. Nishimaru, V. Caldeira, Y. Kunugise, P. Low, *et al.*, 2014 Spinal glutamatergic neurons defined by *Epha4* signaling are essential components of normal locomotor circuits. J. Neurosci. 34: 3841–3853. https://doi.org/10.1523/JNEUROSCI.4992-13.2014

Brown W. P., and A. E. Bell, 1961 Genetic analysis of a "plateaued" population of *Drosophila melanogasteri*. Genetics 46: 407–425. https://doi.org/10.1093/genetics/46.4.407

Bult A., and C. B. Lynch, 2000 Breaking through artificial selection limits of an adaptive behavior in mice and the consequences for correlated responses. Behav. Genet. 30: 193–206. https://doi.org/10.1023/a:1001962124005

Burke M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 467: 587–590. https://doi.org/10.1038/nature09352

Cadney M. D., L. Hiramatsu, Z. Thompson, M. Zhao, J. C. Kay, *et al.*, 2021 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. 234: 113389. https://doi.org/10.1016/j.physbeh.2021.113389

Careau V., M. E. Wolak, P. A. Carter, and T. Garland, Jr., 2013 Limits to behavioral evolution: the quantative genetics of a complex trait under directional selection. Evolution 67: 3102–3119. https://doi.org/10.1111/evo.12200

Castro A. A., and T. Garland, Jr., 2018 Evolution of hindlimb bone dimensions and muscle masses in house mice selectively bred for high voluntary wheel-running behavior. J. Morphol. 279: 766–779. https://doi.org/10.1002/jmor.20809

Castro J. P., M. N. Yancoskie, M. Marchini, S. Belohlavy, L. Hiramatsu, *et al.*, 2019 An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. eLife 8: e40214. https://doi.org/10.7554/eLife.42014

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

Chaouloff F., S. Dubreucq, L. Bellocchio, and G. Marsicano, 2011 Endocannabinoids and Motor Behavior: CB1 Receptors Also Control Running Activity. Physiology 26: 76–77. https://doi.org/10.1152/physiol.00050.2010

Chen J., H. Xu, B. J. Aronow, and A. G. Jegga, 2007 Improved human disease candidate gene prioritization using mouse phenotype. BMC Bioinformatics 8: 392. https://doi.org/10.1186/1471-2105-8-392

Copes L. E., H. Schutz, E. M. Dlugosz, W. Acosta, M. A. Chappell, *et al.*, 2015 Effects of voluntary exercise on spontaneous physical activity and food consumption in mice: Results from an artificial selection experiment. Physiol. Behav. 149: 86–94. https://doi.org/10.1016/j.physbeh.2015.05.025

Copes L. E., H. Schutz, E. M. Dlugsoz, S. Judex, and T. Garland, Jr., 2018 Locomotor activity, growth hormones, and systemic robusticity: An investigation of cranial vault thickness in mouse lines bred for high endurance running. Am. J. Phys. Anthropol. 166: 442–458. https://doi.org/10.1002/ajpa.23446

Corda D., T. Kudo, P. Zizza, C. Iurisci, E. Kawai, *et al.*, 2009 The developmentally regulated osteoblast phosphodiesterase GDE3 is glycerophosphoinositol-specific and modulates cell growth. J. Biol. Chem. 284: 24848–24856. https://doi.org/10.1074/jbc.M109.035444

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

Didion J. P., A. P. Morgan, L. Yadgary, T. A. Bell, R. C. McMullan, *et al.*, 2016 *R2d2* drives selfish sweeps in the house mouse. Mol. Biol. Evol. 33: 1381–1395. https://doi.org/10.1093/molbev/msw036

Dlugosz E. M., H. Schutz, T. H. Meek, W. Acosta, C. J. Downs, *et al.*, 2013 Immune response to a Trichinella spiralis infection in house mice from lines selectively bred for high voluntary wheel running. J. Exp. Biol. 216: 4212–4221. https://doi.org/10.1242/jeb.087361

Dobzhansky T., and B. Spassky, 1969 Artificial and natural selection for two behavioral traits in Drosophila pseudoobscura. Proc. Natl. Acad. Sci. U. S. A. 62: 75–80. https://doi.org/10.1073/pnas.62.1.75

Dottori M., L. Hartley, M. Galea, G. Paxinos, M. Polizzotto, *et al.*, 1998 EphA4 (Sek1) receptor tyrosine kinase is required for the development of the corticospinal tract. Proc. Natl. Acad. Sci. 95: 13248–13253. https://doi.org/10.1073/pnas.95.22.13248

Douhard F., M. Douhard, H. Gilbert, P. Monget, J. Gaillard, *et al.*, 2021 How much energetic trade-offs limit selection? Insights from livestock and related laboratory model species. Evol. Appl. 14: 2726–2749. https://doi.org/10.1111/eva.13320

Dumke C. L., J. S. Rhodes, T. Garland, E. Maslowski, J. G. Swallow, *et al.*, 2001 Genetic selection of mice for high voluntary wheel running: effect on skeletal muscle glucose uptake. J. Appl. Physiol. 91: 1289–1297. https://doi.org/10.1152/jappl.2001.91.3.1289

Falconer D. S., 1989 *Introduction to quantitative genetics*. Longman, Scientific & Technical ; Wiley, Burnt Mill, Harlow, Essex, England : New York.

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

Garland T., C. J. Downs, and A. R. Ives, 2022 Trade-offs (and constraints) in organismal biology. Physiol. Biochem. Zool. 95: 82–112. https://doi.org/10.1086/717897

Garland, Jr. T., and P. W. Freeman, 2005 Selective breeding for high endurance running increases hindlimb symmetry. Evolution 59: 1851–1854.

Garland, Jr. T., H. Schutz, M. A. Chappell, B. K. Keeney, T. H. Meek, *et al.*, 2011a The biological control of voluntary exercise, spontaneous physical activity and daily energy expenditure in relation to obesity: human and rodent perspectives. J. Exp. Biol. 214: 206–229. https://doi.org/10.1242/jeb.048397

Garland, Jr. T., S. A. Kelly, J. L. Malisch, E. M. Kolb, R. M. Hannon, *et al.*, 2011b How to run far: multiple solutions and sex-specific responses to selective breeding for high voluntary activity levels. Proc. R. Soc. B Biol. Sci. 278: 574–581. https://doi.org/10.1098/rspb.2010.1584

Gilley J., R. Adalbert, G. Yu, and M. P. Coleman, 2013 Rescue of peripheral and CNS axon defects in mice lacking *Nmnat2*. J. Neurosci. 33: 13410–13424. https://doi.org/10.1523/JNEUROSCI.1534-13.2013

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

Hesslein D. G. T., J. A. Fretz, Y. Xi, T. Nelson, S. Zhou, *et al.*, 2009 Ebf1-dependent control of the osteoblast and adipocyte lineages. Bone 44: 537–546. https://doi.org/10.1016/j.bone.2008.11.021

Hillis D. A., L. Yadgary, G. M. Weinstock, F. Pardo-Manuel de Villena, D. Pomp, *et al.*, 2020 Genetic basis of aerobically supported voluntary exercise: results from a selection experiment with house mice. Genetics 216: 781–804. https://doi.org/10.1534/genetics.120.303668

Hillis D. A., and T. Garland Jr, 2022 Multiple solutions at the genomic level in response to selective breeding for high locomotor activity. Genetics. https://doi.org/10.1093/genetics/iyac165

Hiramatsu L., J. C. Kay, Z. Thompson, J. M. Singleton, G. C. Claghorn, *et al.*, 2017 Maternal exposure to Western diet affects adult body composition and voluntary wheel running in a genotype-specific manner in mice. Physiol. Behav. 179: 235–245. https://doi.org/10.1016/j.physbeh.2017.06.008

Kelly S. A., P. P. Czech, J. T. Wight, K. M. Blank, and T. Garland, Jr., 2006 Experimental evolution and phenotypic plasticity of hindlimb bones in high-activity house mice. J. Morphol. 267: 360–374. https://doi.org/10.1002/jmor.10407

Kelly S. A., F. R. Gomes, E. M. Kolb, J. L. Malisch, and T. Garland, Jr., 2017 Effects of activity, genetic selection and their interaction on muscle metabolic capacities and organ masses in mice. J. Exp. Biol. 220: 1038–1047. https://doi.org/10.1242/jeb.148759

Kim M.-J., J.-Y. Son, J.-S. Ju, and D.-K. Ahn, 2020 Early blockade of Epha4 pathway reduces trigeminal neuropathic pain. J. Pain Res. Volume 13: 1173–1183. https://doi.org/10.2147/JPR.S249185

Kolb E. M., S. A. Kelly, K. M. Middleton, L. S. Sermsakdi, M. A. Chappell, *et al.*, 2010 Erythropoietin elevates VO2,max but not voluntary wheel running in mice. J. Exp. Biol. 213: 510–519. https://doi.org/10.1242/jeb.029074

Kolb E. M., E. L. Rezende, L. Holness, A. Radtke, S. K. Lee, *et al.*, 2013a Mice selectively bred for high voluntary wheel running have larger midbrains: support for the mosaic model of brain evolution. J. Exp. Biol. 216: 515–523. https://doi.org/10.1242/jeb.076000

Kolb E. M., S. A. Kelly, and T. Garland, Jr., 2013b Mice from lines selectively bred for high voluntary wheel running exhibit lower blood pressure during withdrawal from wheel access. Physiol. Behav. 112–113: 49–55. https://doi.org/10.1016/j.physbeh.2013.02.010

Konczal M., W. Babik, J. Radwan, E. T. Sadowska, and P. Koteja, 2015 Initial molecular-level response to artificial selection for increased aerobic metabolism occurs primarily through changes in gene expression. Mol. Biol. Evol. 32: 1461–1473. https://doi.org/10.1093/molbev/msv038

Konczal M., P. Koteja, P. Orlowska-Feuer, J. Radwan, E. T. Sadowska, *et al.*, 2016 Genomic response to selection for predatory behavior in a mammalian model of adaptive radiation. Mol. Biol. Evol. 33: 2429–2440. https://doi.org/10.1093/molbev/msw121

Labialle S., V. Marty, M. Bortolin-Cavaillé, M. Hoareau-Osman, J. Pradère, *et al.*, 2014 The miR-379/miR-410 cluster at the imprinted *Dlk1-Dio3* domain controls neonatal metabolic adaptation. EMBO J. 33: 2216–2230. https://doi.org/10.15252/embj.201387038

Lau K.-H. W., and M. H.-C. Sheng, 2018 A novel miR17 /protein tyrosine phosphatase-oc/EphA4 regulatory axis of osteoclast activity. Arch. Biochem. Biophys. 650: 30–38. https://doi.org/10.1016/j.abb.2018.05.014

Lerner I. M., and E. R. Dempster, 1951 Attenuation of genetic progress under continued selection in poultry. Heredity 5: 75–94. https://doi.org/10.1038/hdy.1951.4

Li J., W. Dong, X. Gao, W. Chen, C. Sun, *et al.*, 2021 EphA4 is highly expressed in the atria of heart and its deletion leads to atrial hypertrophy and electrocardiographic abnormalities in rats. Life Sci. 278: 119595. https://doi.org/10.1016/j.lfs.2021.119595

Lightfoot J. T., E. J. C. De Geus, F. W. Booth, M. S. Bray, M. Den Hoed, *et al.*, 2018 Biological/genetic regulation of physical activity level: Consensus from GenBioPAC. Med. Sci. Sports Exerc. 50: 863–873. https://doi.org/10.1249/MSS.0000000000001499

Lillie M., C. F. Honaker, P. B. Siegel, and Ö. Carlborg, 2019 Bidirectional selection for body weight on standing genetic variation in a chicken model. Genes|Genomes|Genetics g3.400038.2019. https://doi.org/10.1534/g3.119.400038

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

MacKay H., C. A. Scott, J. D. Duryea, M. S. Baker, E. Laritsky, *et al.*, 2019 DNA methylation in AgRP neurons regulates voluntary exercise behavior in mice. Nat. Commun. 10. https://doi.org/10.1038/s41467-019-13339-3

Mathes W. F., D. L. Nehrenberg, R. Gordon, K. Hua, T. Garland, Jr., *et al.*, 2010 Dopaminergic dysregulation in mice selectively bred for excessive exercise or obesity. Behav. Brain Res. 210: 155–163. https://doi.org/10.1016/j.bbr.2010.02.016

Meek T. H., B. P. Lonquich, R. M. Hannon, and T. Garland, Jr., 2009 Endurance capacity of mice selectively bred for high voluntary wheel running. J. Exp. Biol. 212: 2908–2917. https://doi.org/10.1242/jeb.028886

Middleton K. M., C. E. Shubin, D. C. Moore, P. A. Carter, T. Garland, Jr., *et al.*, 2008 The relative importance of genetics and phenotypic plasticity in dictating bone morphology and mechanics in aged mice: Evidence from an artificial selection experiment. Zoology 111: 135–147. https://doi.org/10.1016/j.zool.2007.06.003

Middleton K. M., B. D. Goldstein, P. R. Guduru, J. F. Waters, S. A. Kelly, *et al.*, 2010 Variation in within-bone stiffness measured by nanoindentation in mice bred for high levels of voluntary wheel running. J. Anat. 216: 121–131. https://doi.org/10.1111/j.1469-7580.2009.01175.x

Mohd-Zin S. W., N.-L. Abdullah, A. Abdullah, N. D. E. Greene, P.-S. Cheah, *et al.*, 2016 Identification of the genomic mutation in *Epha4* $^{rb-2J/rb-2J}$ mice, (M. E. Cristescu, Ed.). Genome 59: 439–448. https://doi.org/10.1139/gen-2015-0142

Nguyen Q. A. T., D. Hillis, S. Katada, T. Harris, C. Pontrello, *et al.*, 2020 Coadaptation of the chemosensory system with voluntary exercise behavior in mice, (H. Matsunami, Ed.). PLOS ONE 15: e0241758. https://doi.org/10.1371/journal.pone.0241758

Nugent A. A., J. G. Park, Y. Wei, A. P. Tenney, N. M. Gilette, *et al.*, 2017 Mutant α2-chimaerin signals via bidirectional ephrin pathways in Duane retraction syndrome. J. Clin. Invest. 127: 1664–1682. https://doi.org/10.1172/JCI88502

Obenchain V., M. Lawrence, V. Carey, S. Gogarten, P. Shannon, *et al.*, 2014 VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. Bioinformatics 30: 2076–2078. https://doi.org/10.1093/bioinformatics/btu168

Palpant N. J., M. L. Szatkowski, W. Wang, D. Townsend, F. B. Bedada, *et al.*, 2009 Artificial selection for whole animal low intrinsic aerobic capacity co-segregates with hypoxia-induced cardiac pump failure, (J. A. L. Calbet, Ed.). PLoS ONE 4: e6117. https://doi.org/10.1371/journal.pone.0006117

Porto A., R. Schmelter, J. L. VandeBerg, G. Marroig, and J. M. Cheverud, 2016 Evolution of the genotype-to-phenotype map and the cost of pleiotropy in mammals. Genetics 204: 1601–1612. https://doi.org/10.1534/genetics.116.189431

Pratt A. J., and I. J. MacRae, 2009 The RNA-induced Silencing Complex: A Versatile Gene-silencing Machine. J. Biol. Chem. 284: 17897–17901. https://doi.org/10.1074/jbc.R900012200

Reeve J. P., 2000 Predicting long-term response to selection. Genet. Res. 75: 83–94. https://doi.org/10.1017/S0016672399004140

Reeve J. P., and D. J. Fairbairn, 2001 Predicting the evolution of sexual size dimorphism: Predicting the evolution of SSD. J. Evol. Biol. 14: 244–254. https://doi.org/10.1046/j.1420-9101.2001.00276.x

Ren Y., L. G. Koch, S. L. Britton, N. R. Qi, M. K. Treutelaar, *et al.*, 2016 Selection-, age-, and exercise-dependence of skeletal muscle gene expression patterns in a rat model of metabolic fitness. Physiol. Genomics 48: 816–825. https://doi.org/10.1152/physiolgenomics.00118.2015

Rezende E. L., F. R. Gomes, M. A. Chappell, and T. Garland Jr., 2009 Running behavior and its energy cost in mice selectively bred for high voluntary locomotor activity. Physiol. Biochem. Zool. 82: 662–679. https://doi.org/10.1086/605917

Rhodes J. S., G. R. Hosack, I. Girard, A. E. Kelley, G. S. Mitchell, *et al.*, 2001 Differential sensitivity to acute administration of cocaine, GBR 12909, and

fluoxetine in mice selectively bred for hyperactive wheel-running behavior. Psychopharmacology (Berl.) 158: 120–131. https://doi.org/10.1007/s002130100857

Rhodes J. S., H. van Praag, S. Jeffrey, I. Girard, G. S. Mitchell, *et al.*, 2003 Exercise increases hippocampal neurogenesis to high levels but does not improve spatial learning in mice bred for increased voluntary wheel running. Behav. Neurosci. 117: 1006–1016. https://doi.org/10.1037/0735-7044.117.5.1006

Rhodes J. S., S. C. Gammie, and T. Garland Jr, 2005 Neurobiology of mice selected for high voluntary wheel-running activity. Integr. Comp. Biol. 45: 438–455. https://doi.org/10.1093/icb/45.3.438

Roberts R. C., 1966 The limits to artificial selection for body weight in the mouse II. The Genetic Nature of the Limits. Genet. Res. 8: 361–375. https://doi.org/10.1017/S0016672300010211

Roberts M. D., L. Gilpin, K. E. Parker, T. E. Childs, M. J. Will, *et al.*, 2012 Dopamine D1 receptor modulation in nucleus accumbens lowers voluntary wheel running in rats bred to run high distances. Physiol. Behav. 105: 661–668. https://doi.org/10.1016/j.physbeh.2011.09.024

Roberts M. D., J. D. Brown, J. M. Company, L. P. Oberle, A. J. Heese, *et al.*, 2013 Phenotypic and molecular differences between rats selectively bred to voluntarily run high vs. low nightly distances. Am. J. Physiol.-Regul. Integr. Comp. Physiol. 304: R1024–R1035. https://doi.org/10.1152/ajpregu.00581.2012

Roberts M. D., R. G. Toedebusch, K. D. Wells, J. M. Company, J. D. Brown, *et al.*, 2014 Nucleus accumbens neuronal maturation differences in young rats bred for low *versus* high voluntary running behaviour. J. Physiol. 592: 2119–2135. https://doi.org/10.1113/jphysiol.2013.268805

Rose M. R., H. B. Passananti, A. K. Chippindale, J. P. Phelan, M. Matos, *et al.*, 2005 The effects of evolution are local: evidence from experimental evolution in drosophila. Integr. Comp. Biol. 45: 486–491. https://doi.org/10.1093/icb/45.3.486

Schlötterer C., R. Kofler, E. Versace, R. Tobler, and S. U. Franssen, 2015 Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity 114: 431–440.

Schmidt S., V. Gawlik, S. M. Hölter, R. Augustin, A. Scheepers, *et al.*, 2008 Deletion of glucose transporter GLUT8 in mice increases locomotor activity. Behav. Genet. 38: 396–406. https://doi.org/10.1007/s10519-008-9208-1

Schwartz N. L., B. A. Patel, T. Garland, Jr., and A. M. Horner, 2018 Effects of selective breeding for high voluntary wheel-running behavior on femoral nutrient canal size and abundance in house mice. J. Anat. 233: 193–203. https://doi.org/10.1111/joa.12830

Smith C. L., and J. T. Eppig, 2009 The mammalian phenotype ontology: enabling robust annotation and comparative analysis. WIREs Syst. Biol. Med. 1: 390–399. https://doi.org/10.1002/wsbm.44

Stephan W., 2016 Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. Mol. Ecol. 25: 79–88. https://doi.org/10.1111/mec.13288

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. https://doi.org/10.1023/A:1021479331779

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. https://doi.org/10.1152/jappl.1998.84.1.69

Swallow J. G., J. P. Hayes, P. Koteja, and T. Garland, 2009 Selection experiments and experimental evolution of performance and physiology, pp. 301–351 in *Experimental evolution: concepts, methods, and applications of selection experiments*, University of California Press, Berkeley.

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

The International Mouse Phenotyping Consortium, M. E. Dickinson, A. M. Flenniken, X. Ji, L. Teboul, *et al.*, 2016 High-throughput discovery of novel developmental phenotypes. Nature 537: 508–514. https://doi.org/10.1038/nature19356

Thompson Z., D. Argueta, T. Garland, Jr., and N. DiPatrizio, 2017 Circulating levels of endocannabinoids respond acutely to voluntary exercise, are altered in mice selectively bred for high voluntary wheel running, and differ between the sexes. Physiol. Behav. 170: 141–150. https://doi.org/10.1016/j.physbeh.2016.11.041

Thyfault J. P., R. S. Rector, G. M. Uptergrove, S. J. Borengasser, E. M. Morris, *et al.*, 2009 Rats selectively bred for low aerobic capacity have reduced hepatic mitochondrial oxidative capacity and susceptibility to hepatic steatosis and injury:

Aerobic capacity and hepatic mitochondria. J. Physiol. 587: 1805–1816.
https://doi.org/10.1113/jphysiol.2009.169060

Travisano M., and R. G. Shaw, 2013 Lost in the map. Evolution 67: 305–314.
https://doi.org/10.1111/j.1558-5646.2012.01802.x

Wallace I. J., K. M. Middleton, S. Lublinsky, S. A. Kelly, S. Judex, *et al.*, 2010
Functional significance of genetic variation underlying limb bone diaphyseal
structure. Am. J. Phys. Anthropol. 143: 21–30. https://doi.org/10.1002/ajpa.21286

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

Wallace I. J., and T. Garland, Jr., 2016 Mobility as an emergent property of biological
organization: Insights from experimental evolution: Mobility and biological
organization. Evol. Anthropol. Issues News Rev. 25: 98–104.
https://doi.org/10.1002/evan.21481

Wang M. M., and R. R. Reed, 1993 Molecular cloning of the olfactory neuronal
transcription factor Olf-1 by genetic selection in yeast. Nature 364: 121–126.
https://doi.org/10.1038/364121a0

Wang S. S., J. W. Lewcock, P. Feinstein, P. Mombaerts, and R. R. Reed, 2004 Genetic
disruptions of *O/E2* and *O/E3* genes reveal involvement in olfactory receptor
neuron projection. Development 131: 1377–1388.
https://doi.org/10.1242/dev.01009

Wang Y., C. Wen, G. Xie, and L. Jiang, 2021 Blockade of spinal Epha4 reduces chronic
inflammatory pain in mice. Neurol. Res. 43: 528–534.
https://doi.org/10.1080/01616412.2021.1884798

Wang Z., A. Emmerich, N. J. Pillon, T. Moore, D. Hemerich, *et al.*, 2022 Genome-wide
association analyses of physical activity and sedentary behavior provide insights
into underlying mechanisms and roles in disease prevention. Nat. Genet. 54:
1332–1344. https://doi.org/10.1038/s41588-022-01165-1

Waters R. P., R. B. Pringle, G. L. Forster, K. J. Renner, J. L. Malisch, *et al.*, 2013
Selection for increased voluntary wheel-running affects behavior and brain
monoamines in mice. Brain Res. 1508: 9–22.
https://doi.org/10.1016/j.brainres.2013.01.033

Wisløff U., S. M. Najjar, Ø. Ellingsen, P. M. Haram, S. Swoap, *et al.*, 2005 Cardiovascular Risk Factors Emerge After Artificial Selection for Low Aerobic Capacity. Science 307: 418–420. https://doi.org/10.1126/science.1108177

Wood A. R., The Electronic Medical Records and Genomics (eMERGE) Consortium, The MIGen Consortium, The PAGE Consortium, The LifeLines Cohort Study, *et al.*, 2014 Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46: 1173–1186. https://doi.org/10.1038/ng.3097

Xie Y., W. Pan, and A. B. Khodursky, 2005 A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. Bioinformatics 21: 4280–4288. https://doi.org/10.1093/bioinformatics/bti685

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

Yanaka N., Y. Imai, E. Kawai, H. Akatsuka, K. Wakimoto, *et al.*, 2003 Novel membrane protein containing glycerophosphodiester phosphodiesterase motif is transiently expressed during osteoblast differentiation. J. Biol. Chem. 278: 43595–43602. https://doi.org/10.1074/jbc.M302867200

Zamer W. E., and S. M. Scheiner, 2014 A conceptual framework for organismal biology: linking theories, models, and data. Integr. Comp. Biol. 54: 736–756. https://doi.org/10.1093/icb/icu075

Zinski A. L., S. Carrion, J. J. Michal, M. A. Gartstein, R. M. Quock, *et al.*, 2021 Genome-to-phenome research in rats: progress and perspectives. Int. J. Biol. Sci. 17: 119–133. https://doi.org/10.7150/ijbs.51628

CONCLUDING REMARKS

Locomotor behavior has important evolutionary and medical implications. The High Runner (HR) mouse selection experiment has demonstrated that the voluntary wheel-running behavior of outbred laboratory house mice responds strongly to selection, with the four replicate HR lines running 2.5- to 3-fold more revolutions per day than the four non-selected control lines when they reach selection limits (Careau *et al.* 2013). Such a strong response to selection demonstrates a strong genetic component to individual variation in the base population (Swallow *et al.* 1998). With respect to the specific genes that account for the genetic effects on wheel running, previous studies have mapped QTL and eQTL (Kelly *et al.* 2012, 2014) and identified genomic regions differentiated between the HR and control linetypes (Xu and Garland 2017). This dissertation continues this line of research by conducting analyses with more complete data sets and delving more deeply into the biological functions of the implicated regions and genes.

Chapter 1

Chapter 1 utilizes mixed model comparisons of whole-genome sequences (with nearly 6 million SNPs polymorphic across the 8 lines), along with haplotype data and nonstatistical tests, and focuses on regions consistently identified by all three differentiation tests. These 13 "consistent" regions include genes that are associated with many systems controlling running motivation and ability, both of which are known to differ between the HR and control lines (Rhodes *et al.* 2005; Rhodes and Kawecki 2009; Swallow *et al.* 2009; Garland, Jr. *et al.* 2011; Wallace and Garland, Jr. 2016). These

systems include neural differentiation, limb bud development, metabolism, and other biological functions relevant for physical activity. However, results in chapter 3 do raise concerns regarding the possibly coincidental nature of these intuitive systems (see below).

Additionally, multiple differentiated regions contained genes associated with the vomeronasal and non-vomeronasal olfactory systems. The vomeronasal system functions similarly to traditional olfactory systems in that it detects particles in the environment that are associated with a stimulus, such as food or other animals. The major distinction between vomeronasal and non-vomeronasal olfaction is that the latter identifies particles in the air, while the vomeronasal system requires physical contact (i.e., the sensors in the nose must make contact with a surface containing the particles). Chapters 2 and 3 find that genes related to these olfactory systems are among the most consistently reoccurring throughout the various analyses of this dissertation. The likely explanation is that a mouse's ability to smell other mice (either in the room with non-vomeronasal olfaction or on the wheel with vomeronasal olfaction) directly affects its motivation to run on the wheel. Dewan et al. (2019) found that mice ran less on wheels that had previously been used by a mouse, an effect that was enhanced when the previous mouse had been a male as opposed to a female. Given the social nature of mice, it is not a surprise that cues indicating the presence or absence of other mice in the area would alter a mouse's physical activity. Furthermore, HR and control mice have been shown to have differential genes expression in several sensory receptors in specific receptor clusters related to the vomeronasal system (Nguyen *et al.* 2020).

Chapter 2

Chapter 2 utilized the same genomic data as chapter 1 and the analyses performed (SNP, haplotype, and nonstatistical tests) were similar. The major difference here is that analyses were repeated dropping one line at a time, with a particular focus on line HR3, which has become fixed for the mini-muscle allele that has drastic effects on muscle mass, other muscle phenotypes, and a variety of other pleiotropic effects on traits that should affect exercise ability (Garland, Jr. *et al.* 2002; Guderley *et al.* 2006, 2008; Hannon *et al.* 2008; Bilodeau *et al.* 2009; Kelly *et al.* 2013). These analyses allowed us to test for differences in response to selection among the HR lines ("multiple solutions" sensu Garland, Jr. *et al.* 2011).

To identify a locus as differentiated, the control lines need to have the allele frequencies for a given locus either not change much or shift in a similar fashion (e.g., become fixed for the same allele) while simultaneously the HR lines need to respond to selection at this locus in a similar fashion to each other that is different from the controls (e.g., become fixed for the opposite allele as the controls). Thus, in the normal 4 vs 4 analyses, if any line's allele frequency deviates from the other lines within its linetype, that locus will not be detected as differentiated by our analyses (regardless of its relevance to the selection response). This makes the 4 vs 4 analyses inadequate for detecting differences in response to selection among the HR lines. Dropping one of the HR lines enables the detection of differentiation of the remaining 3 HR lines from the 4 control lines, where the 4th HR line had not followed suit.

These analyses were performed with the expectation that HR3 would share the least in common with the other HR lines. This expectation is based on HR3 being fixed for a recessive allele ($Myo4^{minimsc}$) of a gene of major effect that causes systemic change throughout several relevant systems (see Chapter 2 for more details), thus altering the genetic background of the line. As seen in the results of chapter 1 (Table 3), HR3 had more heterozygosity than any other line, potentially related to the presence of the $Myo4^{minimsc}$ allele. Dropping any line enabled the identification of new genomic regions of differentiation; however, as expected, dropping HR3 produced more such differentiated regions, implying that this line does diverge from the other HR lines more than they do each other. Admittedly, I cannot rule out the possibility that HR3 is uniquely different from the other HR lines simply because it has higher heterozygosity, but even if that were the case it is also possible that the presence of $Myo4^{minimsc}$ has caused the higher heterozygosity.

Interestingly, 3 of the 4 HR lines, when dropped, resulted in significant differentiation of a previously unidentified olfactory (non-vomeronasal) region. Moreover, each region identified when dropping an HR line was different from those identified when dropping a different HR line. This potentially implies that selection for a similar olfactory phenotype may have resulted in a genetic response at different loci (i.e., multiple solutions to selection on olfaction, specifically). These olfactory regions are in addition to identifying genes related to dopamine signaling, hippocampus morphology, heart size, body size, etc., that are new as compared with the results of chapter 1.

Chapter 3

Chapter 3 analyzed pooled sequence data from each of the 8 lines at generation 22, which

is when the HR lines were reaching selection limits (Careau *et al.* 2013). The expectation

was that we might detect more selection signatures at generation 22 than at 61, because

random genetic drift in the control lines would decrease the statistical ability to detect

differences between the two linetypes as successive generations passed. These pooled

sequence data amount to only 8 data points for each SNP locus, rather than 80, and so

could not be analyzed with the mixed model method used in the previous chapters. Thus,

we applied 3 different approaches comparing the 4 control lines to the 4 HR lines (1)

standard t-test assuming unequal variance, (2) regularized t-test described by Baldwin-

Brown et al. (2014) (see also Baldi and Long 2001), and (3) a windowed variant of the

regularized t-test (WRT) developed for this chapter. The WRT test was intended to

provide two benefits: (1) reduce Type I error rate caused by sampling error producing

low among-line variance (like the regularized t-test) and (2) allow for greater statistical

power for those genomic regions that have genuinely low among-line variance expected

to be produced by uniform selection (Burke 2012). In practice, the WRT test was

computationally highly intensive and produced results very similar to the regularized t-

test.

For comparison to generation 22 results, generation 61 was reanalyzed as pooled

sequence data. The comparison of these two generations showed that we had almost

completely divergent regions being identified in each. For example, analyses of

generation 22 identified 436 differentiated regions (FDR = 0.01), whereas analyses of

generation 61 identified 21 differentiated regions (FDR = 0.01), with little congruence

between them. This came as such a surprise that allele frequencies pertaining to

generation 22 were recalculated starting with the raw sequence (fastq) files to confirm

their accuracy. Simulations involving sampling error demonstrated that methodological

differences could explain some of the discrepancies between the two generations, but

were insufficient to explain how some regions of differentiation at generation 61

containing hundreds of differentiated SNPs could be virtually undetected at generation

22.

One hypothesis for a biological cause of these differences would be a

physiological constraint limiting the full potential each mouse's running behavior (this

was also described in chapter 2 as a possible cause of "multiple solutions"). Simulations

of selection over 61 generations comparing a model with no constraint on wheel running

with a model that contained a constraint showed that the constraint could reduce the

consistency in detected selection signatures between generations 22 and 61. However, in

these simulations the constraint did not alter consistency by nearly enough to explain the

differences observed in the real data. This result implies that other factors that were not

included in the model (e.g., linkage disequilibrium, dominance, epistasis, interactions,

sampling error) may be essential for identifying the cause of these differences. Attempts

were made to model dominance by treating every fifth locus as being completely

dominant for the pro-running allele, but the dominance limited our ability to observe a

realistic response to selection in wheel running levels, implying that its role in the

genetics of wheel running may be more complex than we had expected. Interestingly,

both the constrained and unconstrained models demonstrated that loci with large effect size on wheel running were much easier to detect as differentiated at earlier generations (proportionate to the effect size) than at later generations, when detection of loci with lower effect sizes becomes more common than earlier generations.

In chapters 1 and 2, determination of potential biological functions of implicated genes was done using a review of literature and knockout phenotypes. To better assess which knockout phenotypes were overrepresented among the genes identified in chapter 3, the number of genes associated with a given phenotype (e.g., cardiac development) were counted from among the implicated genes and this was compared to the number of genes associated with that same phenotype from the whole genome. The relative proportion of genes associated with the given phenotype among the differentiated genes was compared to the relative proportion of the genes associated with the given phenotype among all genes in the database using a chi-squared test. Ultimately, the only generation 22 phenotype to be overrepresented was associated with lung development, a trait that has been little studied in the HR mice (Meek *et al.* 2009; Kolb *et al.* 2010; Vaanholt *et al.* 2010; Dlugosz *et al.* 2013; Kelly *et al.* 2017; Garland Jr *et al.* unpublished manuscript; Schwartz *et al.* in review). Repeating this test on the generation 61 results (both pooled and chapter 1 results) did not produce any nominally overrepresented phenotypes among those tested (Supplemental Table S5 in Chapter 3). This implies that the intuitive phenotypes described in chapters 1 and 2 may have been coincidental. This means that such phenotypes (e.g., cardiac, neurological, dopamine) are not necessarily overrepresented in the regions identified in chapters 1 and 2. However, this does not

diminish the possible biological role that these genes may play in wheel running described in these chapters. Importantly, olfactory (vomeronasal and non-vomeronasal) genes were not analyzed in this way due to their underrepresentation in the knockout database. In addition, gene ontology tests in chapter 1 would imply that such olfaction genes are likely overrepresented, but this merits further analyses to confirm.

Gene ontology analyses performed in chapter 3 implicate genes related to reward pathways and glucose metabolism in prenatal development, as well as posttranscriptional and translational processing, as overrepresented functions among the genes identified. Olfactory genes are again largely omitted from these ontological analyses. However, one genomic region identified among the 13 consistent regions by the individual mouse analyses performed in chapter 1 was also identified among the most differentiated at generation 22 (chr14:51,198,229-53,776,455), and it contains both vomeronasal and non-vomeronasal olfactory receptor genes.

Suggestions for Future Research

In conclusion, the analyses of these three chapters indicate numerous potential genes influencing wheel-running behavior in the HR mice. Furthermore, we see evidence that the genes favored by selection can vary from line to line, and across generations. These findings indicate that the genetic response to selection in complex traits is itself complex, as has been suggested for other traits and systems (Burke *et al.* 2010; Lillie *et al.* 2019; Crawford *et al.* 2020). Among the genes identified (see Chapter 3) are several with influence over systems that are intuitively involved in aerobic exercise and have been

demonstrated to be differentiated between the HR and control lines (e.g., reward pathways involving dopamine precursors and GABAnergic neurons). Additionally, these results suggest organs and systems that may benefit from further exploration (e.g., lungs).

The olfactory system is perhaps the most consistently implicated across all three chapters (see also Dewan *et al.* 2019; Nguyen *et al.* 2020) and could use further study in the HR mice. With so much variation from generation 22 to 61, analyses of more generations of the HR mouse genome could yield additional understanding of the genomic architecture of wheel running. Although individual mouse genotypes allow for more powerful analyses (Xu and Garland 2017), simulations involving sampling error in chapter 3 indicate that pooled sequencing is potentially sufficient (though increased read depth is valuable for reducing sampling error). Furthermore, loci with the greatest effect size on wheel running may be best detected before the selection limit is achieved (~generations 10-15). Additionally, future studies should include functional analyses of the genes most likely to have been favored by selection. For example, *Dach1* , a gene associated with eye and leg development (Mardon *et al.* 1994), is identified at generation 61 in a differentiated region with no other genes present. Simulations in chapter 3 might implicate that *Dach1* may not be a gene with a large effect size (as many selection signatures at generation 61 will have small effect sizes), but it is unlikely to be a hitchhiker for another gene.

Given the importance of physical activity for both animals in the wild and biomedicine, further exploration of its genetic underpinnings would be of great value to science and society (Manley 1996; Dickinson *et al.* 2000; Lightfoot *et al.* 2018).

REFERENCES

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

Baldwin-Brown J. G., A. D. Long, and K. R. Thornton, 2014 The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. Mol. Biol. Evol. 31: 1040–1055. https://doi.org/10.1093/molbev/msu048

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

Burke M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 467: 587–590. https://doi.org/10.1038/nature09352

Burke M. K., 2012 How does adaptation sweep through the genome? Insights from long-term selection experiments. Proc. R. Soc. B Biol. Sci. 279: 5029–5038. https://doi.org/10.1098/rspb.2012.0799

Careau V., M. E. Wolak, P. A. Carter, and T. Garland, Jr., 2013 Limits to behavioral evolution: the quantative genetics of a complex trait under directional selection. Evolution 67: 3102–3119. https://doi.org/10.1111/evo.12200

Crawford D. L., P. M. Schulte, A. Whitehead, and M. Oleksiak, 2020 Evolutionary physiology and genomics in the highly adaptable killifish (*Fundulus heteroclitus*). Compr. Physiol. 10. https://doi.org/DOI:10.1002/cphy.c190004

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

Dickinson M. H., C. T. Farley, R. J. Full, M. A. R. Koehl, R. Kram, *et al.*, 2000 How animals move: an integrative view. Science 288: 100–106. https://doi.org/10.1126/science.288.5463.100

Dlugosz E. M., H. Schutz, T. H. Meek, W. Acosta, C. J. Downs, *et al.*, 2013 Immune response to a Trichinella spiralis infection in house mice from lines selectively bred for high voluntary wheel running. J. Exp. Biol. 216: 4212–4221. https://doi.org/10.1242/jeb.087361

Garland, Jr. T., M. T. Morgan, J. G. Swallow, J. S. Rhodes, I. Girard, *et al.*, 2002 Evolution of a small-muscle polymorphism in lines of house mice selected for high activity levels. Evolution 56: 1267–1275. https://doi.org/10.1111/j.0014-3820.2002.tb01437.x

Garland, Jr. T., S. A. Kelly, J. L. Malisch, E. M. Kolb, R. M. Hannon, *et al.*, 2011 How to run far: multiple solutions and sex-specific responses to selective breeding for high voluntary activity levels. Proc. R. Soc. B Biol. Sci. 278: 574–581. https://doi.org/10.1098/rspb.2010.1584

Garland Jr T., B. Wallau, I. Girard, J. S. Rhodes, and S. F. Perry, unpublished manuscript Experimental evolution and phenotypic plasticity of the lung and maximal oxygen consumption in high-activity house mice. (g22b_Lung_VO2max_1_Perry_11-06.doc)

Guderley H., P. Houle-Leroy, G. M. Diffee, D. M. Camp, and T. Garland, 2006 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. 144: 271–282. https://doi.org/10.1016/j.cbpb.2006.02.009

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

Hannon R. M., S. A. Kelly, K. M. Middleton, E. M. Kolb, D. Pomp, *et al.*, 2008 Phenotypic effects of the "Mini-Muscle" allele in a large HR x C57BL/6J mouse backcross. J. Hered. 99: 349–354. https://doi.org/10.1093/jhered/esn011

Kelly S. A., D. L. Nehrenberg, K. Hua, T. Garland, Jr., and D. Pomp, 2012 Functional genomic architecture of predisposition to voluntary exercise in mice: expression QTL in the brain. Genetics 191: 643–654. https://doi.org/10.1534/genetics.112.140509

Kelly S. A., T. A. Bell, S. R. Selitsky, R. J. Buus, K. Hua, *et al.*, 2013 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 195: 1385–1395. https://doi.org/10.1534/genetics.113.154476

Kelly S. A., D. L. Nehrenberg, K. Hua, T. Garland, Jr., and D. Pomp, 2014 Quantitative genomics of voluntary exercise in mice: transcriptional analysis and mapping of expression QTL in muscle. Physiol. Genomics 46: 593–601. https://doi.org/10.1152/physiolgenomics.00023.2014

Kelly S. A., F. R. Gomes, E. M. Kolb, J. L. Malisch, and T. Garland, Jr., 2017 Effects of activity, genetic selection and their interaction on muscle metabolic capacities and organ masses in mice. J. Exp. Biol. 220: 1038–1047. https://doi.org/10.1242/jeb.148759

Kolb E. M., S. A. Kelly, K. M. Middleton, L. S. Sermsakdi, M. A. Chappell, *et al.*, 2010 Erythropoietin elevates VO2,max but not voluntary wheel running in mice. J. Exp. Biol. 213: 510–519. https://doi.org/10.1242/jeb.029074

Lightfoot J. T., E. J. C. De Geus, F. W. Booth, M. S. Bray, M. Den Hoed, *et al.*, 2018 Biological/genetic regulation of physical activity level: Consensus from GenBioPAC. Med. Sci. Sports Exerc. 50: 863–873. https://doi.org/10.1249/MSS.0000000000001499

Lillie M., C. F. Honaker, P. B. Siegel, and Ö. Carlborg, 2019 Bidirectional selection for body weight on standing genetic variation in a chicken model. Genes|Genomes|Genetics g3.400038.2019. https://doi.org/10.1534/g3.119.400038

Manley A. F., 1996 *Physical activity and health: a report of the Surgeon General*. DIANE Publishing.

Mardon G., N. M. Solomon, and G. M. Rubin, 1994 *dachshund* encodes a nuclear protein required for normal eye and leg development in *Drosophila*. Development 120: 3473–3486.

Meek T. H., B. P. Lonquich, R. M. Hannon, and T. Garland, Jr., 2009 Endurance capacity of mice selectively bred for high voluntary wheel running. J. Exp. Biol. 212: 2908–2917. https://doi.org/10.1242/jeb.028886

Nguyen Q. A. T., D. Hillis, S. Katada, T. Harris, C. Pontrello, *et al.*, 2020 Coadaptation of the chemosensory system with voluntary exercise behavior in mice, (H. Matsunami, Ed.). PLOS ONE 15: e0241758. https://doi.org/10.1371/journal.pone.0241758

Rhodes J. S., S. C. Gammie, and T. Garland Jr, 2005 Neurobiology of mice selected for high voluntary wheel-running activity. Integr. Comp. Biol. 45: 438–455. https://doi.org/10.1093/icb/45.3.438

Rhodes J. S., and T. J. Kawecki, 2009 Behavior and Neurobiology, pp. 263–300 in *Experimental evolution: concepts, methods, and applications of selection experiments*, University of California Press, Berkley.

Schwartz N. E., M. P. McNamara, J. M. Orozco, J. O. Rashid, A. P. Thai, *et al.*, in review A test of the aerobic capacity model for vertebrate energetics: selective breeding

for high voluntary exercise in mice increases maximal (VO2max), but not basal metabolic rate. J. Exp. Biol.

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

Swallow J. G., J. P. Hayes, P. Koteja, and T. Garland, 2009 Selection experiments and experimental evolution of performance and physiology, pp. 301–351 in *Experimental evolution: concepts, methods, and applications of selection experiments*, University of California Press, Berkeley.

Vaanholt L. M., S. Daan, T. Garland Jr., and G. H. Visser, 2010 Exercising for life? Energy metabolism, body composition, and longevity in mice exercising at different intensities. Physiol. Biochem. Zool. 83: 239–251. https://doi.org/10.1086/648434

Wallace I. J., and T. Garland, Jr., 2016 Mobility as an emergent property of biological organization: Insights from experimental evolution: Mobility and biological organization. Evol. Anthropol. Issues News Rev. 25: 98–104. https://doi.org/10.1002/evan.21481

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