Initial Molecular-Level Response to Artificial Selection for Increased Aerobic Metabolism Occurs Primarily through Changes in Gene Expression

Mateusz Konczal,*[1] Wiesław Babik,[1] Jacek Radwan,[2] Edyta T. Sadowska,[1] and Paweł Koteja[1]

[1]Institute of Environmental Sciences, Jagiellonian University, Kraków, Poland

[2]Institute of Environmental Biology, Faculty of Biology, Adam Mickiewicz University, Poznań, Poland

***Corresponding author:** E-mail: mateusz.konczal@uj.edu.pl.

Associate editor: Michael Nachman

Abstract

Experimental evolution combined with genome or transcriptome resequencing (Evolve and Resequence) represents a promising approach for advancing our understanding of the genetic basis of adaptation. Here, we applied this strategy to investigate the effect of selection on a complex trait in lines derived from a natural population of a small mammal. We analyzed the liver and heart transcriptomes of bank voles (*Myodes* [=*Clethrionomys*] *glareolus*) that had been selected for increased aerobic metabolism. The organs were sampled from 13th generation voles; at that point, the voles from four replicate selected lines had 48% higher maximum rates of oxygen consumption than those from four control lines. At the molecular level, the response to selection was primarily observed in gene expression: Over 300 genes were found to be differentially expressed between the selected and control lines and the transcriptome-wide pattern of expression distinguished selected lines from controls. No evidence for selection-driven changes of allele frequencies at coding sites was found: No single nucleotide polymorphism (SNP) changed frequency more than expected under drift alone and frequency changes aggregated over all SNPs did not separate selected and control lines. Nevertheless, among genes which showed highest differentiation in allele frequencies between selected and control lines we identified, using information about gene functions and the biology of the selected phenotype, plausible targets of selection; these genes, together with those identified in expression analysis, have been prioritized for further studies. Because our selection lines were derived from a natural population, the amount and the spectrum of variation available for selection probably closely approximated that typically found in populations of small mammals. Therefore, our results are relevant to the understanding of the molecular basis of complex adaptations occurring in natural vertebrate populations.

Key words: **selection experiment, mammals, bank vole, standing genetic variation, RNA-Seq, maximum metabolic rate.**

Introduction

One of the central goals of evolutionary biology is to understand the genetic mechanisms by which organisms evolve new, adaptive phenotypes under natural selection and thus diverge phenotypically (Stapley et al. 2010; Butlin et al. 2012). Despite decades of research, detecting and deciphering the molecular changes underlying adaptation remain challenging tasks to which researchers have applied various approaches, such as study of candidate genes, genome-wide scans for positive selection or experimental evolution (Sabeti et al. 2007; Garland and Rose 2009; Stapley et al. 2010; Barrett and Hoekstra 2011; Fournier-Level et al. 2011). Recently, however, the Evolve and Resequence (E&R) approach has been gaining popularity. E&R studies provide better control over confounding factors than other approaches and allow investigators to choose the traits under selection (Turner et al. 2011; Kawecki et al. 2012). This approach involves genetic analyses of populations of organisms that are either adapted to specific, experimentally controlled ambient conditions, or that are selected for increased performance with respect to a specific behavioral, life-history, or morphophysiological trait. Such studies have helped to answer questions concerning adaptation (Tenaillon et al. 2012; Soria-Carrasco et al. 2014), the importance of new mutations (Burke et al. 2010), and the genomic patterns of a recent response to selection (Johansson et al. 2010; Pettersson et al. 2013).

The sources of adaptive variation appear to vary among evolutionary lineages. For example, extensive work on microorganisms has contributed to our understanding of adaptation scenarios that are driven by selection acting on new mutations: A substantial number of de novo mutations are expected during the course of experiments in such organisms as a result of the large population sizes involved (Herring et al. 2006; Barrick et al. 2009; Tenaillon et al. 2012). However, in multicellular, sexually reproducing species (the subject of this study), standing genetic variation is the main source of variation at the initial stages of adaptive evolution (Barrett and Schluter 2008; Burke et al. 2010). The E&R approach has been used to comprehensively and successfully investigate some traits in *Drosophila melanogaster* (Teotónio et al. 2009; Burke et al. 2010; Turner et al. 2011); for example, Burke et al. (2010) studied flies that had been selected for accelerated development for 600 generations. They concluded that the probability of fixation of selected variants is

relatively low and that selection does not readily expunge genetic variation (Burke et al. 2010). Subsequent studies using fruit flies have confirmed the complex evolutionary trajectories of selected variants and emphasized the importance of epistatic interactions (Huang et al. 2012; Orozco-Terwengel et al. 2012).

The genetic response to selection at the early stages of adaptation is less well understood in vertebrates, which usually have smaller population sizes (Johansson et al. 2010; Chan et al. 2012; Pettersson et al. 2013). Interestingly, however, the results from the few experiments performed thus far contrast with those obtained from *D. melanogaster*. For instance, in lines of chickens that had been selected for high and low body mass, Johansson et al. (2010) observed many genetic regions with fixed differences; likewise, signals of classical, hard selective sweeps were detected in a mouse line that had been selected for high body mass (Chan et al. 2012). In these studies, the high number of regions detected that were presumably under the influence of divergent selection (50 in chickens and 67 in mice) suggests that the initial phase of selection substantially increases divergence between lines while simultaneously reducing polymorphism within them. However, these results may reflect the differences in experimental setup rather than true contrast between vertebrates and *Drosophila*. Both vertebrate experiments utilized crossed inbred lines as a base population. In such cases long haplotype blocks are present at the beginning of the experiment and they may be fixed rapidly in small experimental populations, mimicking the effects of hard sweeps. Such situations are less likely in nature, where at the early stage of adaptation standing genetic variation is subject to selection. To understand the basis of adaptive processes occurring in the wild, it is therefore crucial to conduct selection experiments that control for the effect of genetic drift and utilize lines derived from natural populations. Although selection experiments on nonmodel organisms were always possible to perform, in practice they were rarely undertaken, partly because until recently uncovering molecular genetic mechanisms of the evolution in nonmodel organisms was often not possible. This has changed with the advent of high throughput sequencing (Schlötterer et al. 2014).

In this study, we used high-throughput transcriptome sequencing to test whether recent, intense selection acting over multiple generations in mammalian populations would result in repeatable changes in the frequencies of variants in protein-coding genes and/or patterns of gene expression. This study was performed using an experimental evolution model system, with four lines of bank voles (*Myodes* [=*Clethrionomys*] *glareolus*), selectively bred for high swim-induced aerobic metabolism (A lines) and four unselected control lines (C lines; Sadowska et al. 2008). The experiment has been established as a tool for testing hypotheses concerning correlated evolution of aerobic locomotor performance and basal metabolic rates, which is believed to have been a crucial element in evolution of terrestrial vertebrates (literature cited in Sadowska et al. 2005, 2008). Thus, the model is likely to illuminate many ecophysiological questions concerning physiological genomics and the evolution of endothermy (Nespolo et al. 2011; Pérusse et al. 2013). The swim-induced maximum rate of oxygen consumption differed significantly between the selected and control lines already in generation 2 (Sadowska et al. 2008), and in generation 13 it was 48% higher in A-line voles than in C-line voles (mean ± SD: 5.32 ± 0.64 ml O_2/min vs. 3.59 ± 0.57 ml O_2/min, respectively; Chrzascik et al. 2014; Stawski et al. 2015; see also supplementary material S1.3, Supplementary Material online). Voles from the A lines (also referred to as "selected" lines) differed significantly from control voles not only in the trait directly under selection but also in their basal metabolic rate and a number of other behavioral and morphophysiological traits (supplementary material S1.3, Supplementary Material online). This experiment presented a unique opportunity to study the genetic basis of the response to selection in mammals thanks to a combination of several factors: 1) Selection could operate on the natural genetic variation directly derived from a wild population, 2) known pedigrees allowed for the exact calculation of drift expectations, 3) the trait under selection was complex and ecologically important, and 4) the replicated lines provided an appropriate system to study the role of drift in phenotypic and genetic differentiation.

The eight lines (four selected and four control) were sequenced using a pooled RNA-Seq approach (Konczal et al. 2014). We used transcriptome analysis as a convenient way to determine whether the response to selection at the molecular level was dominated by gene expression or structural changes. King and Wilson (1975) proposed that adaptive evolutionary change is largely due to changes in gene expression, and there is empirical evidence from genetic mapping and interspecies comparisons that both support (Wray 2007; Jones et al. 2012) and contradict this view (Hoekstra and Coyne 2007). A recent study of patterns of polymorphism and divergence in murid rodents suggested that most of adaptive changes appear in regulatory regions. On the other hand, wider regions of reduced diversity around exons than around conserved noncoding elements may be interpreted as a result of substantially larger effects of adaptive substitutions (Halligan et al. 2013). However, it is unclear whether rapid adaptation from standing genetic variation produces similar patterns. To address the question of the relative importance of coding mutations versus changes in expression levels, we studied the transcriptomes of two organs: The heart, which plays a crucial role in an organism's aerobic capacity (Bye et al. 2008), and the liver, which, as a central metabolic organ (Malarkey et al. 2005; Konczal et al. 2014), was a promising target for investigations of the molecular mechanisms that were responsible for the increased basal metabolism observed in selected lines. The scale of the project limited the possibility of detecting significant responses to selection in allele frequencies in coding regions or gene expression only to loci of large effects. However, we could still infer the role of many loci of small effect if selection changes allele frequencies in coding regions or gene expression of many genes in replicable way (across the four selected and four control lines). In such case, the aggregate effect of these changes should result in multidimensional differentiation of selected lines from

controls (Turchin et al. 2012), although covariances of allele frequencies, resulting from between-population component of LD (Linkage Disequilibrium; Ohta 1982), may weaken this effect (Storz 2005; Le Corre and Kremer 2012).

We identified over 300 differentially expressed genes that are associated with diverse molecular functions; many of these functions appeared to be highly relevant to the phenotypic response to selection for increased aerobic metabolism. This result, combined with significant clustering of genome-wide transcriptional profiles, highlights the role of rapid changes in gene expression at the early stages of adaptive evolution. In contrast, allele frequency changes in coding sequences appear to play, at best, a minor role: The differences observed in the allele frequencies between the selected and control lines could be entirely explained by drift and the aggregate effect of allele frequency changes does not separate selected lines from controls. Nevertheless, among the genes that showed the highest differentiation in allele frequencies, we identified, on the basis of their molecular function, a set of candidates, which may possibly contribute to phenotypic changes between the selected and control lines. These genes should be prioritized as a target for future research.

Results

Single Nucleotide Polymorphisms

From each sample, an average of 37.1 (±10.6 SD) million 1 × 100 bp reads were obtained; of these, 75.9% were uniquely mapped to the bank vole liver and heart reference transcriptomes (table 1). After several steps of data filtering (see Materials and Methods), we identified 172,246 single nucleotide polymorphisms (SNPs). The vast majority of identified variants were found in putative protein-coding genes, with an average of 3.95 and 3.48 SNPs per kilobase in open reading frames (ORFs) and untranslated regions (UTRs) of SNP-containing contigs (table 2).

To estimate effective population sizes, the mean inbreeding coefficient (F) was calculated from pedigree for each of the four selected (A) and four control (C) lines in each generation. The degree of inbreeding increased slightly faster in the selected lines, probably reflecting a subtle difference in the breeding scheme between the selected and control lines (see Materials and Methods). The mean effective population size (N_e) was about 16.4% lower in the selected than in the control lines (56.1 vs. 67.1; $P = 0.06$, t-test; fig. 1A). To evaluate the effect of differences in N_e between lines on the amount of genetic variation, we examined the allele frequency spectra (fig. 1B). Specifically, we calculated for each line the number of such SNPs which were polymorphic in the entire data set but showed little or no variation (minor allele frequency, MAF < 0.05) within the line. An analysis of covariance (ANCOVA) was used to examine how well N_e (covariate) and treatment (selected vs. control lines) explained the number of such SNPs. We found a significant effect of N_e ($F(1,5) = 6.92$, $P = 0.047$), but no effect of treatment ($F(1, 5) = 0.14$, $P = 0.72$; fig. 1D).

For each SNP, F_{ST} values were calculated between all pairs of lines. Mean pairwise F_{ST} distances did not reveal any clustering of selected or control lines ($F(1, 6) = 0.97$, $P = 1.00$, randomization test; fig. 1C); and variation among selected lines (calculated as a mean distance to centroids) was slightly, but nonsignificantly higher than that between control lines ($F(1, 6) = 1.17$, $P = 0.32$; analysis of variance [ANOVA]). The two control lines (C1, C3) with the largest effective population sizes were least distant from each other, suggesting that drift played the dominant role in the differentiation of allele frequencies among lines.

Additionally, a principal components analysis was performed to look for correlated changes in allele frequencies in various subsets of SNPs; such changes could reflect the response of multiple genes to the same selection pressure. None of the eight PCs clearly differentiated between selected and control lines (supplementary fig. S2.1, Supplementary Material online).

In the next step, folded allele frequency spectra were compared both between selection regimes and with expectations generated from simulations of genetic drift over the course of the experiment. Forward simulations were performed using known pedigrees; for the initial allele frequency spectrum, these simulations used the average spectrum calculated from control lines.

The allele frequency spectra were less skewed in the simulated data than in the observed data (fig. 1B), which could have been caused by two effects: Bias in the estimation of the initial allele frequency spectrum or selection against slightly deleterious alleles. We assessed the overall effect of deleterious alleles by comparing the allele frequency spectra of synonymous and nonsynonymous sites. Minor allele frequencies were lower for nonsynonymous SNPs than for synonymous SNPs (synonymous median MAF = 0.091, nonsynonymous median MAF = 0.068; $P = 10^{-16}$; KS (Kolmogorov Smirnov) tests), indicating the presence of purifying selection. For

Table 1. Overview of the Assembly of Bank Vole Transcriptomes.

	Liver	Heart
No. of genes	146,758	252,281
No. of genes > 1 kb	23,512	24,825
N50 gene length (bp)	1,225	650
No. of genes within N50	19,101	47,439
No. of genes with likely CDS	18,050	11,110
N50 of genes with likely CDS	3,296	3,081
No. of bases (Mb)	103.1	134.9

Note.—N50, 50% of the assembly length is in genes of the length of N50 bp or longer; genes, TGMs contain both coding and noncoding sequences; genes with likely CDS, genes containing successfully annotated ORFs.

Table 2. Overview of SNPs Used for Analyses.

No. of SNPs	172,246
No. of genes with SNPs	15,043
No. of nonsynonymous SNPs	22,963
No. of synonymous SNPs	44,844
No. of UTR-located SNPs	71,657
No. of SNPs in noncoding genes	32,782

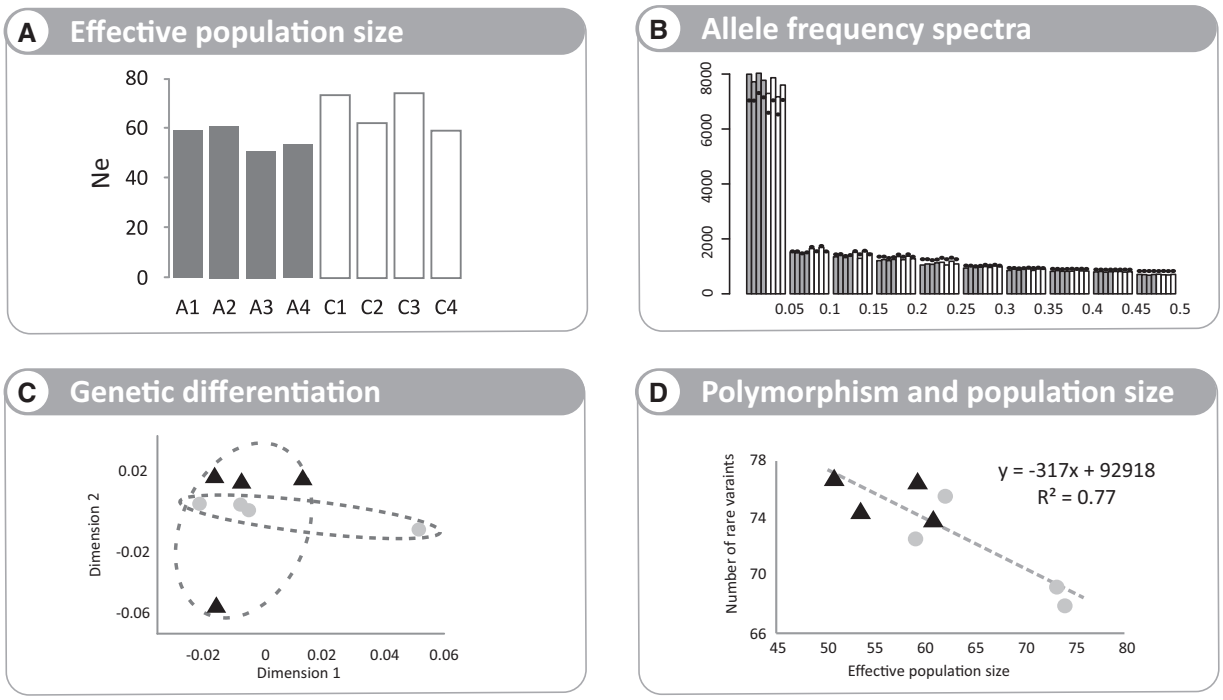

Fig. 1. Effect of selection and population size on allele frequency changes in the bank vole selection experiment. (*A*) Effective population sizes (N_e) of selected (gray) and control (white) lines, calculated from pedigrees for 13 generations of the selection experiment. (*B*) Folded allele frequency spectra for selected (gray) and control (white) lines, compared with pedigree-based simulations (dots—expectations from simulations). (*C*) Multidimensional scaling plot (MDS) of genetic distances (pairwise F_{ST}) between selected and control lines. Triangles represent selected lines; circles represent controls. The MDS plot drawn using mean pairwise F_{ST} values calculated for all SNPs. (*D*) Regression of the number of SNPs with rare variants (MAF < 0.05) on effective population size. Number of rare variants in thousands.

synonymous sites, the difference in the percentage of rare variants between simulated and observed sites was 2.7%; in contrast, the difference was 7.4% for nonsynonymous sites.

If the same alleles contribute to the response to selection in all lines, SNP frequencies in the selected lines should diverge from those in the control lines to a greater degree than expected under neutrality. To investigate whether such effect occurred, we identified variants that had ranges of allele frequencies nonoverlapping between the selected and the control lines (3,233 [1.88%] SNPs in 1,873 genes). The number of SNPs with nonoverlapping allele frequencies was significantly lower than expected from drift simulations ($P = 0.01$; randomization test), but this effect was not significant for subsets of synonymous (1.97%, $P = 0.33$, randomization test) or nonsynonymous SNPs (1.96%, $P = 0.78$, randomization test). For each of these sites, the minimum allele frequency difference between the 16 possible A–C comparisons (diffStat) was used as a composite statistic (Turner et al. 2011). The distribution of diffStat values did not differ between the data and drift simulations, and we did not observe overrepresentation of high diffStat values (supplementary fig. S2.2, Supplementary Material online).

The relatively small population sizes decrease the population recombination rates, which may cause entire long haplotypes to drift. To control for the effect of linkage within genes, we used the following procedure. First, we generated 1,000 data sets consisting of SNPs sampled randomly one per gene. Then, for each data set the number of SNPs with the ranges of allele frequencies nonoverlapping between selection and control lines was calculated. Finally, we recorded the proportion of data sets in which the number of SNPs with nonoverlapping allele frequencies was higher than expected under drift (upper 10% of the distribution from simulations). None of the data sets fell into the upper 10% of the distribution, and the relative number of differentiated SNPs was slightly lower than expected from simulations ($P < 10^{-50}$; *t*-test). In coding regions this effect was mostly explained by nonsynonymous sites ($P < 10^{-50}$, *t*-test), whereas the fraction of synonymous SNPs with nonoverlapping allele frequencies closely followed drift expectations ($P = 0.15$; *t*-test; supplementary fig. S2.3, Supplementary Material online).

To get some insight about the power to detect variants under selection in our experiment, we performed pedigree-based simulations of selection. These simulations were used to estimate the probability of obtaining nonoverlapping allele frequencies between the selected and control lines, depending on the strength of selection and initial allele frequency. With increasing selective advantage the probability of obtaining nonoverlapping frequencies increased considerably ($s = 0.05–5.5\%$; $s = 0.2–41.6\%$, averaged over the range of initial allele frequencies; supplementary fig. S2.4, Supplementary Material online). The probability of obtaining nonoverlapping frequencies after 13 generations was highest when the favored allele initially segregated at an intermediate frequency (initial frequency 0.05–5.8%; 0.5–29.9%; 0.9–4.0%, averaged over the range of selection coefficients). This probably reflects the fact that rare positively selected variants will often be lost due to

drift in some of selected lines. Similarly, selected variants at high initial frequencies will often become fixed in at least some control lines.

Genes that harbored differentiated SNPs (diffStat > 0) had a higher density of polymorphisms ($P < 10^{-6}$, randomization test) which in turn exhibited more equal allele frequencies ($P = 10^{-12}$; KS test). This effect was present for nonsynonymous SNPs ($P = 1.6 \times 10^{-5}$, randomization test), but not for synonymous ($P = 0.38$, randomization test), what may mean either that highly polymorphic genes are more likely to be targets of selection or that they are more likely to differentiate by drift because of their effective neutrality.

To explore whether some of the genes with differentiated nonsynonymous SNPs (supplementary table S1, Supplementary Material online) were somehow associated with phenotypic differences between selected and control lines, we investigated their functions using relevant databases, and the most intriguing cases are described in Discussion.

Overall, these results did not provide evidence that selection for increased maximum metabolic rate caused allele frequency changes at coding SNPs. The changes in allele frequencies that we did observe can be explained by the actions of two other evolutionary forces, namely drift and purifying selection, that acted in the same way in all lines.

Gene Expression

To determine differences in expression levels between the selected and control lines, we investigated all expressed genes with at least ten mapped reads and performed ordination of the lines using a multidimensional scaling analysis that was based on estimates of pairwise similarity in expression levels. In contrast with the SNP results, this analysis found that the selected lines and control lines clustered separately for the liver samples; for heart samples clustering was not significant (liver: $P = 0.002$; heart: $P = 0.384$; randomization tests; fig. 2*A* and *B*). Thus, it appears that similar changes in gene expression in the most important metabolic organ, the liver, might have occurred in all selected lines, distinguishing them from controls.

In the heart samples, 79 genes were differentially expressed between selected and control lines (52 genes were overexpressed and 20 were underexpressed in selected lines; false discovery rate [FDR] = 0.05; fig. 2*C*). Many more genes were differentially expressed in the liver (278 genes at FDR = 0.05; 123 genes were overexpressed and 155 were underexpressed in selected lines; fig. 2*D*). We annotated 110 differentially expressed genes (28 in heart and 82 in liver), all putatively protein coding (supplementary tables S2 and S3, Supplementary Material online). As an additional assessment of these differentially expressed genes we performed *t*-test on FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values and calculated the proportion of genes with FPKM values nonoverlapping between the selected and control lines. Sixty-three per cent of coding genes differentially expressed in liver and 61% in heart showed statistically significant result of the *t*-test ($P < 0.05$) and 55% and 63% of them, respectively, had nonoverlapping expression level values. The molecular functions of some of these 110 genes are discussed below.

Discussion

Differentiation at the Molecular Level

Using replicate selected and control lines derived from a natural population of a small mammal, we experimentally quantified the responses to artificial selection at the molecular level. The trait investigated here, maximum metabolic rate during exercise, likely has a complex genetic basis (Hagberg et al. 2011; Roth et al. 2012; Pérusse et al. 2013; Wolfarth et al. 2014) and as the genomic basis of evolutionary change in complex traits is still poorly understood (Rockman 2012), our results are of interest from a broad evolutionary perspective. This experiment thus addresses a question of general evolutionary and physiological relevance by applying the strict criteria associated with the design of E&R studies (replicates and use of control populations). In doing so, it has provided insight into the genetic patterns of adaptation that arise from standing genetic variation in populations of mammals. Here, we clearly observed that the effects of artificial selection were visible at both the phenotypic and molecular levels. However, although the artificial selection applied in this experiment resulted in reproducible changes in the expression levels of many genes, it did not cause appreciable changes in allele frequencies at coding SNPs, which were instead influenced predominately by drift.

The importance and contribution of expression changes and coding mutations to adaptation has long been a topic of great interest (King and Wilson 1975; Hoekstra and Coyne 2007; Wray 2007; Stern and Orgogozo 2008; Fraser 2013). Recent findings in human populations suggest that adaptation in regulatory elements, likely affecting gene expression, is ten times more frequent than in protein-coding parts of the genome (Fraser 2013). Similar evidence has been obtained from diverse taxa: Since the split between marsupials and placental mammals, many more new regulatory elements than coding exons have emerged to differentiate the two groups (Mikkelsen et al. 2007), and in the evolution of rodents, most adaptive mutations have occurred in regulatory elements rather than in protein-coding exons (Halligan et al. 2013). Our results are consistent with these findings in showing that, during a relatively short period of selection in small populations, the pattern of expression of multiple genes can change rapidly and in a reproducible manner. Our two main observations—that more than 300 genes were differentially expressed between selected and control lines and that changes in allele frequencies were caused predominantly by drift—support the hypothesis that changes in gene expression, rather than changes in allele frequencies of coding regions, play a central role in adaptation. Other genetic analyses of rodent selection experiments found hundreds of differentially expressed genes by either eQTL (expression Quantitative Trait Loci) investigations (Kelly et al. 2012, 2014) or by comparing expression profiles between treatments (Bye et al. 2008; Roberts et al. 2013). These observations suggest that expression analyses may be among the most

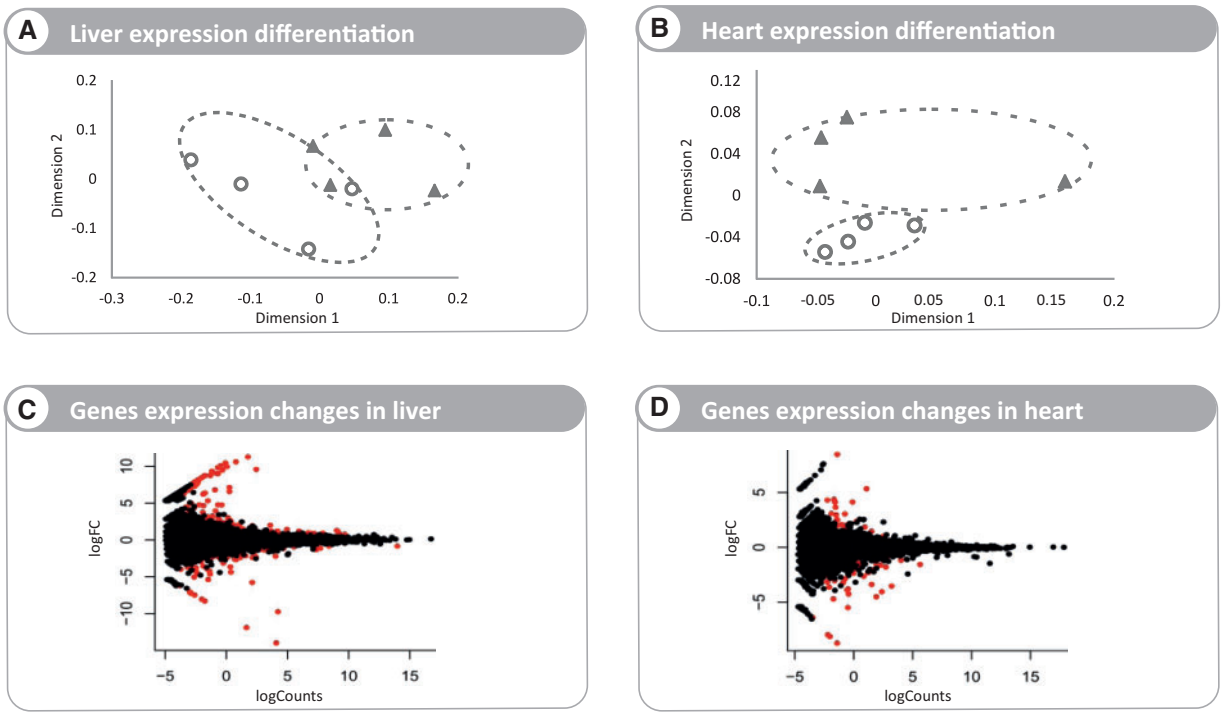

Fig. 2. Effect of selection on expression changes in the bank vole selection experiment. (*A*, *B*) Multidimensional scaling plot (MDS) of transcriptome distances (in terms of the BCV) between selected and control lines of the bank vole experiment. Triangles represent selected lines; circles represent controls. The MDS plots were drawn using the expression level values of 91,760 liver and 108,656 heart genes. (*C*, *D*) Log-fold-change expression versus log abundance of gene expression in liver and heart samples. Gene expression data are TMM-normalized. Genes that qualified as significantly differentially expressed (FDR 0.05) are in red.

promising strategies to identify the molecular basis of phenotypic differences.

The contrast between expression and frequency changes of coding alleles may be explained by selection acting mostly on alleles in regulatory elements. However other factors cannot be ignored here. Gene expression can be thought of as a first-order phenotype. Because many different SNPs, possibly located in many different genes may affect expression level of a particular gene it may be easier to observe the effects of selection on gene expression levels than on the frequency changes of individual alleles. Thus significant and reproducible changes in expression levels may result from the combined effects of a number of subtle (and not necessary repeatable among selected lines) allele frequency changes in regulatory elements. For many genes expression is essentially a polygenic trait and as such might be less prone to drift. Therefore random variation between lines in gene expression might be lower than in SNPs, increasing statistical power to detect subtle changes. On the other hand, if the artificial selection in this experiment had resulted in subtle allele frequency changes in many coding SNPs, then this pattern should have been detected by the multidimensional scaling analysis: The selected lines would have formed clusters separated from the controls. However, we did not observe such clustering for the allele frequency data; it was only seen in the gene expression data. This suggests that the changes in gene expression, but not repeatable changes in coding variants, underlie the observed response to selection.

Multidimensional scaling and additional analyses showed that the differences in allele frequencies in coding regions between selected and control lines were driven mainly by genetic drift. These results do not necessarily mean that selection does not affect variation in coding regions; however, they are not compatible with a scenario in which widespread positive selection on coding genes shapes the genomic patterns of polymorphism within, and divergence between, selection regimes. The fact that we did not find SNPs differentiated more than expected under drift effectively rules out the possibility that any genetic variant of large effect was repeatedly selected for. Therefore, we suspect that if positive selection affected the coding sequences in our experiment, it acted on a limited number of variants that provided a small-to-moderate fitness advantage. In this respect, our study contrasts with other selection experiments performed on vertebrates which showed large genomic regions being fixed for alternative variants between treatments (Johansson et al. 2010; Chan et al. 2012; Pettersson et al. 2013). This difference may be a consequence of differences in the genetic architecture of the traits investigated or result from differences in experimental setup, in particular the origin and genetic makeup (e.g., presence of linkage disequilibria) of the base population. However, many previous studies relied on observations of reduced genetic diversity as evidence for the effects of selection and, in doing so, may have suffered from the confounding effects of genetic drift. For example, Johansson et al. (2010) selected chicken lines for high body mass and

interpreted the decrease in heterozygosity as reflecting the operation of selection. However, the high-body-mass selected line had an effective population size that was reduced by around 10% (44.5 vs. 49.3). Our experimental design allowed us to ascribe reduced polymorphism in the selected lines to their reduced effective population sizes, and we were able to show that even such minor differences can significantly affect polymorphism.

Our study differed from many other E&R studies (especially these performed on vertebrates) in the nature of the standing genetic variation available at the onset of the experiment. We directly utilized genetic variation that was segregating in a natural population. This has not been the case in many other experiments, in which source populations were created by crossing inbred or isofemale lines (Johansson et al. 2010; Chan et al. 2012; Orozco-Terwengel et al. 2012; Turner and Miller 2012). As a result, experimental populations may not have adequately reflected the standing genetic variation available for positive selection in natural populations. For example, inbred lines are likely to have been cleaned of large-effect recessive deleterious mutations but to have fixed many slightly deleterious ones. When inbred lines are crossed, slightly deleterious mutations become common, and the initial allele frequency spectrum is expected to depart from that observed in nature. In nature, most deleterious variants are rare and the shape of allele frequency spectrum depends on effective population size. Here, we inferred that negative selection is an important force that might shape allele frequencies, even in populations of small N_e.

However, a small N_e is a limitation inherent to E&R studies in vertebrates, and because of this, our study had limited power to detect the effects of selection on SNPs. Pedigree-based selection simulations demonstrated that, due to the effect of drift in relatively small experimental populations, only strongly selected ($s \sim 0.2$) variants segregating at appreciable frequencies in the base population can be detected with high probability. Therefore, the effective size of experimental populations has critical consequences for the E&R approach. Several theoretical studies have examined the effect of population size on analyses of artificial selection, and all of them have found that N_e is a crucial factor that influences the power of such analyses (Baldwin-Brown et al. 2015; Kofler and Schlötterer 2015; Kessner D, Novembre J, unpublished data; http://dx.doi.org/10.1101/005892, last accessed August 15, 2014). Specifically, Baldwin-Brown et al. (2015) argue that, to localize causative SNPs with at least 80% success, researchers should use a population size of 1,000 diploid individuals. This is obviously not feasible for most laboratory experiments involving vertebrates, and therefore only variants with large effects can be detected with high probability (Baldwin-Brown et al. 2015). The same situation is however observed in nature—many vertebrate populations are small, having effective population sizes comparable to those reported here, which makes distinguishing effects of drift and selection a challenging task (Palstra and Ruzzante 2008). Additionally, population recombination rate is low in small populations, which increases the rate of false positives because drift affects entire long haplotypes and leads to correlated allele frequency changes in multiple SNPs. We partially controlled for the effect of linkage by sampling one SNP per gene. However, this problem needs to be considered in future E&R studies.

An alternative explanation for the lack of considerable changes in allele frequencies is that adaptation is due to different variants in different lines. Repeatability of adaptation is however surprisingly high on the gene level both in experimental evolution experiments and in natural populations (Conte et al. 2012; Tenaillon et al. 2012; Martin and Orgogozo 2013). Because of that observation and the fact that initial standing genetic variation was similar in all selected lines derived from a single base population, many SNPs initially in moderate frequencies should be repetitively selected. Because the number of repetitively selected coding SNPs was probably modest, and the power to detect them was limited, we attempted to identify potential candidates by exploring the molecular functions of differentiated genes. We also carried out a similar analysis on genes with significantly different expression levels. This strategy is often used in experimental selection surveys (Bye et al. 2008; Kelly et al. 2012, 2014; Roberts et al. 2013) allowing to pinpoint the most promising candidates for future investigations. Below, we very briefly discuss the molecular processes associated with these plausible candidates.

Molecular Function of Plausible Candidates

To assess which biological pathways have possibly changed in response to selection, we investigated the genes that had been identified as having nonoverlapping allele frequencies between selected and control lines (despite the overall lack of support for a role of selection in allele frequency changes, some variants may nevertheless be weakly selected for) and those that were differentially expressed in at least one organ. We refer to these genes as "plausible candidates" and list them in supplementary tables S1–S3, Supplementary Material online. None GO (Gene Ontology) category was significantly overrepresented relative to all GO categories (FDR < 0.05). We argue, though, that some of these plausible candidates are more likely than others to explain some phenotypic changes. We narrowed down the list of candidates based on their functions and present the most interesting genes below.

The stromal interaction molecule 1 (*STIM1*) gene showed highest differences in allele frequency between the selected and control lines, that is, harbored nonsynonymous SNPs with the highest diffStat values. STIM1 senses exhaustion of Ca^{2+} in the endoplasmic reticulum and activates an ion channel in the plasma membrane, causing continuous influx of the extracellular Ca^{2+} (Kurosaki and Baba 2010). Heterozygous mutations in human *STIM1* cause tubular aggregate myopathy (Bohm et al. 2013) and sotormorken syndrome (Misceo et al. 2014). In tubular aggregate myopathy all patients were characterized by mild and slowly progressive lower limb muscle weakness causing frequent falls and running difficulties (Bohm et al. 2013), which suggests that mutations in *STIM1* may play an important role in swimming performance.

Another gene of great interest is that of glycogen phosphorylase (*PYGL*). The physiological role of this liver phosphorylase is to ensure constant supply of glucose for extrahepatic tissues by catalyzing the rate-limiting step in glycogenolysis (Newgard et al. 1989; Bollen et al. 1998). Nonsynonymous mutations in human *PYGL* cause glycogen storage disease type VI. In substantial number of patients with such disease mild hypotonia, delayed motor development and muscle weakness and cramps were observed (Beauchamp et al. 2007). Interestingly, another nonsynonymous SNP that showed significant differences between selected and control lines was located in the gene that encodes the glycogen-debranching enzyme AGL, which acts together with PYGL to mobilize glucose from glycogen reserves. Mutations in human *AGL* cause Glycogen Storage Disease type III affecting calves and peroneal muscles (Lucchiari et al. 2007).

The gene characterized by the highest number of differentiated nonsynonymous SNPs was *MYO18B*, encoding unconventional myosin XVIIIb. Previous studies have demonstrated the important role of this gene in myocardic structures (Ajima et al. 2008), as well as its contribution to cognitive phenotypes (Purcell et al. 2009; Ludwig et al. 2013).

Another interesting gene is insulin-like growth factor 2 (*IGF2*) its expression increases in response to endurance training and extent of this change differs between humans with highest and lowest improvement in aerobic capacity (Keller et al. 2011). Next gene with an interesting function is the one that encodes fibroblast growth factor 21, which stimulates glucose uptake in adipocytes and plays a critical role in the regulation of lipid homeostasis (Badman et al. 2007). We also identified changes in other genes involved in lipid metabolism (e.g., *ABCG1*, *CYP17A*, *APOB*, *LIPA*, *APOA1*, *APOA2*, *CYP4A14*), the formation and proper functioning of the heart (e.g., *XIRP2*, *KDM4A*, *JPH2*), and stress responses (e.g., *IRGM*, *DELE*, *PARP*, *HSP70*, *HSP105*). All these genes may be involved in response to selection for aerobic performance.

In liver tissue, we found significant differences in expression of the gene that encodes retinoblastoma-like protein-2 (RBL2). RBL2 acts as a transcriptional repressor of the enzymes DNMT3A and DNMT3B, which catalyze the transfer of methyl groups to specific CpG structures in DNA, a process called DNA methylation (Benetti et al. 2008). Also on the list of differentially expressed liver genes are the genes encoding heterogeneous nuclear ribonucleoprotein H2 (HNRPH2), which plays an important role in pre-mRNA processing (Alkan et al. 2006), and methyl-CpG-binding domain protein 4 (MBD4), which takes part in the active demethylation process (Roloff et al. 2003). Additionally, one of the genes whose allele frequencies differed the most between selected and control lines was that coding for lysine-specific demethylase 4A (KDM4A), which plays a central role in modifying the "histone code" (Tan et al. 2011). Taken together, these observations suggest that genes associated with epigenetic changes might represent important targets of selection.

One of the most significant changes in expression level was observed for the gene that encodes aphrodisin—a protein that transports pheromones that stimulate copulatory behavior (Briand et al. 2004; Stopková et al. 2010). Genes coding for aphrodisin-like proteins in bank voles may be used in chemical communication among individuals and thus may play an important role in aggression, dominance, and mate choice (Stopková et al. 2010). Changes in expression of this gene are interesting in the context of differences in reproductive success between the selected and control lines. Already in previous generations of the selection experiment we observed that voles from the selected lines produced litters sooner after the mating (Koteja et al. 2010). Also, in generation 12 and 13 (parents and siblings, respectively, of the voles used in transcriptome analysis), the proportion of mated pairs that produced offspring was significantly higher in the selected than in the control lines (generation 12—selected: 93.2%, control: 70.1%, $P = 0.011$; generation 13—selected: 92.9%, control: 68.2%, $P = 0.010$; GLIMMIX procedure in SAS 9.3). It is tempting to speculate that changes in the expression of aphrodisin may have been the underlying mechanism.

Conclusions

We characterized, through transcriptome sequencing, the response to selection for increased aerobic metabolism in lines derived from a natural population of the bank vole. We showed that the initial response to selection occurs mainly through changes in gene expression. After applying a rigorous control for the effect of drift, no repeatable changes in allele frequencies at coding SNPs could be unambiguously attributed to directional selection. These results differ from a handful of previous analyses of selection experiments in birds and mammals, in which signals of multiple selective sweeps were detected by resequencing of genomes. Because our selection lines were derived from a natural population, the amount and spectrum of variation available for selection probably closely approximates these typically found in populations of small mammals. Therefore our results are relevant to the understanding of the molecular basis of complex adaptations occurring in vertebrate populations. By combining transcriptome analyses, information about gene functions, and knowledge about selected traits and phenotypes, we identified genes and pathways that could be the targets of selection for increased aerobic metabolism. To further investigate the patterns uncovered here, novel methods that combine knowledge from both population genetics and molecular biology should be developed and exploited in order to effectively characterize the candidate genes that were identified during this experiment.

Materials and Methods

Selection Experiment

This study was performed using individuals from the 13th generation of a laboratory colony of bank voles (*Myodes* [=*Clethrionomys*] *glareolus*) that was subjected to selection for improved aerobic metabolism. The rationale for the selection experiment as well as detailed breeding and selection protocols are described elsewhere (Sadowska et al. 2008; supplementary materials S1.1 and S1.2, Supplementary

Material online). Briefly, the colony was founded using approximately 320 voles captured in 2000 and 2001 in the Niepołomice Forest in southern Poland. For 6–7 generations, the animals were bred randomly, and the colony was used for quantitative genetic analyses of metabolic rates (Sadowska et al. 2005). In 2004, a multidirectional selection experiment was established. In the A-lines analyzed here, the selection criterion was the maximum mass-independent (residual from regression) 1-min rate of oxygen consumption achieved during 18 min of swimming. The swim test was conducted at 38°C so that no thermoregulatory burden was imposed, and animals were tested when they were around 75–85 days old (see supplementary materials S1.1 and S1.2, Supplementary Material online, for details of the protocol and results of the selection).

Estimating Inbreeding Effective Population Size

To explore breeding differences among lines, individual inbreeding coefficients were calculated for each line using the R package "pedigree." Changes in inbreeding over time were calculated as

$$\Delta F_i = \frac{F_i - F_{i-1}}{1 - F_{i-1}},$$

where F_i is the mean inbreeding coefficient in generation i (Falconer and Mackay 1996). The effective population size N_e was calculated for each line according to the formula:

$$N_e = \frac{1}{2x\Delta F},$$

where ΔF is the mean change in inbreeding over time.

Sampling, RNA Extraction, and Sequencing

For the transcriptome analysis, five males and five females of 75–80 days in age were sampled from each line; each individual came from a different family. These individuals had not been previously used in the swimming trials or for any other specific measurements (except routine measurements of body mass). Voles were euthanized by being placed one by one in a jar containing isoflurane (Aerane) fumes; this process took place between 8.00 AM and 2.00 PM. The animals were then weighed, and a small part of their left liver lobes and hearts were immediately excised and placed in RNAlater (Sigma). Samples were stored overnight at 4°C and then frozen at −20°C.

Total RNA was extracted using RNAzol (Molecular Research Center) in accordance with the manufacturer's instructions. RNA concentration and quality were measured using Nanodrop and Agilent 2100 Bioanalyzer, respectively. All samples had an RNA Integrity Number higher than 7.0 and were thus suitable for poly-A selection and cDNA library preparation.

For each organ, we prepared one pooled sample per line—using equal amounts of total RNA from each individual—for a total of 16 samples. Residual DNA was removed from pooled samples using a DNA-free Kit (Ambion). RNA quality and concentration following the DNAse treatment were assessed as described above.

Poly-A selection, reverse transcription, and the preparation of barcoded cDNA libraries with the TrueSeq RNA kit were performed by the Georgia Genomics Facility, USA. Liver samples from one control line (C3) were pair-end (2 × 100 bp) sequenced on an Illumina HiSeq 2000 and used for reference transcriptome construction (Konczal et al. 2014). For the remaining 15 pools, single-end (1 × 100 bp) sequencing was performed. The reads were deposited in Sequence Read Archive (Bioproject PRJNA267038).

Reference Transcriptome Reconstruction and Annotation

We first trimmed low-quality reads using DynamicTrim, removed adaptors with Cutadapt, and removed reads shorter than 20 bp with LengthSort (Cox et al. 2010; Grabherr et al. 2011; Martin 2011). As references, we used a previously assembled liver transcriptome (Konczal et al. 2014) and the heart reference transcriptome generated by Kaczyńska, Konczal, Babik and Niedziałkowska (unpublished data), which had been assembled for other purposes. The transcriptomes were processed by merging transcripts that were likely derived from the same genomic locations. This produced transcriptome-based gene models (TGMs), which we refer to here as "genes" (Stuglik et al. 2014).

We did not assemble one transcriptome from pooled reads of two organs to avoid the problem of potential redundancy of the reference transcriptome. Transcriptome complexity negatively affects de novo assembly and TGMs reconstruction (Vijay et al. 2013), and it is known that most alternative splicing occurs between organs (Wang et al. 2008). Redundancy of the reference transcriptome has serious implications for SNP calling, because reads mapping equally well to multiple locations are filtered out during this procedure. Because we assembled transcriptomes of each organ separately, this problem is reduced to within-organ splice variant variation and even if it occurs for one transcriptome, SNPs may be still identified using reference transcriptome from the other organ (see below).

TGMs were annotated using Trinotate software. Trinotate makes use of a number of methods for functional annotation (e.g., homology search to Swissprot database, protein domain identification, protein signal prediction) of likely coding regions (likely CDSs). Likely CDSs were identified using a pipeline implemented in Trinity, but this approach did not successfully annotate all of them. Nonannotated genes represent either errors, fast evolving genes, or genes whose homologs are not present in the Swissprot database.

Mapping and Identification of SNPs

Filtered reads were mapped to the reference transcriptomes using Bowtie2 (Langmead and Salzberg 2012). For liver samples of line C3, we subsampled reads to obtain a comparable number of single-end sequences. Reads mapped into multiple locations were removed from analyses. We mapped all reads from both organs together to the liver and heart

transcriptomes to increase accuracy of allele frequency estimation.

SNP calling was performed in two steps. First, we identified SNPs with samtools (mpileup with options: -Q 10, -E), which is dedicated to diploid genomes (Li et al. 2009). SNPs that contained more than two variants in samtools output were discarded. In the second step, we applied PoPoolation2 (Kofler et al. 2011) to filter data and estimate allele frequencies. Only SNPs with a minimum of 10× coverage in each sample and a minimum of three reads that supported minor allele were considered. Additional to multiallelic SNPs removal, two procedures were applied to identify and exclude similar paralogs that had been assembled into single genes: 1) We removed most polymorphic genes (more than five SNPs per 100 bp using a minimum of 10× coverage) and 2) we discarded all genes that contained SNPs with an excess of observed heterozygotes or had BLASTN hits with E value $< 10^{-150}$ to such genes. This procedure was based on those developed for the individually sequenced liver transcriptomes of 10 voles (Konczal et al. 2014) and the heart transcriptomes of 20 voles (Kaczyńska et al., unpublished data)—these studies excluded SNPs for which more than 8/10 or 14/20 samples, respectively, were heterozygotes. Using custom python scripts, SNPs were classified as being synonymous, nonsynonymous, UTRs, or localized in putative nonprotein-coding genes.

The above procedure was performed for both liver and heart transcriptomes and yielded highly overlapping sets of SNPs. The differences resulted from differences in reference transcriptomes (lack of an SNP in one transcriptome may be caused by incompletely assembled or nonassembled genes or by splice variants which were not collapsed into a single gene model). To remove the redundancy in the SNP data set, we preformed the following procedure. First, we clustered liver and heart SNPs-containing genes using reciprocal BLAST searches (BLASTN hits with E value $< 10^{-100}$ and $> 99\%$ identity). Genes which did not form clusters were apparently expressed in one organ only and were retained (liver: 5,786 genes, heart: 1,759 genes). Genes with significant hits in the other transcriptome were reduced using the criterion of completeness. From clusters with one-to-one relation (containing a single sequence from each transcriptome; 6,082 clusters) we retained the longer one. The relation one to many (939 clusters) was mainly caused by fragmentation of the gene in one of the assemblies; therefore, we retained SNPs from the transcriptome with the single assembled sequence. For clusters containing greater than one sequence from liver and greater than one from heart (many to many, 200 clusters), we included in analyses sequences from this transcriptome in which the total length of sequences was larger. SNPs identified in thus selected genes were used for all analyses. For genes that contained at least one SNP, F_{ST} was calculated using PoPoolation2. F_{ST} was calculated for each SNP using the formula $F_{ST} = (\pi_T - \pi_W)/\pi_T$. Mean F_{ST} values between each pair of lines were calculated, and this matrix of pairwise F_{ST} was used to test 1) whether the extent of variation among lines within treatments differed between selected and control lines and 2) whether selected and control lines cluster separately. Multivariate homogeneity of group dispersion was tested using betadisper function from vegan package, (Oksanen et al. 2013) followed by an ANOVA. To test for separate clustering of selected and control lines, we calculated the ratio of between treatment to within treatment variance using adonis function (vegan package) and assessed its statistical significance through 1,000 randomizations. Randomized matrices of mean F_{ST} were obtained by shuffling pairwise F_{ST} values for each gene independently. The original mean pairwise F_{ST} matrix was visualized using nonmetric multidimensional scaling.

Simulations of Allele Frequency Distribution under Drift and Positive Selection

To obtain the allele frequency distributions that would be expected under drift, we performed forward drift simulations on known pedigrees. Simulations were performed separately for allele frequency spectra derived from all, synonymous and nonsynonymous SNPs.

The simulations were divided into four parts and were repeated 10 million times (steps 2–4):

1. *Estimation of the initial allele frequency distribution*. As we did not know the allele frequencies in the ancestral population, we had to estimate them using data from the control lines. For each SNP, we calculated the mean allele frequency from the four control lines. If control lines diverge mainly due to drift (a reasonable assumption for most polymorphisms), such averages are unbiased estimates of allele frequencies in the ancestral population, which may then be used to reconstruct the allele frequency spectrum in the ancestral population.
2. *Simulation of the genotypes of "generation 0" individuals*. We simulated the genotype of each individual in the ancestral population by randomly choosing one initial allele frequency (p_0) from the set of frequencies estimated in step 1. Then, for each individual, we sampled from a binomial distribution with $n = 2$ and $P = p_0$ (n, number of draws; P, probability of success), thus obtaining the number of allele copies (0, 1, 2) for each individual.
3. *Simulation of the effect of drift on known pedigrees*. Based on known pedigrees, we simulated genotypes for each individual by randomly choosing one chromosome from each of the parents. We then obtained genotypes for ten individuals (that were selected for sequencing for each line) and calculated allele frequencies.
4. *Simulation of pooling and sequencing error*. Pooling and sequencing cause inaccuracy in allele frequency estimation. Therefore, we decided to add relevant variation to the simulated allele frequencies using the relative errors of allele frequency estimation that had been previously calculated (Konczal et al. 2014). For a given MAF class, a gamma function was fitted to the distribution of experimentally obtained relative errors. Then, one value of estimation error was randomly chosen from the fitted gamma distribution and incorporated into the simulation results.

To assess power to detect selected variant given its selective advantage and initial frequency, we used the approach similar as in drift simulations. Several initial allele frequencies ($f = 0.05, 0.1, 0.25, 0.5, 0.75, 0.9$) and four different values of selection advantage ($s = 0, 0.05, 0.1, 0.2$) were used. In the course of pedigree based simulations, in selectively bred lines advantageous allele was passed from heterozygote parents to offspring with higher probability ($0.5 + 1/2\ s$) than the alternative variant ($0.5 - 1/2\ s$). For each combination of f and s, we performed 100,000 iterations and recorded the fraction iterations with diffStat > 0.

Scripts used for simulations are available at http://www.molecol.eko.uj.edu.pl (last accessed March 9, 2014).

SNP Analyses, Polymorphism, and Divergence between Lines/Selection Regimes

To study the effect of differences in N_e between lines on the amount of genetic variation, we examined the allele frequency spectra. Specifically we calculated for each line the number of such SNPs which were polymorphic in the entire data set but showed little or no variation (MAF < 0.05) within the line. We used an ANCOVA in which N_e was a covariate and treatment (control vs. selected) was a fixed effect; the interaction between the two was included in order to check the assumption that the slopes were homogeneous between treatments. The interaction was not significant ($F(1, 4) = 0.84$, $P = 0.41$). As a consequence, we used a simple model without interactions to study the general effect of N_e and treatment on allele frequency spectra.

To study differentiation between the selected and control lines, we investigated repeatable changes in allele frequencies. SNPs with frequencies that were either always higher or always lower in selected lines as compared with control lines (nonoverlapping allele frequencies) were considered to be potential targets of selection (plausible candidates). For each such site, we calculated the diffStat statistic, which is the smallest difference in allele frequency between selected and control lines (Turner et al. 2011). The distribution of the number of unlinked candidate SNPs was estimated by sampling one SNP per gene 1,000 times. We then sampled the same number of SNPs (the number of genes with at least one SNP) from simulated pedigrees 600 times; in these simulations, drift was the only evolutionary force in operation. We subsequently compared the two sets of results. The difference between these two distributions should reveal the genome-wide effects of selection. These analyses were performed on the set of all SNPs, as well as separately for each class of SNPs (synonymous, nonsynonymous, UTR, noncoding).

The biological functions and molecular processes associated with the differentiated genes were studied using custom scripts and Gowinda software (Kofler and Schlötterer 2012).

Estimation and Comparison of Gene Expression Levels

To identify differentially expressed genes, we mapped reads onto reference transcriptomes with bowtie and used the EdgeR Bioconductor and RSEM packages (Robinson et al. 2010). The matrix of expected counts over all samples was used for EdgeR analyses. Only genes for which the sum of expected counts over all samples was higher than 10 were counted. Using the standard EdgeR procedure, we normalized counts for library size and RNA composition. We performed multidimensional scaling (biological coefficient of variation [BCV] method, EdgeR package) over all genes to analyze general expression patterns within tissues. We also estimated dispersion and calculated exact tests for genes that were differentially expressed between control and selected lines. The FDR was calculated as per Benjamini and Hochberg (1995).

The GO terms associated with the differentially expressed genes were investigated with GOrilla software (Eden et al. 2009).

To statistically test for separate clustering of transcriptional profiles of selected and control lines, we developed a procedure analogous to that used for the F_{ST} matrix. We used table of expression values (FPKM, TMM normalized) which included only transcripts with the total FPKM > 1. For this table, we calculated distance matrix (dist() function) and the ratio of between treatment to within treatment variance (adonis function, vegan package). The statistical significance of this ratio was assessed through 1,000 randomizations. Randomized matrices of mean gene expression distances were obtained by shuffling expression values of individual gene between lines. Differences between lines in genome-wide transcriptional profiles were visualized with multidimensional scaling (plotMDS function, edgeR).

Supplementary Material

Supplementary materials S1 and S2, figures S1.1, S2.1–2.4, and tables S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was funded by the Polish Ministry of Science (N N303 816740 to P.K.) and Jagiellonian University (DS 757 and DS 762). The authors are grateful to the many technicians and students, especially to K. Baliga-Klimczyk, K. Chrzascik, G. Dheyongera, A. Rudolf, and C. Stawski, for their contribution to animal maintenance, selection protocols and help with dissections.

References

Ajima R, Akazawa H, Kodama M, Takeshita F, Otsuka A, Kohno T, Komuro I, Ochiya T, Yokota J. 2008. Deficiency of Myo18B in mice results in embryonic lethality with cardiac myofibrillar aberrations. *Genes Cells* 13:987–999.

Alkan S, Martincic K, Milcarek C. 2006. The hnRNPs F and H2 bind to similar sequences to influence gene expression. *Biochem J.* 393: 361–371.

Badman MK, Pissios P, Kennedy AR, Koukos G, Flier JS, Maratos-Flier E. 2007. Hepatic fibroblast growth factor 21 is regulated by PPARα and is a key mediator of hepatic lipid metabolism in ketotic states. *Cell Metab.* 5:426–437.

Baldwin-Brown JG, Long AD, Thornton KR. 2015. The power to detect quantitative trait loci using resequenced, experimentally evolved populations of diploid, sexual organisms. *Mol Biol Evol.* 31(4):1040–1055.

Barrett RDH, Hoekstra HE. 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nat Rev Genet.* 12:767–780.

Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. *Trends Ecol Evol.* 23:38–44.

Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461:1243–1247.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological).* 289–300.

Beauchamp NJ, Taybert J, Champion MP, Layet V, Heinz-Erian P, Dalton A, Tanner MS, Pronicka E, Sharrard MJ. 2007. High frequency of missense mutations in glycogen storage disease type VI. *J Inherit Metab Dis.* 30(5):722–734.

Benetti R, Gonzalo S, Jaco I, Muñoz P, Gonzalez S, Schoeftner S, Murchison E, Andl T, Chen T, Klatt P. 2008. A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat Struct Mol Biol.* 15:268–279.

Bohm J, Chevessier F, Maues De Paula A, Koch C, Attarian S, Feger C, Hantai D, Laforet P, Ghorab K, Vallat J.-M, et al. 2013. Constitutive activation of the calcium sensor STIM1 causes tubular-aggregate myopathy. *Am J Hum Genet.* 92:271–278.

Bollen M, Keppens S, Stalmans W. 1998. Specific features of glycogen metabolism in the liver. *Biochem J.* 336:19–31.

Briand L, Trotier D, Pernollet J-C. 2004. Aphrodisin, an aphrodisiac lipocalin secreted in hamster vaginal secretions. *Peptides* 25: 1545–1552.

Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, Long AD. 2010. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* 467:587–U111.

Butlin R, Debelle A, Kerth C, Snook RR, Beukeboom LW, Castillo Cajas RF, Diao W, Maan ME, Paolucci S, Weissing FJ, et al. 2012. What do we need to know about speciation. *Trends Ecol Evol.* 27:27–39.

Bye A, Langaas M, Høydal MA, Kemi OJ, Heinrich G, Koch LG, Britton SL, Najjar SM, Ellingsen.Ø, Wisløff U. 2008. Aerobic capacity-dependent differences in cardiac gene expression. *Physiol Genomics.* 33: 100–109.

Chan YF, Jones FC, McConnell E, Bryk J, Bünger L, Tautz D. 2012. Parallel selection mapping using artificially selected mice reveals body weight control loci. *Curr Biol.* 22:794–800.

Chrzascik KM, Sadowska ET, Rudolf A, Koteja P. 2014. Learning ability in bank voles selected for high aerobic metabolism, predatory behaviour and herbivorous capability. *Physiol Behav.* 135: 143–151.

Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences* 279(1749): 5039–5047.

Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11:485.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10:48.

Falconer DS, Mackay TFC 1996. Introduction to quantitative genetics. Harlow (United Kingdom): Addison Wesley Longman.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM. 2011. A map of local adaptation in *Arabidopsis thaliana*. *Science* 334:86–89.

Fraser HB. 2013. Gene expression drives local adaptation in humans. *Genome Res.* 23:1089–1096.

Garland T Jr, Rose MR, editors. 2009. Experimental evolution: concepts, methods, and applications of selection experiments. Berkeley (CA): University of California Press.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29:644–652.

Hagberg JM, Rankinen T, Loos RJF, Pirusse L, Roth SM, Wolfarth B, Bouchard C. 2011. Advances in exercise, fitness, and performance genomics in 2010. *Med Sci Sports Exerc.* 43:743–752.

Halligan DL, Kousathanas A, Ness RW, Harr B, Eöry L, Keane TM, Adams DJ, Keightley PD. 2013. Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. *PLoS Genet.* 9:e1003995.

Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, et al. 2006. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet.* 38: 1406–1412.

Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61:995–1016.

Huang W, Richards S, Carbone MA, Zhu D, Anholt RRH, Ayroles JF, Duncan L, Jordan KW, Lawrence F, Magwire MM, et al. 2012. Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci U S A.* 109:15553–15559.

Johansson AM, Pettersson ME, Siegel PB, Carlborg.Ö. 2010. Genome-wide effects of long-term divergent selection. *PLoS Genet.* 6: e1001188.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484: 55–61.

Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. 2012. Experimental evolution. *Trends Ecol Evol.* 27:547–560.

Keller P, Vollaard NB, Gustafsson T, Gallagher IJ, Sundberg CJ, Rankinen T, Britton SL, Bouchard C, Koch LG, Timmons JA. 2011. A transcriptional map of the impact of endurance exercise training on skeletal muscle phenotype. *J Appl Physiol.* 110(1):46–59.

Kelly SA, Nehrenberg DL, Hua K, Garland T, Pomp D. 2012. Functional genomic architecture of predisposition to voluntary exercise in mice: expression QTL in the brain. *Genetics* 181((2):643–654.

Kelly SA, Nehrenberg DL, Hua K, Garland T, Pomp D. 2014. Quantitative genomics of voluntary exercise in mice: transcriptional analysis and mapping of expression QTL in muscle. *Physiol Genomics.* 46(16):593–601.

King M-C, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–116.

Kofler R, Pandey RV, Schlötterer C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27:3435–3436.

Kofler R, Schlötterer C. 2012. Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics* 28: 2084–2085.

Kofler R, Schlötterer C. 2015. A guide for the design of evolve and resequencing studies. *Mol Biol Evol.* 31(2):474–483.

Konczal M, Koteja P, Stuglik MT, Radwan J, Babik W. 2014. Accuracy of allele frequency estimation using pooled RNA-Seq. *Mol Ecol Resour.* 14:381–392.

Koteja P, Chrzascik KM, Sadowska ET, Ołdakowski Ł. 2010. Correlated responses to a multi-directional selection in bank voles (Myodes glareolus): reproductive parameters. Annual Main Meeting of the Society for Experimental Biology, Prague, June 30–July 3 Programme and Abstract Book: 98.

Kurosaki T, Baba Y. 2010. Ca2+ signaling and STIM1. *Prog Biophys Mol Biol.* 103(1):51–58.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9:357–359.

Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci under local adaptation. *Mol Ecol.* 21(7):1548–1566.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Lucchiari S, Santoro D, Pagliarani S, Comi GP. 2007. Clinical, biochemical and genetic features of glycogen debranching enzyme deficiency. *Acta Myol.* 26(1):72.

Ludwig K, Sämann P, Alexander M, Becker J, Bruder J, Moll K, Spieler D, Czisch M, Warnke A, Docherty S. 2013. A common variant in Myosin-18B contributes to mathematical abilities in children with dyslexia and intraparietal sulcus variability in adults. *Transl Psychiatry*. 3:e229.

Malarkey DE, Johnson K, Ryan L, Boorman G, Maronpot RR. 2005. New insights into functional aspects of liver morphology. *Toxicol Pathol.* 33:27–34.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17:10–12.

Martin A, Orgogozo V. 2013. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* 67(5):1235–1250.

Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A. 2007. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* 447:167–177.

Misceo D, Holmgren A, Louch WE, Holme PA, Mizobuchi M, Morales RJ, De Paula AM, Stray-Pedersen A, Lyle R, Dalhus B, et al. 2014. A dominant STIM1 mutation causes Stormorken syndrome. *Hum Mutat.* 35:556–564.

Nespolo RF, Bacigalupe LD, Figueroa CC, Koteja P, Opazo JC. 2011. Using new tools to solve an old problem: the evolution of endothermy in vertebrates. *Trends Ecol Evol.* 26:414–423.

Newgard CB, Hwang PK, Fletterick RJ. 1989. The family of glycogen phosphorylases: structure and function. *Crit Rev Biochem Mol Biol.* 24:69–99.

Ohta T. 1982. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceedings of the National Academy of Sciences* 79(6):1940–1944.

Oksanen JF, Blanchet G, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2013. vegan: Community Ecology Package. R package version 2.0-10. Available from: http://CRAN.R-project.org/package=vegan .

Orozco-Terwengel P, Kapun M, Nolte V, Kofler R, Flatt T, Schlätterer C. 2012. Adaptation of *Drosophila* to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Mol Ecol.* 21:4931–4941.

Palstra FP, Ruzzante DE. 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol Ecol.* 17: 3428–3447.

Pérusse L, Rankinen T, Hagberg JM, Loos RJF, Roth SM, Sarzynski MA, Wolfarth B, Bouchard C. 2013. Advances in exercise, fitness, and performance genomics in 2012. *Med Sci Sports Exerc.* 45:824–831.

Pettersson ME, Johansson AM, Siegel PB, Carlborg.Ö. 2013. Dynamics of adaptive alleles in divergently selected body weight lines of chickens. *G3* 3:2305–2312.

Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P, Ruderfer DM, McQuillin A, Morris DW. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460:748–752.

Roberts MD, Brown JD, Oberle LP, Heese AJ, Toedebusch RG, Wells KD, Cruthirds CL, Knouse JA, Ferreira JA, Childs TE, 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(11):R1024–R1035.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* 26:139–140.

Rockman MV. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* 66:1–17.

Roloff TC, Ropers HH, Nuber UA. 2003. Comparative study of methyl-CpG-binding domain proteins. *BMC Genomics* 4:1.

Roth SM, Rankinen T, Hagberg JM, Loos RJF, Pérusse L, Sarzynski MA, Wolfarth B, Bouchard C. 2012. Advances in exercise, fitness, and performance genomics in 2011. *Med Sci Sports Exerc.* 44:809–817.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918.

Sadowska ET, Baliga-Klimczyk K, Chrzascik KM, Koteja P. 2008. Laboratory model of adaptive radiation: a selection experiment in the bank vole. *Physiol Biochem Zool.* 81:627–640.

Sadowska ET, Labocha MK, Baliga K, Stanisz A, Wróblewska AK, Jagusiak W, Koteja P. 2005. Genetic correlations between basal and maximum metabolic rates in a wild rodent: consequences for evolution of endothermy. *Evolution* 59:672–681.

Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU. 2014. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity doi: 10.1038/hdy.2014.86.

Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* 344:738–742.

Stapley J, Reger J, Feulner PGD, Smadja C, Galindo J, Ekblom R, Bennison C, Ball AD, Beckerman AP, Slate J. 2010. Adaptation genomics: the next generation. *Trends Ecol Evol.* 25:705–712.

Stawski C, Koteja P, Sadowska ET, Jefimow M, Wojciechowski MS. 2015. Selection for high activity-related aerobic metabolism does not alter the capacity of non-shivering thermogenesis in bank voles. *Comp Biochem Physiol A Mol Integr Physiol.* 180:51–56.

Stern DL, Orgogozo V. 2008. The loci of evolution: how predictable is genetic evolution? *Evolution* 62:2155–2177.

Stopková R, Zdráhal Z, Ryba.Š, Šedo O, Šandera M, Stopka P. 2010. Novel OBP genes similar to hamster Aphrodisin in the bank vole, *Myodes glareolus*. *BMC Genomics* 11:45.

Storz JF. 2005. INVITED REVIEW: Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol.* 14(3):671–688.

Stuglik MT, Babik W, Prokop Z, Radwan J. 2014. Alternative reproductive tactics and sex-biased gene expression: the study of the bulb mite transcriptome. *Ecol Evol.* 4:623–632.

Tan M-KM, Lim H-J, Harper JW. 2011. SCFFBXO22 regulates histone H3 lysine 9 and 36 methylation levels by targeting histone demethylase KDM4A for ubiquitin-mediated proteasomal degradation. *Mol Cell Biol.* 31:3687–3699.

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012. The molecular diversity of adaptive convergence. *Science* 335:457–461.

Teotónio H, Chelo IM, Bradić M, Rose MR, Long AD. 2009. Experimental evolution reveals natural selection on standing genetic variation. *Nat Genet.* 41:251–257.

Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, Hirschhorn JN. Genetic Investigation of ANthropometric Traits (GIANT) Consortium. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet.* 44(9):1015–1019.

Turner TL, Miller PM. 2012. Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. *Genetics* 191:633–642.

Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. 2011. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genet.* 7:e1001336.

Vijay N, Poelstra JW, Künstner A, Wolf JB. 2013. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. *Mol Ecol.* 22(3):620–634.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456(7221):470–476.

Wolfarth B, Rankinen T, Hagberg JM, Loos RJF, Pérusse L, Roth SM, Sarzynski MA, Bouchard C. 2014. Advances in exercise, fitness, and performance genomics in 2013. *Med Sci Sports Exerc.* 46:851–859.

Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet.* 8:206–216.