HHS Public Access

Author manuscript

Behav Genet. Author manuscript; available in PMC 2018 March 01.

Published in final edited form as:

Behav Genet. 2017 March ; 47(2): 227–243. doi:10.1007/s10519-016-9819-x.

Genetic and Genomic Response to Selection for Food Consumption in *Drosophila melanogaster*

Megan E. Garlapow[1,2,3], **Logan J. Everett**[1,2,3,4], **Shanshan Zhou**[1,2,3,4], **Alexander W. Gearhart**[2,3], **Kairsten A. Fay**[2,3], **Wen Huang**[1,2,3,4], **Tatiana V. Morozova**[2,3], **Gunjan H. Arya**[2,3], **Lavanya Turlapati**[2,3], **Genevieve St. Armour**[2,3], **Yasmeen N. Hussain**[2,3], **Sarah E. McAdams**[2,3], **Sophia Fochler**[2,3,5], and **Trudy F. C. Mackay**[1,2,3,4]

[1]Program in Genetics, North Carolina State University, Raleigh, NC 27695-7614

[2]Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695-7614

[3]W. M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, NC 27695-7614

[4]Initiative for Biological Complexity, North Carolina State University, Raleigh, NC 27695-7614

[5]School of Biosciences and Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK

Abstract

Food consumption is an essential component of animal fitness; however, excessive food intake in humans increases risk for many diseases. The roles of neuroendocrine feedback loops, food

Corresponding Author: Trudy F. C. Mackay, trudy_mackay@ncsu.edu, Tel: 919-515-5810, Fax: 919-515-3355.

Availability of Supporting Data

Raw DNA sequences have been deposited in the NCBI SRA database (SRP062265), and raw RNA sequences in the NCBI GEO database (GSE71966).

Compliance With Ethical Standards

The authors declare that they have no competing interests. This article does not contain any studies with human participants or animals performed by any of the authors.

Authors' Contributions

MEG and TFCM conceived and designed the experiments. MEG, AWG and KAF performed the selection experiment. MEG performed the analysis of body size. SZ performed the analyses of molecular metabolites. SZ and SEM obtained all DNA sequence data. GHA, LT, GSA and YNH obtained all RNA sequence data. LJE analyzed the DNA and RNA sequence data and performed the drift simulations. TVM and SF performed functional validations of candidate genes. WH created and maintained the AIP. MEG, LJE, SZ, WH, TVM and TFCM wrote the manuscript.

Electronic Supplementary Material

Additional file 1 (.xlsx). Genetic divergence of selection lines. If a variant is within 1 kb of more than one gene, all are given.
Additional file 2 (.xlsx). Gene Ontology enrichment analysis of significant genes from genetic divergence of selection lines. (**A**) Significant (Benjamini FDR < 0.05) GO terms. (**B**) Significant (enrichment score > 2) GO clusters.
Additional file 3 (.xlsx). Raw read counts from RNA sequencing analysis. F: females; M: males.
Additional file 4 (.xlsx). Analysis of differential gene expression in High (H) and Low (L) selection lines. F: females M: males.
Additional file 5 (.xlsx). Differentially expressed genes in common between males and females.
Additional file 6 (.xlsx). Gene Ontology cluster enrichment analyses for genes with significant differences in expression between high and low selection lines. (A) Sexes Pooled. (B) Females. (C) Males.
Additional file 7 (.xlsx). Genes that overlap between the genetic and gene expression divergence analyses. If more than one significant variant from the genetic divergence analyses is located in the 5' or 3' UTR, the one with the lowest *P*-value is given. Other putative regulatory variants ± 1 kb of the genes are given. n/a: not applicable (variant is > 1 kb from the nearest gene).
Additional file 8 (.xlsx). Summary statistics, GO category and human orthologs for functional validation of candidate genes associated with food intake.

sensing modalities, and physiological state in regulating food intake are well understood, but not the genetic basis underlying variation in food consumption. Here, we applied ten generations of artificial selection for high and low food consumption in replicate populations of *Drosophila melanogaster*. The phenotypic response to selection was highly asymmetric, with significant responses only for increased food consumption and minimal correlated responses in body mass and composition. We assessed the molecular correlates of selection responses by DNA and RNA sequencing of the selection lines. The high and low selection lines had variants with significantly divergent allele frequencies within or near 2,081 genes and 3,526 differentially expressed genes in one or both sexes. A total of 519 genes were both genetically divergent and differentially expressed between the divergent selection lines. We performed functional analyses of the effects of RNAi suppression of gene expression and induced mutations for 27 of these candidate genes that have human orthologs and the strongest statistical support, and confirmed that 25 (93%) affected the mean and/or variance of food consumption.

Keywords

realized heritability; DNA-seq; RNA-seq; feeding behavior; CAFE assay

Background

All animals consume food to obtain the calories and nutrients necessary for survival, reproduction, growth, and other biological processes. In humans, consumption of excessive calories is associated with an increased incidence of type 2 diabetes, obesity, cardiovascular disease, and other disorders and diseases (Azadbakht et al. 2012; Goncalves et al. 2012; Naja et al. 2012); while insufficient caloric intake is correlated with abnormal liver function and other disorders (Harris et al. 2012; Strumia 2013). In spite of this, little is known about the genetic architecture underlying naturally occurring variation in food intake. Food consumption is a typical quantitative trait, with phenotypic variation in populations due to multiple segregating loci with alleles with small and environmentally sensitive effects as well as direct environmental effects (Falconer and Mackay 1996; Garlapow et al. 2015). Dissecting the genetic and environmental contributions to variation in food consumption is challenging in human populations, not least due to the difficulty in accurately quantifying food intake (Basiotis et al. 1987; Schoeller 1990; Schoeller 1995; Kaczkowski et al. 2000; Bray et al. 2008; Champagne et al. 2013). These challenges can be more readily overcome using genetically tractable model organisms in which food consumption can be rapidly and accurately quantified, environmental conditions controlled, and different and complementary experimental designs for genotype-phenotype mapping applied. Inferences about the genetic architecture of food consumption gleaned from studies of model organisms may be relevant to human health given the evolutionary conservation of basic biological processes across eukaryotic taxa.

Single gene studies in *D. melanogaster* using mutations or manipulation of gene expression are yielding insights regarding the multiple genetic factors affecting food consumption in this species. These include genes affecting chemosensation, assessment of food quality and caloric content, and integration of feeding behavior with physiological state (Rajan and

Perrimon 2013; Padmanabha and Baker 2014; Tennessen et al. 2014a; Garlapow et al. 2015). However, little is known about the extent to which alleles of these genes affect phenotypic variation in feeding behavior in natural populations. Recently, we performed a genome wide association study (GWAS) to identify genes and genetic variants affecting food intake in the sequenced inbred lines of the *Drosophila melanogaster* Genetic Reference Panel (DGRP) (Mackay et al. 2012; Huang et al. 2014; Garlapow et al. 2015). We identified 74 common (minor allele frequency (MAF) ≥0.05) variants in or near 54 genes associated with mean food intake and 160 variants in or near 101 genes associated with within-line variance of food intake at a nominal significance threshold of $P \leq 10^{-5}$. We functionally validated many of the candidate genes using RNAi suppression of gene expression. The majority of these genes were not previously implicated in *Drosophila* feeding behavior *per se*, but most were plausible candidates based on functions of mammalian orthologs, tissue-specific expression patterns, and correlations in gene expression with environmental perturbation of feeding. From these data we infer that the genetic architecture of natural variation in food consumption is polygenic, that many of the effects may be mediated *via* transcriptional regulation; and that analysis of natural variation complements single gene perturbation studies.

Here, we complement the GWAS in the DGRP by combining sequence divergence with analysis of divergence in genome wide gene expression (Toma et al. 2002; Tabakoff et al. 2003; Mackay et al. 2005; Dierick and Greenspan 2006; Edwards et al. 2006; Mulligan et al. 2006; Jordan et al. 2007; Morozova et al. 2007; Sørensen et al. 2007; Telonis-Scott et al. 2009; Malmendal et al. 2013; Konczal et al. 2015) to identify strong candidate genes affecting transcriptional regulation of food consumption and to independently functionally validate the DGRP association mapping results. We performed ten generations of replicated, divergent artificial selection for high and low food consumption using an advanced intercross population (AIP) derived from a subset of DGRP lines as the base population. We performed whole genome DNA and RNA sequencing of pools of individuals from the lines at selection generations (G) G9 and G10. We confirmed the effects of 29 genes previously identified by association mapping in the DGRP. We also performed functional tests of RNAi suppression of gene expression and mutations of 27 additional candidate genes implicated by this analysis, and confirmed effects on the mean and/or variance of food consumption for 25 (93%) of these genes. This analysis highlights the importance of *Malic enzyme (Men)* in regulating food consumption. *Men* is thus a focal point for beginning to understand the complex and evolutionary conserved relationships between metabolism, food intake, sleep, lifespan and disease.

Results

Response to selection for food consumption

We used a modified version of the Capillary Feeding (CAFE) Assay (Ja et al. 2007) to artificially select for divergent feeding behavior from an AIP derived from 37 DGRP (Mackay et al. 2012; Huang et al. 2014) lines with maximal genetic diversity and minimal heterozygosity that were not infected with *Wolbachia* and that did not contain common segregating inversions. We derived two high (H1 and H2) and two low (L1 and L2) food

consumption selection lines by 10 generations of selecting the most extreme 20 males and females from 100 scored of each sex. We also maintained an unselected control (C) population at the same census population size. Responses to selection were asymmetric, with response to selection for increased, but not decreased food consumption (Fig. 1A). The realized heritability ($h^2 \pm$ SE) of food intake (calculated from the regressions of cumulative response ($\sum R$) on cumulative selection differential ($\sum S$) (Falconer and Mackay 1996)) was $h^2 = 0.156 \pm 0.018$ ($P < 0.0001$) and $h^2 = 0.148 \pm 0.031$ ($P = 0.0013$) for H1 and H2, respectively; and $h^2 = 0.057 \pm 0.038$ ($P = 0.17$) and $h^2 = 0.0200 \pm 0.044$ ($P = 0.66$) for L1 and L2, respectively (Fig. 1B, 1C). Realized heritabilities from the divergence of high and low lines were $h^2 = 0.120 \pm 0.013$ ($P < 0.0001$) and $h^2 = 0.097 \pm 0.026$ ($P = 0.0056$) for replicate 1 and replicate 2, respectively (Fig. 1D).

Correlated response to selection for food consumption

Causes of genetic correlation include linkage and net directional pleiotropy, whereby segregating polymorphisms in a pleiotropic gene under selection affect not only the selected trait but also the correlated trait (Falconer and Mackay 1996). Body mass and body composition can be correlated with food composition (Jumbo-Lucioni et al. 2010; Reed et al. 2010). We assessed whether body mass, protein, glycogen, and/or triglyceride levels changed as correlated responses to selection. We collected flies from G9 and G10 for the selection and control lines and measured wet weight before flash freezing the same flies for body composition measurements, assessing males and females separately. We pooled the data for both generations to increase our ability to detect consistent correlated selection responses, and used Tukey's honestly significant difference (HSD) tests to assess differences between lines at $P < 0.05$. We did not observe consistent correlated responses in body mass (Fig. 2A, 2B), total protein (Fig. 2C, 2D), female glycogen (Fig. 2E) or triglycerides (Fig. 2G, 2H) between the high and low selection lines. However, glycogen was increased in males in the high selection lines (Fig. 2F) and the trend was for increased male body mass in the high lines and reduced male body mass in the low lines relative to the control (Fig. 2B).

Sequence divergence between selection lines

To identify genomic regions that are targets of selection, we obtained genome sequences of pools of females from the high and low lines at G9 and G10, and from G0. We identified 1,042,033 biallelic SNPs segregating in the populations meeting coverage, frequency, and strand bias criteria (see Methods). We computed allele frequencies in each population, defining the minor allele as the less frequent allele in the AIP. We identified SNPs with significant allele frequency differences in the same direction in both replicate high and low selection lines that were unlikely to be due to sequencing error. However, these differences could have arisen by genetic drift given the small population size of the selection lines. To identify alleles with frequency changes that could be attributable to drift, we first identified SNPs not likely to be under selection as those with high *P*-values for the difference in allele frequency at G10. We binned these SNPs into groups defined by their allele frequency in the base population, and used the variance in allele frequency across all SNPs in the frequency bin across all selection lines to compute the variance effective population size (N_e) (Barton et al. 2007). The median N_e across all initial allele frequencies was 18.7; therefore, we used a census size of $N = 10$ males and 10 females to simulate drift. We validated that our

simulation provides an accurate estimate of drift for this experiment by comparing the values of N_e computed from our simulations to those originally computed from our observed data (Fig. 3). Note that this population size is roughly half of the actual number of progenitors used in the selection experiment and therefore provides a conservative estimate of drift. We then performed 10,000 drift simulations for starting allele frequencies between 0.01 and 0.5 to determine an upper bound on the expected divergence of allele frequencies due to genetic drift, separately for autosomal and *X*-linked loci (Fig. 4). For each significant SNP at each generation, we compared the observed absolute average difference in frequency to that computed from the simulations with the same starting allele frequency, for the same chromosome type. Only SNPs with an absolute average difference in frequency greater than the upper bound computed from the simulation at either G9 or G10 were inferred to be under selection for feeding behavior.

We identified 5,544 SNPs likely to be under selection, of which 1,117 were ≥1 kb from the nearest gene, and 4,427 were near or within 2,081 genes (Fig. 5, Additional file 1). We performed gene ontology (GO) enrichment analysis (Huang et al. 2009a; Huang et al. 2009b) on the significant genes (Table S2). The most significantly enriched GO terms and clusters involved development and function of the nervous system, as expected for a behavior; and the genes were enriched for immunoglobulin domains (Additional file 2). The most significant SNP within a gene is a synonymous T (major)/C (minor) polymorphism in *Sorting nexin 1 (Snx1)* with $P = 9.94 \times 10^{-51}$ with an average allele frequency of 0.113 in the high lines and 0.995 in the low lines. The most significant SNP ≥1 kb from a gene is a G (major)/A (minor) polymorphism at *3L*_6289981, 35,999 bp downstream of *Ecdysone-inducible gene L3 (ImpL3)* and 23,730 bp upstream of *Odorant receptor 65a (Or65a)* ($P = 9.27 \times 10^{-43}$). Given the location of the latter SNP in a gene desert, it is a putative enhancer. *ImpL3* encodes a lactate dehydrogenase involved in carbohydrate metabolism (St. Pierre et al. 2014) and that regulates glucose metabolism in the glycolytic pathway to coordinate metabolism with growth (Nirala et al. 2013; de la Cova et al. 2014; Tennessen et al. 2014b) and nutrient availability (Teleman et al. 2008). The mammalian orthologs of *ImpL3* are selectively down-regulated in healthy pancreatic β cells and up-regulated in type 2 diabetes to induce excessive insulin secretion (Pullen and Rutter 2013).

Gene expression divergence between selection lines

Genes with significant changes in expression between the replicate selection lines are candidate genes affecting variation in food consumption. We assessed genome wide gene expression in the high and low selection lines at G9 and G10, separately for males and females, using RNA sequencing. All flies were at the same age and physiological state as during selection. We analyzed aligned read counts for all 12,991 known genes with sufficient expression across all samples (Additional file 3) using the generalized linear model (GLM)-based methods (McCarthy et al. 2012) in the EdgeR package (Robinson et al. 2010). We included all relevant primary and interaction effect terms in our model (see Methods), and fit three separate models encompassing: (1) females only; (2) males only; (3) females and males combined with primary sex effects and secondary sex interactions included. For each model, we defined differentially expressed genes as those passing a 5%

FDR threshold from the EdgeR likelihood ratio test on the high vs. low lines across both replicate lines at both generations (Additional file 4).

In total, 811 (1,341) genes were down-regulated and 456 (980) were up-regulated in the high line females (males) relative to the low lines. In the analysis pooled across sexes, 1,033 genes were down-regulated and 598 were up-regulated in the high lines, 338 of which are unique to the pooled analysis. Of the 1,267 differentially expressed genes in females and the 2,321 differentially expressed genes in males, 400 are shared between both sexes (Additional file 5), which are more than predicted by chance (hypergeometric test $P = 4 \times 10^{-36}$). Of the 400 shared differentially expressed genes, 74 had sexually antagonistic gene expression differences between high and low lines: 52 genes are down-regulated in high line females and up-regulated in high line males, and 22 are up-regulated in high line females and down-regulated in high line males (Additional file 5).

We used GO cluster analyses (Huang et al. 2009b) to infer processes and functions enriched in up- and down-regulated genes, separately for the pooled sexes, female and male analyses (Additional file 6). In all analyses, down-regulated genes in the high lines were enriched for terms involved in mitochondrial function while up-regulated genes in the high lines were enriched for terms involving protein synthesis and the cell cycle, mitosis, meiosis and DNA replication. In females, genes involved in egg production and muscle function and development were down-regulated in the high lines while in males, genes involved in metabolic processes and detoxification of xenobiotics were up-regulated in high lines. Thus, although the genes with divergent gene expression are largely different between males and females, many belong to the same enriched categories while others give enrichment clusters specific to each sex. These results suggest a complex genetic regulatory mechanism by which similar feeding outcomes are achieved between the sexes.

Integrating genetic and gene expression divergence

Causal variants affecting food consumption are among those with divergent allele frequencies between the high and low selection lines. However, selection introduces linkage disequilibrium (LD) between the selected variants and nearby genes, and hence it is not possible to determine which of many variants in LD are causal. Candidate genes affecting food consumption behavior *via* a transcriptional mechanism are also among the genes with differences in expression between the high and low selection lines. However, gene expression analyses alone cannot discriminate between gene expression changes causing phenotypic divergence in food consumption from those that are a consequence of phenotypic divergence; and a *cis*-regulatory change in expression of one gene can cause *trans*-regulatory changes in gene expression of other genes, leading to correlated gene expression modules (Sieberts and Schadt 2007; Ayroles et al. 2009; Jumbo-Lucioni et al. 2010; Civelek and Lusis 2014; Reed et al. 2014). We hypothesize that the inclusion of additional non-causal genes from LD in the genetic divergence analysis and from transcriptional co-regulation from the transcript divergence analysis led to the lack of correspondence between the enriched GO categories from the two analyses. Integrating the results of genetic and gene expression divergence should enable us to identify the most likely candidate genes affecting variation in food consumption. We found a total of 519 genes that overlap between the two

analyses (Additional file 7), which is greater than expected by chance (hypergeometric test, $P = 0.0029$).

Functional validation

We identified the top 28 candidate genes (Additional file 8) for functional evaluation using RNAi-mediated reduction in gene expression or analyses of insertional mutations, based on public availability of reagents, the strength of the signal in the genetic divergence and transcriptional divergence studies, existence of a one:one human ortholog, GO classification and gene expression pattern. We also required that at least one variant in the 3' and 5' UTR region was significant from the genetic divergence analysis, because most *cis*-eQTLs are located in these regions (Massouras et al. 2012). We performed RNAi knockdown of gene expression of 26 candidate genes using a weak ubiquitously expressed *GAL4* driver (*Ubi156-GAL4*, Garlapow et al. 2015); all but one of the candidate genes (*Ribosomal protein L30*) yielded viable F1 progeny. We also assessed effects of two *Minos*-element insertional mutations.

Causal variants associated with response to selection for food consumption could exert their effects on mean food intake, but also through an increased within-genotype (residual) variance in food intake. Indeed, our previous GWAS for food intake using the DGRP lines revealed many candidate genes affecting the magnitude of the residual variance of this trait. Therefore we evaluated the effects of the candidate genes on both mean and variance of food consumption, relative to the appropriate control genotypes (Additional file 8). In total, 22 of the 27 RNAi and mutant genotypes had significant effects on mean food consumption in at least one sex (Fig. 6A, 6B), while 15 of the candidate genes had significant effects on variance of food intake (Additional file 8). Importantly, several of the candidate genes that did not have significant effects on mean food consumption were significantly different from the control at the level of the within-genotype variance in both sexes (*CG5335* and *Cyt-c-p*) or only in one sex (*CG15160*, *CG10924*, *CG10469, OstDelta*, *Rab9* and *Rae1*, Additional file 8). Thus, in total 25 of the 27 candidate genes tested (93%) affected the mean and/or variance of food intake. None of these genes have been previously associated with food consumption in *Drosophila*. They are involved in a range of biological processes, including metabolic processes (*CG10467*, *CG6629*, *Fdh*, *Men*, *Prat2*), neurogenesis (*CG3964* and *Ciao1*), proteolysis (*CG12374*, *CG10469* and *Mmp2*) and oxidative phosphorylation (*wal*) (Additional file 8).

RNAi suppression of gene expression of *Men* was associated with decreased mean food consumption in both sexes. Previously, we had shown that a *P{GT1}* insertional allele of *Men* (*Men*^{BG12824}) was associated with increased resistance to the inebriating effects of alcohol (Morozova et al. 2011). We therefore assessed mean food consumption for *Men*^{BG12824},two precise revertant lines that we generated previously (Morozova et al. 2011) and the Canton S B co-isogenic control line (Fig. 6C). We found that *P*-element insertion in the first exon of *Men* caused a significant decrease in food consumption compared to the Canton S B control similar to the effect of *Men* RNAi knockdown (Figure 6A) in a completely different genetic background. Precise excision of the *P*-element restored mean

food consumption to control or near-control values. Thus, we have confirmed an important role of *Men* in regulating food consumption.

Discussion

The response to artificial selection for food consumption in *D. melanogaster* is asymmetric, with significant selection response in the direction of increased, but not decreased food consumption. The low realized heritability ($h^2 = 0.15$) in the high lines and from the high-low divergence ($h^2 = 0.11$ average) as well as the asymmetric selection response is the hallmark of a trait closely related to fitness (Frankham 1990; Falconer and Mackay 1996). Food intake affects fitness (Lee et al. 2008; Reddiex et al. 2013). Traits that are components of fitness are expected to have reduced additive genetic variance (Falconer and Mackay 1996; Fisher 1999), mostly attributable to alleles segregating due to mutation-selection balance (Falconer and Mackay 1996; Fisher 1999). Thus, there will be little selection response in the direction of increased fitness, but response in the direction of decreased fitness (Cunningham and Siegel 1978; Mackay et al. 2005; Vishalakshi and Singh 2009; Dobler and Hosken 2010). From this, we infer that increased food consumption is deleterious in *D. melanogaster*.

We did not observe consistent correlated responses to selection for body mass, total protein, triglycerides, and female glycogen content. It is possible, though unlikely, that genes affecting food consumption do not have directional pleiotropic effects on these traits and selection response is not directly related to body size or lipid energy storage. However, it is important to note that correlated responses to selection depend on the narrow sense heritability of both the directly selected and correlated traits, the genetic correlation between these traits, the phenotypic standard deviation of the correlated trait and the selection intensity (Falconer and Mackay 1996). Thus, if the heritabilities of both the selected and correlated traits are low, we do not expect much correlated response after only ten generations of direct selection even with moderately high genetic correlations. We did observe a correlated response in male glycogen content, suggesting a sex-specific effect of selection on increased energy stores; and a trend in the direction of increased male body size in the high-consuming lines and in the direction of decreased male body size in the low-consuming lines, suggesting the latter explanation is correct. Glycogen levels are both sex- and sucrose-concentration-dependent (Rovenko et al. 2015). While we did not vary the sucrose concentration, the individuals of the high selection lines consumed more sucrose. At G10 the average food consumption (± SD) was 2.12 ± 0.75 μL/fly for the high lines vs. 0.64 ± 0.42 μL/fly for the control and low lines, a threefold difference.

To gain insight into the genetic and genomic correlates of response to selection for food consumption, we assessed genetic divergence between the high and low selection lines at G9 and G10 by DNA sequencing of pools of females (which are diploid for all chromosomes) and the differences in genome wide gene expression by RNA sequencing of males and females. The two approaches are complementary. Causal variants associated with food consumption are among the genetically divergent variants, but there is additional noise from genetic drift and LD arising from both drift and selection. Genes affecting food consumption *via* a transcriptional mechanism are among those with differences in expression between the

selection lines, but a single *cis*-regulatory variant can affect the expression of modules of co-regulated genes. Therefore we also integrated information from genetic divergence and divergence in gene expression to identify candidate genes with putative *cis*-regulatory variants affecting natural genetic variation in food consumption. The candidate genes implicated in each of these analyses indicate that genes previously known to affect feeding behavior by analysis of induced mutations are also associated with natural genetic variation in food consumption; confirm our previous GWAS of food consumption in the DGRP inbred lines; and implicate novel and plausible candidate genes for future study.

Some of the genetic mechanisms affecting food consumption in *D. melanogaster* have been revealed from effects of induced mutations and manipulation of gene expression (Rajan and Perrimon 2013; Padmanabha and Baker 2014; Tennessen et al. 2014a). Notably, variants in some of these genes had different allele frequencies in the high and low selection lines, suggesting these genes also affect natural variation in food consumption. Specifically, we identified genetic divergence in genes previously known to affect food intake, such as two genetically divergent SNPs in the *Insulin Receptor (InR)* gene, four in the gene encoding the *Leucokinin receptor (Lkr)*, one upstream of the gene encoding *Tryptophan hydroxylase (Trh)*, one in the gene encoding *5-hydroxytryptamine receptor 1A (5-HT1A)*, two near the *Serotonin Transporter (SerT)* gene (one upstream, one in the 3'UTR), two in the gene encoding *short neuropeptide F (sNPF)*, and one in the gene encoding the *sNPF* receptor, *sNPF-R* (Britton et al. 2002; Lee et al. 2004; Wu et al. 2005; Al-Anzi et al. 2010; Neckameyer 2010; Hong et al. 2012; Huser et al. 2012; Kapan et al. 2012; Nässel et al. 2013; Luo et al. 2014; Liu et al. 2015). These results suggest that variants in genes previously known to affect feeding from single gene studies may partially underlie natural genetic variation in food intake.

We did not identify any of these variants in genes previously known to affect feeding behavior in our GWAS for food consumption in the DGRP lines (Garlapow et al. 2015). This could be because the alleles were too rare to be included in the GWAS, which used common (MAF > 0.05) variants only. However, effects of low frequency alleles can be detected by selection if they are present in the AIP and have large effects. A second possibility is that the alleles are common in the DGRP, but their effects are too small to be detected in a population of ~200 lines, or they interact epistatically with other loci (Huang et al. 2012; Mackay 2015; Shorter et al. 2015). In the latter case, differences in allele frequency at the interacting loci between the DGRP and AIP used to derive the selection lines can cause a difference in the additive effects of the focal allele between the two populations (Huang et al. 2012; Mackay 2015; Shorter et al. 2015). We therefore compared the frequency of the genetically divergent SNPs in these genes between the DGRP and the AIP. One of four SNPs in *Lkr, 3L*_5519044 (minor allele frequency = 0.029 in the DGRP), and one of two SNPs in *sNPF, 2L*_20038844 (minor allele frequency = 0.041 in the DGRP), are not sufficiently frequent to be evaluated for association in the DGRP. The remaining eleven SNPs have high *P*-values ($P < 10^{-1}$ to $P < 10^{-2}$) and correspondingly small effect sizes in the DGRP. The selection experiment may have had increased power to detect these variants, in part because we used a DGRP-derived AIP such that all founder haplotypes are known (Franssen et al. 2015), and we have a smaller multiple testing penalty than with GWAS; alternatively, the effects of these SNPs in the AIP could be greater in the AIP than the DGRP

because allele frequencies of epistatic partners are different between the two populations (Huang et al. 2012; Mackay 2015; Shorter et al. 2015).

The genetic and genomic responses to selection can also serve as a functional validation for genes and additive variants implicated by the GWAS for food consumption using the inbred DGRP lines. Of the 54 candidate genes identified in the GWAS for mean food consumption, 24 were identified in one or both analyses reported here. Eleven of the DGRP candidate genes were identified in the analysis of genetic divergence (*CG13229*, *CG2121*, *CG30287*, *lilli*, *LRP1*, *Mp*, *mspo*, *Nipped-A*, *SKIP*, *Spn*, *tinc*); nine in the analysis of expression divergence (*CG1136*, *CG11929*, *CG15653*, *CG32107*, *G34356*, *CG42788*, *Grp*, *tmod*, *trp*); and four in both analyses (*Egfr*, *Lmpt*, *MESR3*, *retn*). In addition, five candidate genes affecting the within line variance in food consumption were identified in both analyses in this study (*CG1136*, *CG42747*, *Cpr97Ea*, *fz*, *kdn*). We had previously assessed the effects on food intake of 13 of these 24 overlapping candidate genes between the DGRP GWAS and this study, and found that 11 affected food consumption (*CG2121*, *CG32107*, *Egfr*, *Grp*, *lilli*, *Lmpt*, *LRP1*, *Spn*, *tinc*, *tmod*, *trp*).

The Epidermal Growth Factor (EGF) signaling pathway previously has been shown to regulate the insulin-like signaling pathway in post-prandial metabolism (Zhang et al. 2011b; Brankatschk et al. 2014), though components of EGF signaling had not been known to affect food intake volume *per se*. Previously, we showed that an RNAi knockdown of the *Egf receptor* (*Egfr*) elicited a male-specific increase in food intake (Garlapow et al. 2015). Here, we observe down-regulation of *Egfr* expression in the high selection lines in the pooled analysis of differential gene expression (FDR = 0.011), and five intronic SNPs within *Egfr* have divergent allele frequencies (the most significant is SNP *2R*:17419950, $P = 1.6 \times 10^{-09}$). Similarly, *LDL receptor protein 1* (*LRP1*), a component of the EGF signaling pathway, was identified in the DGRP GWAS and elicited increases in male food intake when knocked down with RNAi (Garlapow et al. 2015). Here, we observe thirteen SNPs (twelve intronic, one non-synonymous with lowest $P = 7.25 \times 10^{-24}$) of *LRP1* with differences in allele frequency consistent with selection. *LRP1* is involved in modulating post-prandial insulin signaling in *Drosophila* (Brankatschk et al. 2014) and mammals (Hofmann et al. 2007; Liu et al. 2011) and here further implicated in underlying natural variation in food consumption.

The remaining genes identified by divergence in allele frequencies between the high and low selection lines and/or that overlap between the genetic and genomic analyses are plausible and interesting candidate genes for future analyses. Two of these genes further implicate aspects of EGF signaling with the regulation of food consumption. The most significant SNP from the DNA sequencing analysis lies in *Snx1*, which is also up-regulated in females from the high selection lines. Snx1 is an essential component of the retromer complex and is a positive regulator of Wnt protein secretion (Zhang et al. 2011a). In mammals, SNX1 interacts with the EGFR as part of the mechanism by which EGFR is degraded (Kurten et al. 1996). We also observed ten SNPs (nine intronic, one non-synonymous, smallest $P = 7.77 \times 10^{-13}$) in *rugose* (*rg*), a known modulator of EGF and Notch signaling (Shamloula et al. 2002). The mammalian orthologues of *rg* are *Neurobeachin* (*Nbea*) and *LPS-responsive, beige-like anchor* (*Lrba*) (Volders et al. 2012). Mice heterozygous for an *Nbea* null allele

increased food intake and white adipose, while in humans, variants in *Nbea* are significantly associated with an increased body mass index (Olszewski et al. 2012). NBEA interacts with the glycine receptor β subunit in the central nervous system, where it may help traffic neurotransmitters (del Pino et al. 2011). In *D. melanogaster*, glycine cleavage has been implicated in the cessation of eating (Zinke et al. 1999; Melcher et al. 2007) and glycine levels in larvae are predictive of a variety of metabolically related phenotypes, including adult heart arrhythmia, sugar levels, and lipid levels (Reed et al. 2014). Recently, *rg* was shown to affect microbiome-dependent triacylglycerol and glucose levels (Dobson et al. 2015). Previous research on *Drosophila* and mammals positions *rg* as a plausible candidate gene directly affecting food intake. Future work is necessary to assess whether *rg* directly regulates food intake and whether it interacts with glycine perception to do so.

Finally, we functionally assessed the effects of 27 top candidate genes that overlapped between the genetic and gene expression divergence analyses, finding that 25 affected the mean and/or residual variance of food consumption. Many of these genes affect aspects of metabolism. In particular, both RNAi suppression of *Men* gene expression and a *P*-element insertional mutation in the first exon of *Men* cause a reduction in food intake in both sexes, and precise revertant alleles restore food consumption at or near control levels. Thus, the effect of *Men* on food intake is supported by association analyses of inbred DGRP lines and outbred populations derived from a subset of these lines; changes in gene expression as a correlated response to selection; and functional tests in two different genetic backgrounds. *Men* encodes Malic enzyme, a key component of intermediary metabolism, and affects lifespan (Paik et al. 2012), regulation of cell death (Yang et al. 2010), sleep (Harbison and Sehgal 2008) and alcohol resistance (Morozova et al. 2011) in *Drosophila*. In mammals, *Malic enzyme 3 (ME3)* is important for appropriate pancreatic β-cell glucose-stimulated insulin secretion (Hasan et al. 2015). Thus, this gene may be one key to understanding the complex and evolutionary conserved relationships between metabolism, food intake, sleep, lifespan and disease.

Methods

Drosophila stocks

We created an advanced intercross outbred population (AIP) from 37 of the 205 sequenced, wild-derived, inbred DGRP lines (Mackay et al. 2012; Huang et al. 2014). These 37 lines are minimally related (polygenic relatedness < 0.08), maximally homozygous (genome-wide heterozygosity < 5%), have the standard karyotype for all common polymorphic inversions, and are not infected with *Wolbachia*. The lines were first crossed in a round robin design (Huang et al. 2012) to ensure equal representation of each genotype. At generation (G) zero, one virgin F1 female and one F1 male from each of the 37 lines was placed into each of 10 bottles. The AIP was maintained by random mating by mixing progeny from all bottles and distributing 40 females and 40 males into each of 10 new bottles, for a census population size of 800. To minimize natural selection via larval competition, egg-laying was restricted to 24 hours. We initiated the selection experiment at AIP G24. The AIP and selection lines were reared in bottles on cornmeal/molasses/agar medium under standard culture conditions (25°C, 12:12 hour light/dark cycle).

We obtained RNAi transgenic fly lines (*Cep97*^{103357}, *CG10467*^{103787} *CG10924*^{107092}, *CG12374*^{101566}, *CG15160*^{107028}, *CG3964*^{102197}, *CG5335*^{101295}, *CG6629*^{106108}, *Ciao1*^{105939}, *Cyt-c-p*^{106759}, *Drip*^{106911}, *Ef1alpha48D*^{104502}, *Fdh*^{110071}, *Ih*^{110274}, *Lk6*^{109663}, *Men*^{104016}, *Mmp2*^{107888}, *OstDelta*^{107068}, *Prat2*^{108948} *Rab9*^{107192}, *Rae1*^{101338} *RpL30*^{101391}, *SelR*^{110755}, *Sod2*^{110547}, *TfIIFβ*^{110569} and *wal*^{103811}) and the corresponding progenitor line with the empty vector site (60010) from the Vienna Drosophila RNAi Center (VDRC) (Dietzl et al. 2007). We crossed these lines to a weak, ubiquitously expressed driver (*Ubiquitin-GAL4*), and we crossed the driver to the progenitor RNAi controls (v60010) to serve as our experimental controls (Garlapow et al. 2015). *Mi{ET1}* mutants (*CG9119*^{25661}, *CG10469*^{27875}) (Bellen et al. 2011) and a *P{GT1}Men*^{12824} mutant (Norga et al. 2003) were obtained from the Bloomington Drosophila stock center. The corresponding *Mi{ET1}* (*w*^{1118}) and *P{GT1}* isogenic control lines (Canton S B) are maintained in our laboratory. *Men* revertant alleles were generated using crossing schemes that preserved the co-isogenic background of the revertant lines by Morozova et al. (2011) and are maintained in our laboratory.

Food consumption assays

We used virgin 3–7 day old flies in all assays, placing each fly in an individual vial with food following CO_2 anesthesia. After a minimum recovery period of 24 hours, we placed each fly into individual vials containing 2 mL 1.5% agar medium and one 5 μL capillary tube (Kimble Glass Inc.) containing a 4% sucrose solution (w/v) inserted through a foam plug. Beginning at G8, we inserted two capillaries per vial for the high consumption replicate populations. Our modified version of the CAFE assay (Ja et al. 2007) includes the agar medium to prevent desiccation from affecting food consumption. We placed the CAFE vials in a transparent plastic container in which high humidity is maintained with open containers of water to minimize evaporation from the capillary tubes. In addition, we assessed evaporation by placing CAFE vials with capillary tubes containing 4% sucrose but without flies in the same humid chamber. We measured total food consumption in each CAFE vial after 24 hours, between 10 A.M. and 12 P.M, and then transferred the flies into individual vials containing standard culture medium to allow them to continue eating while we determined which individuals to select. We adjusted the total amount of food consumed by the average evaporation that occurred in the negative control vials to give the phenotype of each fly.

Mass selection

At G0 for each of the divergent selection replicates, we scored 100 virgin females and males, and selected the 20 individuals of each sex with the highest and lowest food consumption to establish the two high (H1, H2) and two low (L1, L2) selection lines. In each subsequent generation, we selected the 20 highest (or lowest) males and females from 100 scored of each sex as parents of the next generation. We established a control population (C) maintained with 20 unselected males and females each generation. We scored 100 males and females from the control line every two generations. Selection was continued for 10 generations. We estimated h^2, the narrow sense (realized) heritability, for each selection replicate as the regression of the cumulative response to selection (ΣR) on the cumulative selection differential (ΣS) (Falconer and Mackay 1996). At G9 and G10, we assessed

correlated responses to selection for body mass, molecular metabolites, and genome wide genetic and gene expression divergence for all selection lines.

Body weight

We used three-to-five day-old virgin flies for body weight measurements to reduce variability due to egg production and egg-laying in mated flies. Pools of ten flies were measured on an analytical scale in a pre-weighed Eppendorf tube. We measured six replicates per line per sex per generation. We flash-froze the flies after weighing for subsequent body composition analyses.

Molecular metabolites

We measured triglyceride, protein, free glucose, and glycogen in the pooled, weighed tubes of flash-frozen flies. Frozen flies were homogenized in 250 μl PBS buffer using ceramic beads and Tissuelyser (Qiagen.Inc). The homogenate was centrifuged under low speed for 5 minutes at 4°C to remove fly cuticle, and 170 μl supernatant was taken and aliquoted for different measurements. All molecular traits were measured on a SpectraMax Microplate Reader. For measurements of total protein, the homogenate was diluted 10 times and measured using the Quant-iT Protein Assay Kit from Thermo Fisher Scientific. For glycogen measurements, the homogenate was diluted 10 times and heated at 99 °C for 10 minutes to stabilize glycogen. Total glucose hydrolyzed from glycogen including free glucose was measured using the Glycogen Fluorometric Assay kit from BioVision. Baseline free glucose was measured using aliquots of the same sample without adding the hydrolysis enzyme. Glycogen abundance was calculated as total glucose minus free glucose. For triglyceride measurements, undiluted homogenate was measured using Serum Triglyceride Determination Kit from Sigma-Aldrich with longer a incubation time (10 minutes incubation at 37 °C at Step 7, and 60 minutes incubation at 37 °C at Step 9) for a better hydrolysis result. Both free glycerol and total triglycerides were measured from the same sample and the true triglyceride abundance was calculated by subtracting the two measurements.

DNA sequencing

We flash froze 100 female flies from the selection lines at each of generations nine and ten and 173 female flies from the AIP base population used to initiate the selection lines on clean CO_2 pads, producing nine samples. We homogenized flies from each sample using Tissuelyser (Qiagen.Inc) in lysis buffer. Genomic DNA was extracted using magnetic beads and fragmented to 300–400bp using Covaris S220. Fragmented DNA was used to produce barcoded DNA libraries using NEXTflex™ DNA Barcodes (Bioo Scientific, Inc.) with an Illumina TrueSeq compatible protocol. Libraries were quantified using Qubit dsDNA HS Kits (Life Technologies, Inc.) and Bioanalyzer (Agilent Technologies, Inc.) to calculate molarity. Libraries were then diluted to equal molarity and re-quantified, and all samples were pooled together. Pooled library samples were quantified again to calculate final molarity and then denatured and diluted to 14pM. Pooled library samples were clustered on an Illumina cBot and sequenced on two Hiseq2500 high throughput lanes using 125 bp paired-end v4 chemistry.

DNA-seq analysis

Barcoded paired-end sequence reads were demultiplexed using the Illumina pipeline v1.9, and aligned to the *D. melanogaster* reference genome (BDGP5) using BWA v0.7.10 (MEM algorithm with parameters '-v 2 –t 4') (Li and Durbin 2010). GATK (McKenna et al. 2010) was used to locally realign regions around a universal set of indels merged from all libraries, remove duplicate sequence reads, and recalibrate base quality scores. Replicate libraries corresponding to the same population were then merged, and the number of unique base calls for each possible nucleotide at each known SNP was calculated using SAMtools (Li et al. 2009). We considered all biallelic SNPs segregating in the selection lines meeting the following criteria: (1) coverage of Q13 bases ≥10 in each population, including the starting population; (2) at least 80% of coverage had quality of Q13 or better; (3) the two most frequent alleles constituted ≥95% of all observed alleles; (4) the minor allele frequency was ≥5% in at least one selected population; (5) the Chernoff bound of the *P*-value for the null hypothesis that the observed minor alleles were caused by sequencing error (Bansal 2010) was $< 10^{-5}$ whenever the allele frequency was ≥5%; and (6) strand bias was not significant (Fisher's exact test, $P < 10^{-5}$) in any population. For the SNPs passing these criteria, the minor allele was defined as the less frequent allele in the starting population, and allele frequencies in each population were computed as the number of unique base calls corresponding to the minor allele divided by the number of unique base calls corresponding to either the minor or major alleles.

To identify SNPs with significant changes in allele frequency between the high and low lines from each replicate that were unlikely to be due to sequencing error/variation, we performed the following test. Let i denote each pair of high and low selection lines, with allele frequencies $p_H[i]$ and $p_L[i]$, and base coverage $C_H[i]$ and $C_L[i]$, respectively. Let $N_H[i]$ and $N_L[i]$ denote the number of unique chromosomes (2 × number of flies) in each DNA pool.

Let $p_0[i]=\frac{1}{2}(p_H[i]+p_L[i])$ denote the common allele frequency under the null hypothesis, and let $\sigma_H^2[i]=p_0(1-p_0)\left(\frac{1}{C_H[i]}+\frac{1}{N_H[i]}\right)$ denote the expected variance due to sequencing in each population (defined similarly for $\sigma_L^2[i]$), and let $w[i] = (C_H[i]+C_L[i])/\sum(C_H[i] + C_L[i])$ denote the relative weight for each population pair based on base coverage. We then compute test statistic:

$$\chi^2=\frac{\left(\sum_i \frac{p_H[i]-p_L[i]}{w[i]}\right)^2}{\sum_i(\sigma_H^2[i]w[i]^2+\sigma_L^2[i]w[i]^2)}$$

The corresponding *P*-value is calculated using the one-tailed probability density function for the standard χ^2 distribution with one degree of freedom. Individual *P*-values are then corrected for multiple testing by the Bonferroni method. We only rejected the null hypothesis of equal allele frequency in each pair of high and low selection lines when the Bonferroni-corrected *P*-value was ≤0.01. Note that this test only identifies SNPs for which there is a significant difference in allele frequency in one or more population pairs that is

unlikely to be due to random fluctuations in the sequencing process, but it does not address the question of whether the frequency difference occurred due to selection or genetic drift.

To identify allele frequency differences that are not likely to be due to genetic drift among the variants surviving the above test, we applied several additional filters. First, we removed all SNPs for which the sign of $(p_H[i] - p_L[i])$ for at least one pair i of high and low lines differs from that of the other replicate pairs of high and low lines. In other words, we require that the directionality of the allele frequency difference between the high and low lines must be the same in each replicate population, at each generation, which is based on the assumption that selection acts in the same direction for each population, while drift causes changes in allele frequency independently in each population.

Drift simulations

To further estimate the expected allele frequency differences occurring due to drift in our experiment, we performed forward simulations of a single biallelic SNP in a population with a fixed equal number N of male and female flies for 10 generations. Given a starting minor allele frequency q_0, we randomly and independently sample the identity of the allele on each haplotype in each fly. For each subsequent generation, the genotype of each fly is determined by randomly selecting a male and female mating pair from the previous generation, and randomly selecting which haplotype is inherited from each parent. For each simulation, this process is repeated for 4 independent replicates, arbitrarily assigned as high and low lines. We then record the allele frequency for each replicate population at G9 and G10, corresponding to the generations sequenced in our experiment, and compute the average difference in allele frequencies between high and low lines at each generation.

To determine an appropriate population size for simulations of genetic drift, we first estimated effective population size based on the background variance in allele frequency observed across the genome in our DNA-seq samples from G10. We first selected all autosomal SNPs with sufficient minor allele frequency and sequence coverage, as in our primary selection analysis. We then selected the 80% of SNPs with the least significant P-values in our χ^2 test, reasoning that the majority of these SNPs should be free of selection effects. We then binned SNPs by their approximate starting allele frequency (rounded to the nearest hundredth) in the G0 control sample, ranging from 0.05 to 0.49. We excluded SNPs with starting frequency < 0.05 as these tend to have higher technical noise in the estimate of starting frequency. We also excluded SNPs with starting allele frequency = 0.5 because there were too few SNPs in this bin. For each bin with starting allele frequency q_0, we computed the variance σ^2 of allele frequencies across all SNPs and lines (H1, H2, L1, L2) measured at G10. We estimated the variance effective population size using the following equation (Barton et al. 2007):

$$N_e = \frac{-t}{2\ln\left(1 - \frac{\sigma^2}{q_0(1-q_0)}\right)}$$

The median of these estimates across all starting allele frequencies was 18.7. Based on this value, we performed forward simulations of genetic drift for populations of flies with 10 males and 10 females (total $N = 20$).

We repeated the simulation procedure 10,000 times for each starting allele frequency q_0 in a uniform grid of values ranging from 0.01 to 0.5. For each starting allele frequency, we then computed the 0.999 quantile of absolute average difference between high and low selection lines to determine a reasonable upper bound on the expected divergence of allele frequencies due to drift in our experimental setup. This entire simulation procedure was repeated separately for autosomal and *X*-linked loci. For each SNP at each generation, we compared the observed absolute average difference in frequency to that computed from our simulations of the same starting allele frequency, for the same chromosome type. Only SNPs with absolute average difference greater than the upper bound computed from our simulation at either G9 or G10 were inferred as under selection for feeding behavior. These SNPs were then assigned to all genes within 1kb or overlapping.

RNA sequencing

We flash froze at −80 °C three-to-five day-old virgin flies at generations nine and ten, with at least thirty flies per replicate and two replicates per sex per selected population per generation (32 samples) such that each generation of each population and each sex had two biological replicates that were processed separately. All flies were collected between 1 PM and 3 PM. Total RNA was extracted with Trizol using the Quick-RNA MiniPrep kit (Zymo Research; R1055). rRNA was depleted using the Ribo-Zero™ Gold rRNA Removal Kit (Illumina, Inc.) with 5ug total RNA input. Depleted mRNA was fragmented and converted to first strand cDNA. During the synthesis of second strand cDNA, dUTP instead of dTTP was incorporated to label the second strand cDNA. cDNA from each RNA sample was used to produce barcoded cDNA libraries using NEXTflex™ DNA Barcodes (Bioo Scientific, Inc.) with an Illumina TrueSeq compatible protocol. Library size was selected using Agencourt Ampure XP Beads (Beckman Coulter, Inc.) and centered around 250 bp with the insert size ~130 bp. Second strand DNA was digested with Uracil-DNA Glycosylase before amplification to produce directional cDNA libraries. Libraries were quantified using Qubit dsDNA HS Kits (Life Technologies, Inc.) and Bioanalyzer (Agilent Technologies, Inc.) to calculate molarity. Libraries were then diluted to equal molarity and re-quantified. Two pools of 16 libraries were made, one of the 16 libraries from replicate 1 high and low lines; and one of the 16 libraries from replicate 2 high and low lines. Pooled library samples were quantified again to calculate final molarity and then denatured and diluted to 14pM. Pooled library samples were clustered on Illumina cBot; each pool was sequenced on one lane of Illumina Hiseq2500 using 125 bp single-read v4 chemistry.

RNA analysis

Barcoded sequence reads were demultiplexed using the Illumina pipeline v1.9. Adapter sequences were trimmed using cutadapt v1.6 (Martin 2011) and trimmed sequences shorter than 50bp were discarded from further analysis. Trimmed sequences were then filtered for ribosomal RNA sequences by aligning against a database containing the complete 5S, 18S-5p8S-2S-28S, mt:lrRNA, and mt:srRNA sequences using BWA v0.7.10 (MEM

algorithm with parameters '-v 2 –t 4') (Li and Durbin 2010). The remaining sequences were aligned to the *D. melanogaster* genome (BDGP5) and known transcriptome (FlyBase v5.57) using STAR v2.4.0e (Dobin et al. 2013). Read counts were computed for known gene models using HTSeq-count (Anders et al. 2015) with the 'intersection-nonempty' assignment method. Tabulated read counts were then analyzed for all known genes across all samples using EdgeR (Robinson et al. 2010) as follows. First, genes with low expression overall (<10 aligned reads in >75% of the libraries) were excluded from the analysis. Library sizes were recomputed as the sum of reads assigned to the remaining genes, and further normalized using the Trimmed Mean of M-values (TMM) method (Robinson and Oshlack 2010). We then used the generalized linear model (GLM)-based methods (McCarthy et al. 2012) for estimating tagwise dispersion and fit model parameters to the following model design: $X = L + G + L\text{x}G + B + \varepsilon$, where X = observed $\log_2$(read count), L = effect of selection line (H1, H2, L1, L2), G = generation effect (G9 vs G10), $L\text{x}G$ = line by generation interaction effect, B = batch effects (each generation of each population was analyzed using two biological replicates, with the first replicates processed in a separate batch from the second replicates), and ε = model error following a negative binomial distribution with estimated gene-wise dispersion as described in (McCarthy et al. 2012). We applied this model separately to data from males and females. We also fit a full model pooled across males and females: $X = L + S + G + L\text{x}S + L\text{x}G + S\text{x}G + L\text{x}S\text{x}G + B + S\text{x}B + \varepsilon$, where S is the effect of sex and other terms are as defined above. For each model, we selected genes with differential expression as those passing a 5% FDR threshold (based on Benjamini-Hochberg corrected *P*-values) from the EdgeR likelihood ratio test on the contrast of model terms comparing high *vs.* low selection lines across both replicate populations: [(H1+H2)/2-(L1+L2)/2].

Functional validation

We selected 28 genes with top associations for food consumption in the selected lines. We functionally tested 25 genes with available RNAi knock down alleles using the CAFE assay described above, with one alteration. For these experiments, we assessed total consumption for ten groups of eight single sex flies for each sex and genotype. We could not obtain behavioral measurements for $RpL30^{10139}$ due to poor viability of F1 *GAL4-Ubi156/UAS*-RNAi offspring. In addition, we tested two *MiET1* insertional mutations and a *P{GT1}* insertional mutation for *Malic enzyme* gene and its precise revertants. We assessed whether differences in mean food intake between mutant lines and the controls were significant using Dunnett's tests or Student's *t* tests, separately for males and females. We also assessed differences in residual variance between mutant alleles and the controls with pairwise Levene's tests, separately for males and females. Statistical tests were performed using SAS (9.0) solftware.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Stocks obtained from the VDRC and Bloomington Drosophila Stock Center (NIH P40OD018537) were used in this study.

Funding

This work was funded by National Institutes of Health grants R01 GM45146 and R01 AA016560 to TFCM.

References

Al-Anzi B, Armand E, Nagamei P, Olszewski M, Sapin V, Waters C, Zinn K, Wyman RJ, Benzer S. The leucokinin pathway and its neurons regulate meal size in Drosophila. Curr Biol. 2010; 20(11): 969–978. [PubMed: 20493701]

Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; 31(2):166–169. [PubMed: 25260700]

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RR, Mackay TF. Systems genetics of complex traits in *Drosophila melanogaster*. Nat Genet. 2009; 41(3):299–307. [PubMed: 19234471]

Azadbakht L, Haghighatdoost F, Karimi G, Esmaillzadeh A. Effect of consuming salad and yogurt as preload on body weight management and cardiovascular risk factors: a randomized clinical trial. Int J Food Sci Nutr. 2012; 64(4):392–399. [PubMed: 23249429]

Bansal V. A statistical method for the detection of variants from next-generation resequencing of DNA pools. Bioinformatics. 2010; 26(12):318–324.

Barton, NH., Briggs, DEG., Eisen, JA., Goldstein, DB., Patel, NH. Evolution. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2007.

Basiotis PP, Welsh SO, Cronin FJ, Kelsay JL, Mertz W. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. J Nutr. 1987; 117(9):1638–1641. [PubMed: 3655942]

Bellen HJ, Levis RW, He Y, Carlson JW, Evans-Holm M, Bae E, Kim J, Metaxakis A, Savakis C, Schulze KL, Hoskins RA, Spradling AC. The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics. 2011; 188(3):731–743. [PubMed: 21515576]

Brankatschk M, Dunst S, Nemetschke L, Eaton S. Delivery of circulating lipoproteins to specific neurons in the *Drosophila* brain regulates systemic insulin signaling. Elife Oct. 2014; 2:3.

Bray GA, Flatt JP, Volaufova J, Delany JP, Champagne CM. Corrective responses in human food intake identified from an analysis of 7-d food-intake records. Am J Clin Nutr. 2008; 88(6):1504–1510. [PubMed: 19064509]

Britton JS, Lockwood WK, Li L, Cohen SM, Edgar BA. *Drosophila*'s insulin/PI3-kinase pathway coordinates cellular metabolism with nutritional conditions. Dev Cell. 2002; 2(2):239–249. [PubMed: 11832249]

Champagne CM, Han H, Bajpeyi S, Rood J, Johnson WD, Lammi-Keefe CJ, Flatt JP, Bray GA. Day-today variation in food intake and energy expenditure in healthy women: the Dietitian II Study. J Acad Nutr Diet. 2013; 113(11):1532–1538. [PubMed: 24021734]

Civelek M, Lusis AJ. Systems genetics approaches to understand complex traits. Nat Rev Genet. 2014; 15(1):34–48. [PubMed: 24296534]

Cunningham DL, Siegel PB. Response to bidirectional and reverse selection for mating behavior in Japanese quail *Coturnix coturnix japonica*. Behav Genet. 1978; 8(5):387–397. [PubMed: 736872]

de la Cova C, Senoo-Matsuda N, Ziosi M, Wu DC, Bellosta P, Quinzii CM, Johnston LA. Supercompetitor status of *Drosophila* Myc cells requires p53 as a fitness sensor to reprogram metabolism and promote viability. Cell Metab. 2014; 19(3):470–483. [PubMed: 24561262]

del Pino I, Paarmann I, Karas M, Kilimann MW, Betz H. The trafficking proteins Vacuolar Protein Sorting 35 and Neurobeachin interact with the glycine receptor β-subunit. Biochem Biophys Res Commun. 2011; 412(3):435–440. [PubMed: 21821005]

Dierick HA, Greenspan RJ. Molecular analysis of flies selected for aggressive behavior. Nat Genet. 2006; 38(9):1023–1031. [PubMed: 16906161]

Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, Gasser B, Kinsey K, Oppel S, Scheiblauer S, Couto A, Marra V, Keleman K, Dickson BJ. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature. 2007; 448(7150):151–156. [PubMed: 17625558]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013; 29(1):15–21. [PubMed: 23104886]

Dobler R, Hosken DJ. Response to selection and realized heritability of sperm length in the yellow dung fly (*Scathophaga stercoraria*). Heredity (Edinb). 2010; 104(1):61–66. [PubMed: 19639007]

Dobson AJ, Chaston JM, Newell PD, Donahue L, Hermann SL, Sannino DR, Westmiller S, Wong AC, Clark AG, Lazzaro BP, Douglas AE. Host genetic determinants of microbiota-dependent nutrition revealed by genome-wide analysis of *Drosophila melanogaster*. Nat Commun. 2015; 6:6312. [PubMed: 25692519]

Edwards AC, Rollmann SM, Morgan TJ, Mackay TF. Quantitative genomics of aggressive behavior in *Drosophila melanogaster*. PLoS Genet. 2006; 2(9):e154. [PubMed: 17044737]

Falconer, DS., Mackay, TFC. Introduction to Quantitative Genetics. Essex, England: Longman; 1996.

Fisher, RA. The Genetical Theory of Natural Sselection: a Complete Variorum Edition. Oxford: Oxford University Press; 1999.

Frankham R. Are responses to artificial selection for reproductive fitness characters consistently asymmetrical? Genet Res. 1990; 56(1):35–42.

Franssen SU, Nolte V, Tobler R, Schlötterer C. Patterns of linkage disequilibrium and long range hitchhiking in evolving experimental *Drosophila melanogaster* populations. Mol Biol Evol. 2015; 32(2):495–509. [PubMed: 25415966]

Garlapow ME, Huang W, Yarboro MT, Peterson KR, Mackay TF. Quantitative genetics of food intake in *Drosophila melanogaster*. PLoS One. 2015; 10(9):e0138129. [PubMed: 26375667]

Goncalves H, Gonzalez DA, Araujo CP, Muniz L, Tavares P, Assuncao MC, Menezes AM, Hallal PC. Adolescents' perception of causes of obesity: unhealthy lifestyles or heritage? J Adolesc Health. 2012; 51(6 Suppl):S46–S52. [PubMed: 23283160]

Harbison ST, Sehgal A. Quantitative genetic analysis of sleep in *Drosophila melanogaster*. Genetics. 2008; 178(4):2341–2360. [PubMed: 18430954]

Harris RH, Sasson G, Mehler PS. Elevation of liver function tests in severe anorexianervosa. Int J Eat Disord. 2013; 46(4):369–374. [PubMed: 23280717]

Hasan NM, Longacre MJ, Stoker SW, Kendrick MA, MacDonald MJ. Mitochondrial malic enzyme 3 is important for insulin secretion in pancreatic β-cells. Mol Endocrinol. 2015; 29(3):396–410. [PubMed: 25594249]

Hofmann SM, Zhou L, Perez-Tilve D, Greer T, Grant E, Wancata L, Thomas A, Pfluger PT, Basford JE, Gilham D, Herz J, Tschöp MH, Hui DY. Adipocyte LDL receptor-related protein-1 expression modulates postprandial lipid transport and glucose homeostasis in mice. J Clin Invest. 2007; 117(11):3271–3282. [PubMed: 17948131]

Hong SH, Lee KS, Kwak SJ, Kim AK, Bai H, Jung MS, Kwon OY, Song WJ, Tatar M, Yu K. Minibrain/Dyrk1a regulates food intake through the Sir2-FOXO-sNPF/NPY pathway in *Drosophila* and mammals. PLoS Genet. 2012; 8(8):e1002857. [PubMed: 22876196]

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4(1):44–57. [PubMed: 19131956]

Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37(1):1–13. [PubMed: 19033363]

Huang W, Massouras A, Inoue Y, Peiffer J, Rámia M, Tarone A, Turlapati L, Zichner T, Zhu D, Lyman R, Magwire M, Blankenburg K, Carbone MA, Chang K, Ellis L, Fernandez S, Han Y, Highnam G, Hjelmen C, Jack J, Javaid M, Jayaseelan J, Kalra D, Lee S, Lewis L, Munidasa M, Ongeri F, Patel S, Perales L, Perez A, Pu L, Rollmann S, Ruth R, Saada N, Warner C, Williams A, Wu YQ, Yamamoto A, Zhang Y, Zhu Y, Anholt R, Korbel J, Mittelman D, Muzny D, Gibbs R, Barbadilla

A, Johnston S, Stone E, Richards S, Deplancke B, Mackay T. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. Genome Res. 2014; 24(7):1193–1208. [PubMed: 24714809]

Huang W, Richards S, Carbone MA, Zhu D, Anholt RR, Ayroles JF, Duncan L, Jordan KW, Lawrence F, Magwire MM, Warner CB, Blankenburg K, Han Y, Javaid M, Jayaseelan J, Jhangiani SN, Muzny D, Ongeri F, Perales L, Wu YQ, Zhang Y, Zou X, Stone EA, Gibbs RA, Mackay TF. Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. Proc Natl Acad Sci U S A. 2012; 109(39):15553–15559. [PubMed: 22949659]

Huser A, Rohwedder A, Apostolopoulou AA, Widmann A, Pfitzenmaier JE, Maiolo EM, Selcho M, Pauls D, von Essen A, Gupta T, Sprecher SG, Birman S, Riemensperger T, Stocker RF, Thum AS. The serotonergic central nervous system of the *Drosophila* larva: anatomy and behavioral function. PLoS One. 2012; 7(10):e47518. [PubMed: 23082175]

Ja WW, Carvalho GB, Mak EM, de la Rosa NN, Fang AY, Liong JC, Brummel T, Benzer S. Prandiology of Drosophila and the CAFE assay. Proc Natl Acad Sci U S A. 2007; 104(20):8253–8256. [PubMed: 17494737]

Jordan KW, Carbone MA, Yamamoto A, Morgan TJ, Mackay TF. Quantitative genomics of locomotor behavior in *Drosophila melanogaster*. Genome Biol. 2007; 8(8):R172. [PubMed: 17708775]

Jumbo-Lucioni P, Ayroles JF, Chambers MM, Jordan KW, Leips J, Mackay TF, De Luca M. Systems genetics analysis of body weight and energy metabolism traits in *Drosophila melanogaster*. BMC Genomics. 2010; 11:297. [PubMed: 20459830]

Kaczkowski CH, Jones PJ, Feng J, Bayley HS. Four-day multimedia diet records underestimate energy needs in middle-aged and elderly women as determined by doubly-labeled water. J Nutr. 2000; 130(4):802–805. [PubMed: 10736333]

Kapan N, Lushchak OV, Luo J, Nässel DR. Identified peptidergic neurons in the *Drosophila* brain regulate insulin-producing cells, stress responses and metabolism by coexpressed short neuropeptide F and corazonin. Cell Mol Life Sci. 2012; 69(23):4051–4066. [PubMed: 22828865]

Konczal M, Babik W, Radwan J, Sadowska ET, Koteja P. Initial molecular-level response to artificial selection for increased aerobic metabolism occurs primarily through changes in gene expression. Mol Biol Evol. 2015; 32(6):1461–1473. [PubMed: 25739734]

Kurten RC, Cadena DL, Gill GN. Enhanced degradation of EGF receptors by a sorting nexin, SNX1. Science. 1996; 272(5264):1008–1010. [PubMed: 8638121]

Lee KP, Simpson SJ, Clissold FJ, Brooks R, Ballard JW, Taylor PW, Soran N, Raubenheimer D. Lifespan and reproduction in *Drosophila*: New insights from nutritional geometry. Proc Natl Acad Sci U S A. 2008; 105(7):2498–2503. [PubMed: 18268352]

Lee KS, You KH, Choo JK, Han YM, Yu K. *Drosophila* short neuropeptide F regulates food intake and body size. J Biol Chem. 2004; 279(49):50781–50789. [PubMed: 15385546]

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26(5):589–595. [PubMed: 20080505]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–2079. [PubMed: 19505943]

Liu Q, Zhang J, Zerbinatti C, Zhan Y, Kolber BJ, Herz J, Muglia LJ, Bu G. Lipoprotein receptor LRP1 regulates leptin signaling and energy homeostasis in the adult central nervous system. PLoS Biol. 2011; 9(1):e1000575. [PubMed: 21264353]

Liu Y, Luo J, Carlsson MA, Nässel DR. Serotonin and insulin-like peptides modulate leucokinin-producing neurons that affect feeding and water homeostasis in *Drosophila*. J Comp Neurol. 2015; 523(12):1840–1863. [PubMed: 25732325]

Luo J, Lushchak OV, Goergen P, Williams MJ, Nässel DR. *Drosophila* insulin-producing cells are differentially modulated by serotonin and octopamine receptors and affect social behavior. PLoS One. 2014; 9(6):e99732. [PubMed: 24923784]

Mackay TFC. Epistasis for quantitative traits in *Drosophila*. Methods Mol Biol. 2015; 1253:47–70. [PubMed: 25403527]

Mackay TF, Heinsohn SL, Lyman RF, Moehring AJ, Morgan TJ, Rollmann SM. Genetics and genomics of *Drosophila* mating behavior. Proc Natl Acad Sci U S A. 2005; 102(Suppl 1):6622–6629. [PubMed: 15851659]

Mackay TF, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, Richardson MF, Anholt RR, Barrón M, Bess C, Blankenburg KP, Carbone MA, Castellano D, Chaboub L, Duncan L, Harris Z, Javaid M, Jayaseelan JC, Jhangiani SN, Jordan KW, Lara F, Lawrence F, Lee SL, Librado P, Linheiro RS, Lyman RF, Mackey AJ, Munidasa M, Muzny DM, Nazareth L, Newsham I, Perales L, Pu LL, Qu C, Ràmia M, Reid JG, Rollmann SM, Rozas J, Saada N, Turlapati L, Worley KC, Wu YQ, Yamamoto A, Zhu Y, Bergman CM, Thornton KR, Mittelman D, Gibbs RA. The *Drosophila melanogaster* Genetic Reference Panel. Nature. 2012; 482(7384):173–178. [PubMed: 22318601]

Malmendal A, Sørensen JG, Overgaard J, Holmstrup M, Nielsen NC, Loeschcke V. Metabolomic analysis of the selection response of *Drosophila melanogaster* to environmental stress: are there links to gene expression and phenotypic traits? Naturwissenschaften. 2013; 100(5):417–427. [PubMed: 23571708]

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011; 17(1):10–12.

Massouras A, Waszak SM, Albarca-Aguilera M, Hens K, Holcombe W, Ayroles JF, Dermitzakis ET, Stone EA, Jensen JD, Mackay TF, Deplancke B. Genomic variation and its impact on gene expression in *Drosophila melanogaster*. PLoS Genet. 2012; 8(11):e1003055. [PubMed: 23189034]

McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012; 40(10):4288–4297. [PubMed: 22287627]

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–1303. [PubMed: 20644199]

Melcher C, Bader R, Pankratz MJ. Amino acids, taste circuits, and feeding behavior in *Drosophila*: towards understanding the psychology of feeding in flies and man. J Endocrinol. 2007; 192(3): 467–472. [PubMed: 17332516]

Morozova TV, Anholt RR, Mackay TF. Phenotypic and transcriptional response to selection for alcohol sensitivity in *Drosophila melanogaster*. Genome Biol. 2007; 8(10):R231. [PubMed: 17973985]

Morozova TV, Mackay TF, Anholt RR. Transcriptional networks for alcohol sensitivity in *Drosophila melanogaster*. Genetics. 2011; 187(4):1193–1205. [PubMed: 21270389]

Mulligan MK, Ponomarev I, Hitzemann RJ, Belknap JK, Tabakoff B, Harris RA, Crabbe JC, Blednov YA, Grahame NJ, Phillips TJ, Finn DA, Hoffman PL, Iyer VR, Koob GF, Bergeson SE. Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. Proc Natl Acad Sci U S A. 2006; 103(16):6368–6373. [PubMed: 16618939]

éNaja F, Hwalla N, Itani L, Salem M, Azar ST, Zeidan MN, Nasreddine L. Dietary patterns and odds of Type 2 diabetes in Beirut, Lebanon: a case-control study. Nutr Metab (Lond). 2012; 9(1):111. [PubMed: 23270372]

Neckameyer WS. A trophic role for serotonin in the development of a simple feeding circuit. Dev Neurosci. 2010; 32(3):217–237. [PubMed: 20714156]

Nirala NK, Rahman M, Walls SM, Singh A, Zhu LJ, Bamba T, Fukusaki E, Srideshikan SM, Harris GL, Ip YT, Bodmer R, Acharya UR. Survival response to increased ceramide involves metabolic adaptation through novel regulators of glycolysis and lipolysis. PLoS Genet. 2013; 9(6):e1003556. [PubMed: 23818862]

Nässel DR, Kubrak OI, Liu Y, Luo J, Lushchak OV. Factors that regulate insulin producing cells and their output in *Drosophila*. Front Physiol. 2013; 4:252. [PubMed: 24062693]

Norga KK, Gurganus MC, Dilda CL, Yamamoto A, Lyman RF, Patel PH, Rubin GM, Hoskins RA, Mackay TF, Bellen HJ. Quantitative analysis of bristle number in Drosophila mutants identifies genes involved in neural development. Curr Biol. 2003; 13(16):1388–1396. [PubMed: 12932322]

Olszewski PK, Rozman J, Jacobsson JA, Rathkolb B, Strömberg S, Hans W, Klockars A, Alsiö J, Risérus U, Becker L, Hölter SM, Elvert R, Ehrhardt N, Gailus-Durner V, Fuchs H, Fredriksson R, Wolf E, Klopstock T, Wurst W, Levine AS, Marcus C, de Angelis MH, Klingenspor M, Schiöth HB, Kilimann MW. Neurobeachin, a regulator of synaptic protein targeting, is associated with body fat mass and feeding behavior in mice and body-mass index in humans. PLoS Genet. 2012; 8(3):e1002568. [PubMed: 22438821]

Padmanabha D, Baker KD. *Drosophila* gains traction as a repurposed tool to investigate metabolism. Trends Endocrinol Metab. 2014; 25(10):518–527. [PubMed: 24768030]

Paik D, Jang YG, Lee YE, Lee YN, Yamamoto R, Gee HY, Yoo S, Bae E, Min KJ, Tatar M, Park JJ. Misexpression screen delineates novel genes controlling Drosophila lifespan. Mech Ageing Dev. 2012; 133(5):234–245. [PubMed: 22366109]

Pullen TJ, Rutter GA. When less is more: the forbidden fruits of gene repression in the adult β-cell. Diabetes Obes Metab. 2013; 15(6):503–512. [PubMed: 23121289]

Rajan A, Perrimon N. Of flies and men: insights on organismal metabolism from fruit flies. BMC Biol. 2013; 11:38. [PubMed: 23587196]

Reddiex AJ, Gosden TP, Bonduriansky R, Chenoweth SF. Sex-specific fitness consequences of nutrient intake and the evolvability of diet preferences. Am Nat. 2013; 182(1):91–102. [PubMed: 23778229]

Reed LK, Lee K, Zhang Z, Rashid L, Poe A, Hsieh B, Deighton N, Glassbrook N, Bodmer R, Gibson G. Systems genomics of metabolic phenotypes in wild-type *Drosophila melanogaster*. Genetics. 2014; 197(2):781–793. [PubMed: 24671769]

Reed LK, Williams S, Springston M, Brown J, Freeman K, DesRoches CE, Sokolowski MB, Gibson G. Genotype-by-diet interactions drive metabolic phenotype variation in *Drosophila melanogaster*. Genetics. 2010; 185(3):1009–1019. [PubMed: 20385784]

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26(1):139–140. [PubMed: 19910308]

Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010; 11(3):R25. [PubMed: 20196867]

Rovenko BM, Kubrak OI, Gospodaryov DV, Perkhulyn NV, Yurkevych IS, Sanz A, Lushchak OV, Lushchak VI. High sucrose consumption promotes obesity whereas its low consumption induces oxidative stress in *Drosophila melanogaster*. J Insect Physiol. 2015; 79:42–54. [PubMed: 26050918]

Schoeller DA. How accurate is self-reported dietary energy intake? Nutr Rev. 1990; 48(10):373–379. [PubMed: 2082216]

Schoeller DA. Limitations in the assessment of dietary energy intake by self-report. Metabolism. 1995; 44(2 Suppl 2):18–22.

Shamloula HK, Mbogho MP, Pimentel AC, Chrzanowska-Lightowlers ZM, Hyatt V, Okano H, Venkatesh TR. rugose (rg), a *Drosophila* A kinase anchor protein, is required for retinal pattern formation and interacts genetically with multiple signaling pathways. Genetics. 2002; 161(2):693–710. [PubMed: 12072466]

Shorter J, Couch C, Huang W, Carbone MA, Peiffer J, Anholt RR, Mackay TF. Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior. Proc Natl Acad Sci U S A. 2015; 112(27):E3555–E3563. [PubMed: 26100892]

Sieberts SK, Schadt EE. Moving toward a system genetics view of disease. Mamm Genome. 2007; 18(6–7):389–401. [PubMed: 17653589]

St. Pierre SE, Ponting L, Stefancsik R, McQuilton P. the FlyBase Consortium. FlyBase 102 - advanced approaches to interrogating FlyBase. Nucleic Acids Res. 2014; 42(Database issue):D780–D788. [PubMed: 24234449]

Strumia R. Eating disorders and the skin. Clin Dermatol. 2013; 31(1):80–85. [PubMed: 23245978]

Sørensen JG, Nielsen MM, Loeschcke V. Gene expression profile analysis of *Drosophila melanogaster* selected for resistance to environmental stressors. J Evol Biol. 2007; 20(4):1624–1636. [PubMed: 17584255]

Tabakoff B, Bhave SV, Hoffman PL. Selective breeding, quantitative trait locus analysis, and gene arrays identify candidate genes for complex drug-related behaviors. J Neurosci. 2003; 23(11): 4491–4498. [PubMed: 12805289]

Teleman AA, Hietakangas V, Sayadian AC, Cohen SM. Nutritional control of protein biosynthetic capacity by insulin via Myc in *Drosophila*. Cell Metab. 2008; 7(1):21–32. [PubMed: 18177722]

Telonis-Scott M, Hallas R, McKechnie SW, Wee CW, Hoffmann AA. Selection for cold resistance alters gene transcript levels in *Drosophila melanogaster*. J Insect Physiol. 2009; 55(6):549–555. [PubMed: 19232407]

Tennessen JM, Barry WE, Cox J, Thummel CS. Methods for studying metabolism in *Drosophila*. Methods. 2014; 68(1):105–115. [PubMed: 24631891]

Tennessen JM, Bertagnolli NM, Evans J, Sieber MH, Cox J, Thummel CS. Coordinated metabolic transitions during *Drosophila* embryogenesis and the onset of aerobic glycolysis. G3. 2014; 4(5): 839–850. [PubMed: 24622332]

Toma DP, White KP, Hirsch J, Greenspan RJ. Identification of genes involved in *Drosophila melanogaster* geotaxis, a complex behavioral trait. Nat Genet. 2002; 31(4):349–353. [PubMed: 12042820]

Vishalakshi C, Singh BN. Effect of directional selection for body size on fluctuating asymmetry in certain morphological traits in Drosophila ananassae. J Biosci. 2009; 34(2):275–285. [PubMed: 19550043]

Volders K, Scholz S, Slabbaert JR, Nagel AC, Verstreken P, Creemers JW, Callaerts P, Schwärzel M. Drosophila rugose is a functional homolog of mammalian Neurobeachin and affects synaptic architecture, brain morphology, and associative learning. J Neurosci. 2012; 32(43):15193–15204. [PubMed: 23100440]

Wu Q, Zhao Z, Shen P. Regulation of aversion to noxious food by Drosophila neuropeptide Y- and insulin-like systems. Nat Neurosci. 2005; 8(10):1350–1355. [PubMed: 16172603]

Yang CS, Thomenius MJ, Gan EC, Tang W, Freel CD, Merritt TJ, Nutt LK, Kornbluth S. Metabolic regulation of Drosophila apoptosis through inhibitory phosphorylation of Dronc. EMBO J. 2010; 29(18):3196–3207. [PubMed: 20700104]

Zhang P, Wu Y, Belenkaya TY, Lin X. SNX3 controls Wingless/Wnt secretion through regulating retromer-dependent recycling of Wntless. Cell Res. 2011; 21(12):1677–1690. [PubMed: 22041890]

Zhang W, Thompson BJ, Hietakangas V, Cohen SM. MAPK/ERK signaling regulates insulin sensitivity to control glucose metabolism in Drosophila. PLoS Genet. 2011; 7(12):e1002429. [PubMed: 22242005]

Zinke I, Kirchner C, Chao LC, Tetzlaff MT, Pankratz MJ. Suppression of food intake and growth by amino acids in *Drosophila*: the role of *pumpless*, a fat body expressed gene with homology to vertebrate glycine cleavage system. Development. 1999; 126(23):5275–5284. [PubMed: 10556053]

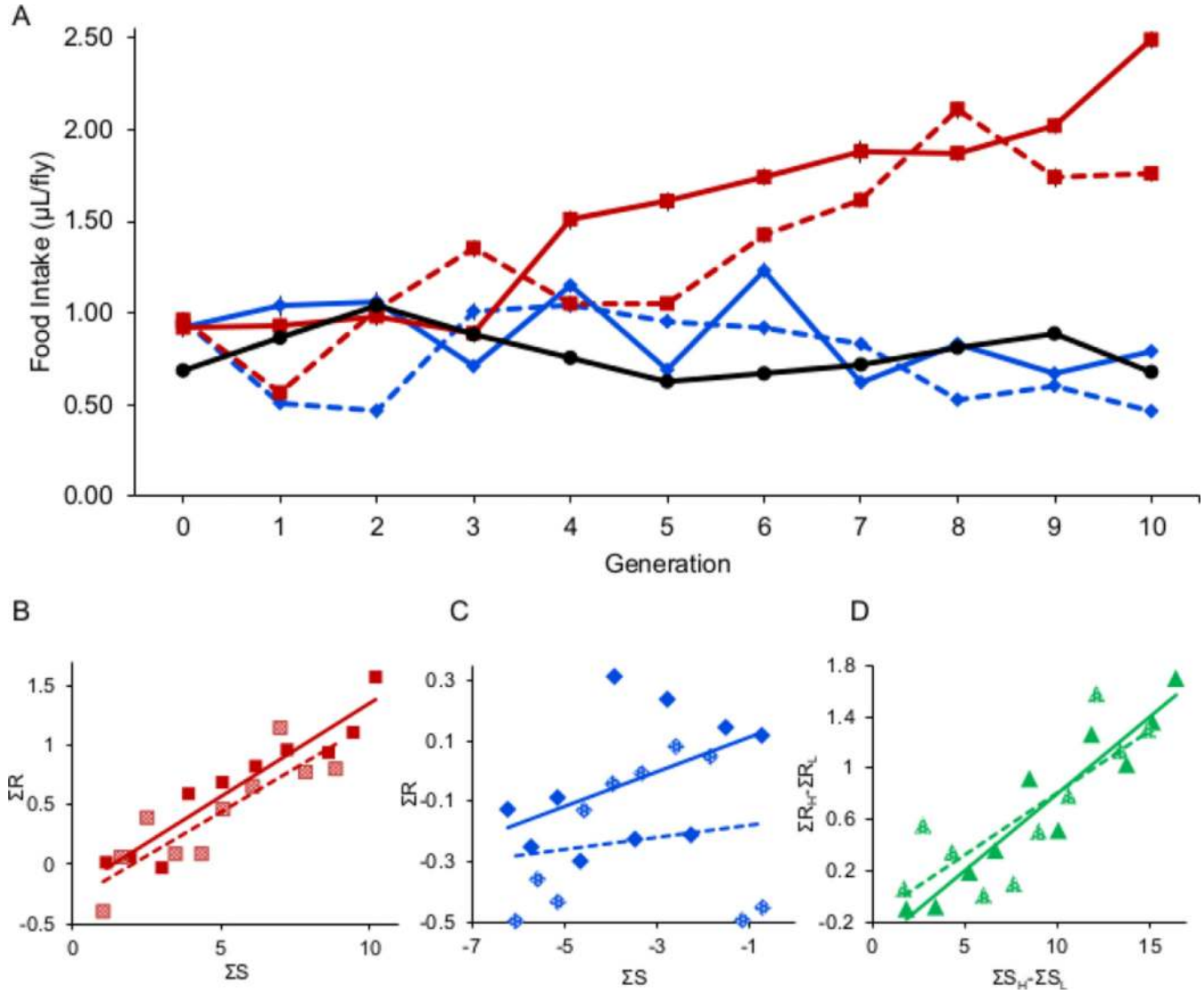

Figure 1. Response to selection for food consumption

(A) Mean food consumption of selection lines. Red squares indicate H lines, blue diamonds indicate L lines, black circles indicate the C line. Solid lines indicate Replicate 1 and dashed lines indicate Replicate 2. Error bars are ±SE. **(B)** Regressions of cumulative response on cumulative selection differential for high lines. The solid line and solid squares indicate Replicate 1, and the dashed line and patterned squares indicate Replicate 2. **(C)** Regressions of cumulative response onto cumulative selection differential for low lines. The solid line and solid diamonds indicate Replicate 1, and the dashed line and patterned diamonds indicate Replicate 2. **(D)** Regressions of high-low divergence of cumulative response on cumulative selection differential for Replicates 1 and 2. The solid line and solid triangles indicate Replicate 1, and the dashed line and patterned triangles indicate Replicate 2.

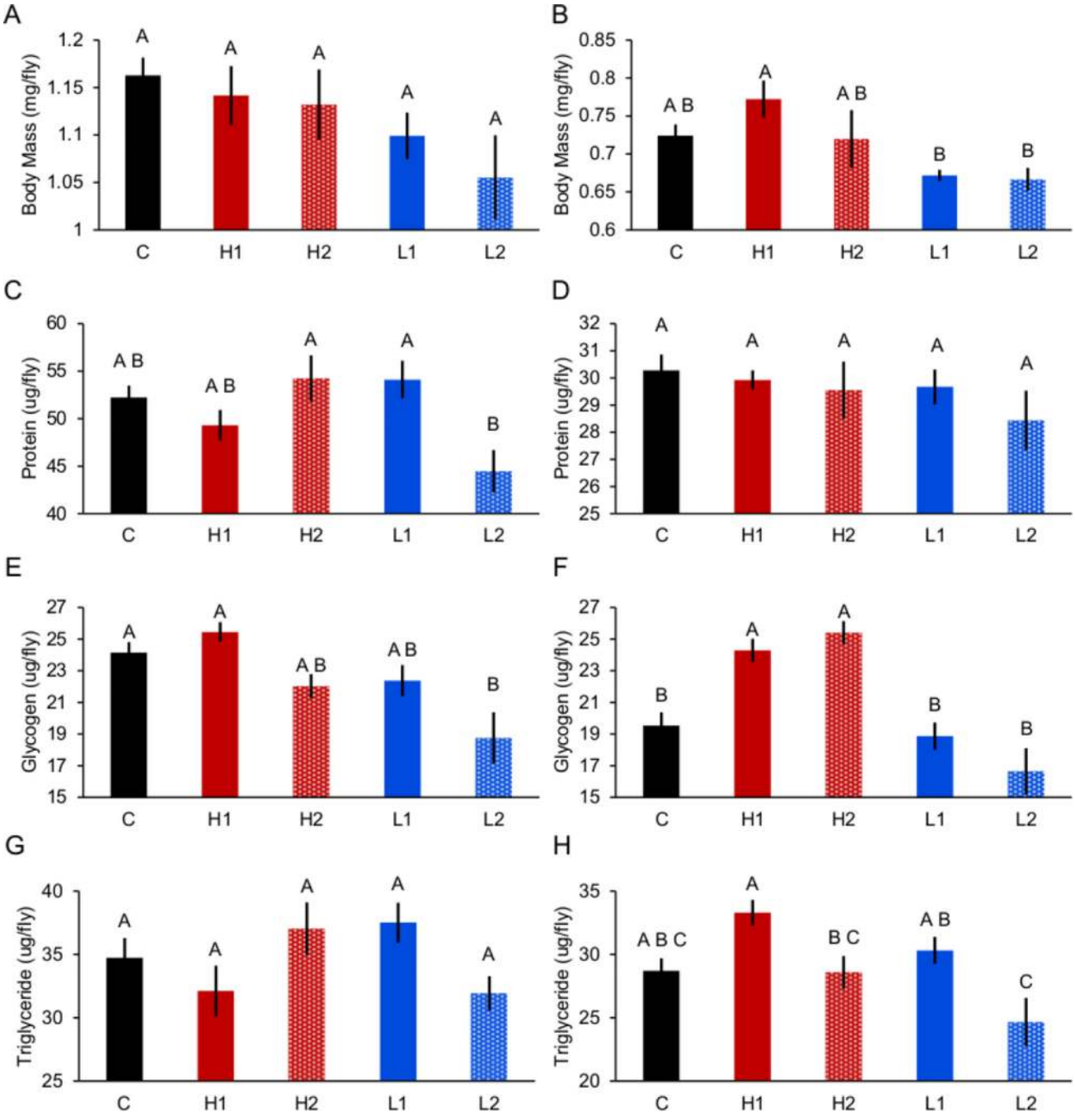

Figure 2. Body mass and molecular metabolites

All scores are pooled over G9 and G10. Red and blue bars denote H and L lines, respectively; the black bar is the C line. Solid bars indicate Replicate 1 and patterned bars indicate Replicate 2. Lines with the same letter are not significantly different from each other at $P < 0.05$. Error bars are ±SE. **(A)** Female body mass (mg/fly). **(B)** Male body mass (mg/fly). **(C)** Female total protein (μg/fly). **(D)** Male total protein (μg/fly). **(E)** Female glycogen (μg/fly). **(F)** Male glycogen (μg/fly). **(G)** Female triglyceride (μg/fly). **(H)** Male triglyceride (μg/fly).

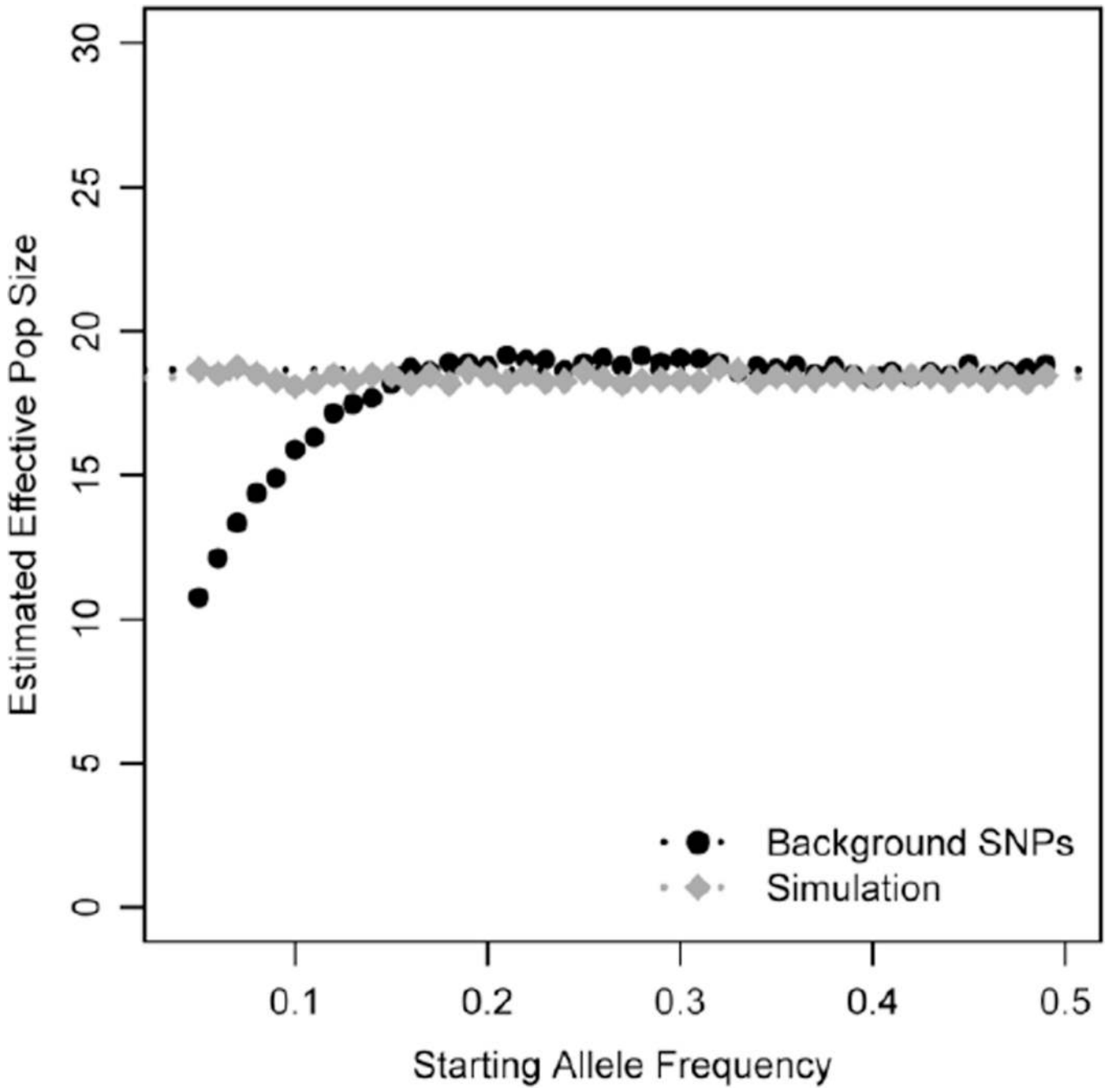

Figure 3. Comparison of effective population size estimated from observed SNP data and simulation

We ordered all segregating SNPs by their χ^2 *P*-value and selected 80% of SNPs with larger *P*-values to represent genomic background behavior in the absence of selection. We then binned these SNPs by their starting allele frequency in the base population and computed the effective population size based on the variance of allele frequency for these SNPs at generation 10 in both replicates. Black dots show the effective population size computed for each bin, and the dashed black line shows the median estimate across all bins. Grey dots show the effective population size based on the observed variance of allele frequencies in

our simulation with $N = 10$ males and 10 females for each starting allele frequency, and the dashed grey line shows the median of these estimates.

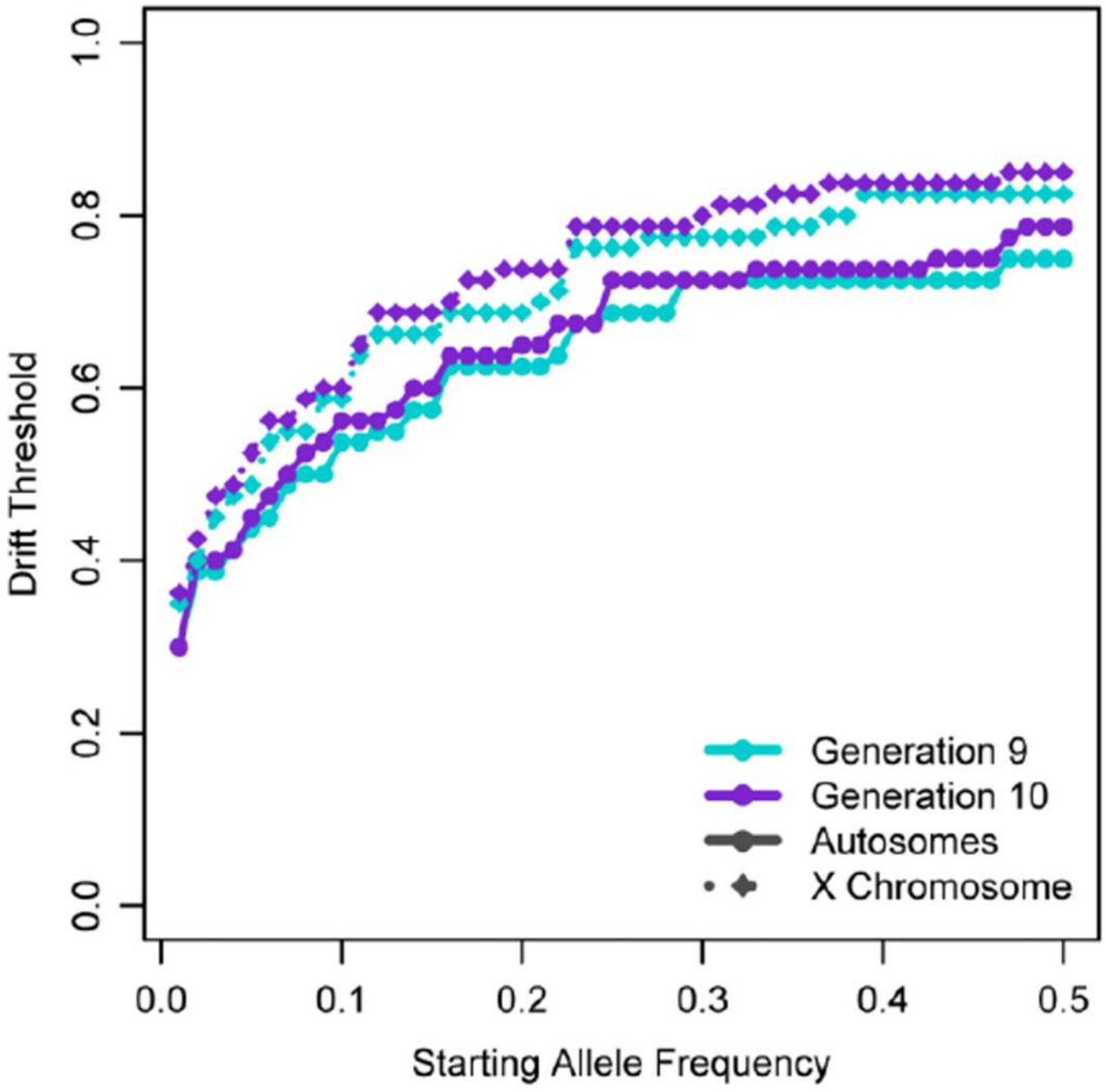

Figure 4. Upper bounds on absolute average frequency difference inferred from drift simulations

For each starting allele frequency, we performed 10,000 simulations of genetic drift in our experimental design for both autosomal and *X*-linked SNPs. For each simulation, we computed the absolute average difference between high and low lines, and used the 0.999 quantile of these values across all simulations as an upper bound on the value expected to occur in the absence of selection. Light blue dots/lines show the inferred upper bounds for each starting allele frequency at G9 for both autosomal (circles/solid lines) and *X*-linked SNPs (diamonds/dashed lines). Similarly, purple dots/lines show the inferred upper bounds for G10.

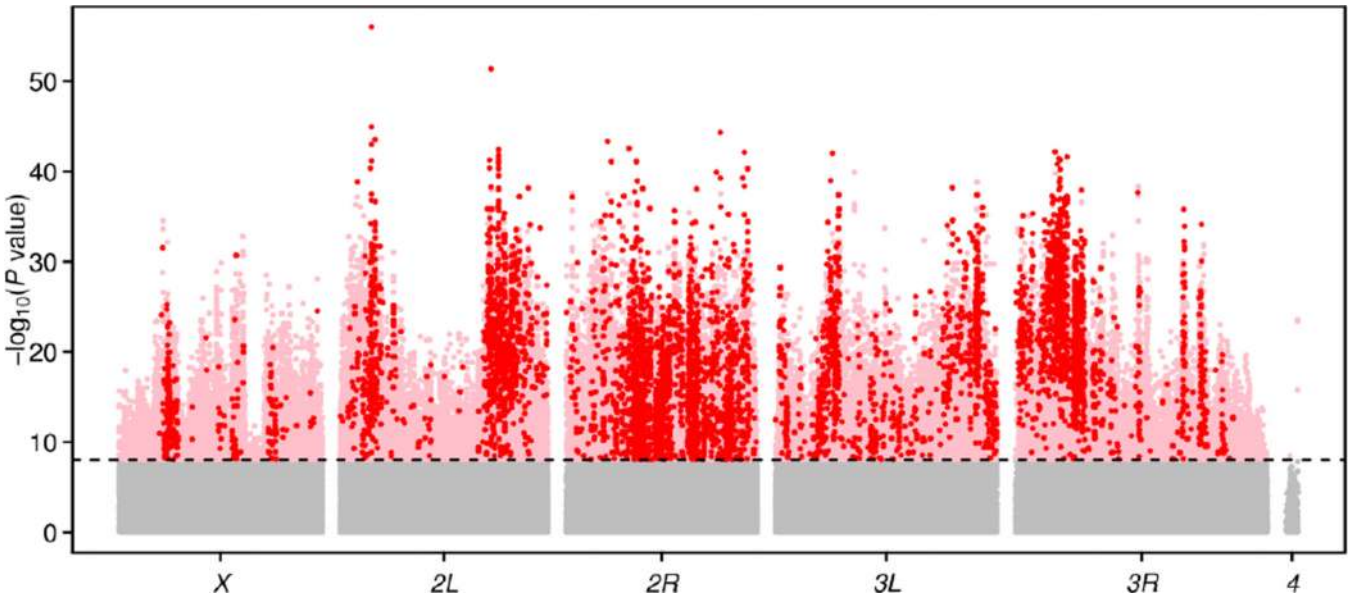

Figure 5. Genetic divergence of selection lines

Each dot represents a SNP, arranged by genomic coordinate on the *x*-axis and -log10(χ^2 *P*-value) on the *y*- axis. Grey dots indicate SNPs that are not significantly different between high and low lines averaged across replicates. Light red dots indicate SNPs that are significantly different between high and low lines, but may be the result of genetic drift based on comparison to drift simulations. Dark red dots indicate SNPs with significant differences in allele frequency between high and low lines averaged across replicates, for which the magnitude of the allele frequency difference exceeds that expected from drift alone.

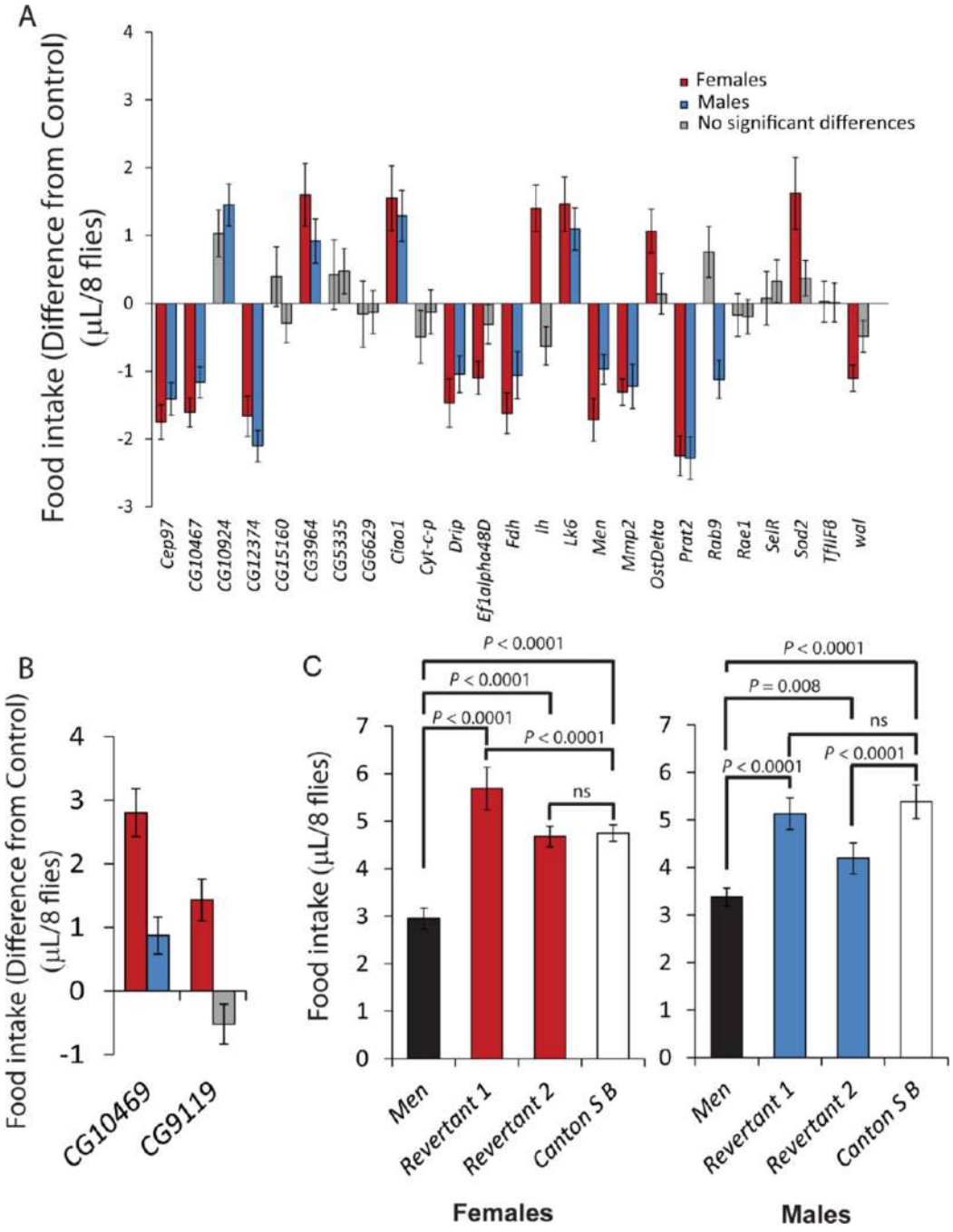

Figure 6. Functional validation of 27 candidate genes associated with mean food intake in selection lines

Grey bars: $P > 0.05$. Red Bars: Red bars: $P < 0.05$ in females. Blue bars: $P < 0.05$ in males. **(A)** RNAi suppression of gene expression. **(B)** *Mi{ET1}* element insertional mutations. **(C)** *P{GT1}* insertional mutation in *Men* and two precise revertant alleles. Error bars are ± SE.