Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in *Drosophila melanogaster*

Thomas L. Turner[1]*, Andrew D. Stewart[1], Andrew T. Fields[2], William R. Rice[1], Aaron M. Tarone[2]

1 Ecology, Evolution, and Marine Biology Department, University of California Santa Barbara, Santa Barbara, California, United States of America, 2 Department of Entomology, Texas A&M University, College Station, Texas, United States of America

Abstract

Body size is a classic quantitative trait with evolutionarily significant variation within many species. Locating the alleles responsible for this variation would help understand the maintenance of variation in body size in particular, as well as for quantitative traits in general. However, successful genome-wide association of genotype and phenotype may require very large sample sizes if alleles have low population frequencies or modest effects. As a complementary approach, we propose that population-based resequencing of experimentally evolved populations allows for considerable power to map functional variation. Here, we use this technique to investigate the genetic basis of natural variation in body size in *Drosophila melanogaster*. Significant differentiation of hundreds of loci in replicate selection populations supports the hypothesis that the genetic basis of body size variation is very polygenic in *D. melanogaster*. Significantly differentiated variants are limited to single genes at some loci, allowing precise hypotheses to be formed regarding causal polymorphisms, while other significant regions are large and contain many genes. By using significantly associated polymorphisms as a priori candidates in follow-up studies, these data are expected to provide considerable power to determine the genetic basis of natural variation in body size.

Introduction

Body size is a critical phenotype for all organisms. In *Drosophila* species, considerable evidence from natural populations indicates size is under spatially varying selection, with larger size selected in higher latitudes [1–4]. Body size is also sexually dimorphic in *D. melanogaster*, yet genetic variation for size is correlated between the sexes, and differences in optimal size have been shown to result in sexual conflict [5]. As size can be correlated with development time [6,7], life span [8,9], sexual attractiveness [10], fecundity [10–12] and other traits [13], natural variation in this complex trait is of considerable interest. Determining which genetic polymorphisms affect body size would elucidate the joint distributions of effect sizes, population frequencies, and network position—the key to understanding how biological systems accommodate change while maintaining performance. It would also allow molecular dissection of trait correlations, sexual correlations, and interactions between genes and the environment, all of which may be crucial factors in maintaining genetic variation in natural populations. However, genome-wide association studies of human variation, even when conducted with hundreds of thousands of individuals, have only explained a modest fraction of heritable variation for height [14], probably due to small effect sizes and/or low population frequencies [14–16]. In model organisms, where strict environmental controls and repeated measures of identical genotypes may reduce non-heritable variation, power is likely to increase substantially [17], but adequate power to comprehensively map functional variation will likely remain challenging. Because of the short generation time and ease of culture of *D. melanogaster*, a complementary approach may be to couple experimental evolution with population-based sequencing of evolved populations.

For over 100 years, experimentally evolved populations of *D. melanogaster* have been used to address fundamental questions in population genetics [18,19], with experiments on body size from at least 1952 [20]. More recently, experimental evolution has been combined with complete genome sequencing to link genotype and phenotype in viral and microbial systems [21–23]. In these systems, evolution generally begins with a single genotype, and adaptation occurs after mutation produces genetic variation. This approach has yielded many insights into both the structure of molecular networks and the adaptive process. By combining experimental evolution and population-based sequencing in *D. melanogaster*, it may be possible to investigate the genetic basis of natural population variation, which has been a primary goal of population genetics since molecular characterization revealed the extent to which genomes vary in natural populations. Partial support for this proposal comes from a previous study which used microarray-based genotyping on individuals derived from populations selected for enhanced stress resistance [24], and a recent

Author Summary

Understanding the causes and consequences of natural genetic variation is crucial to the characterization of biological evolution. Moreover, natural genetic variation is comprised of millions of perturbations, which are partially randomized across genotypes such that a small number of individuals can be used to combinatorially analyze a large number of differences, facilitating mechanistic understanding of biological systems. Here we demonstrate a powerful technique to parse genomic variation using artificial selection. By selecting replicate populations of *Drosophila* flies to become bigger and smaller, and then determining the evolutionary response at the genomic level, we have mapped hundreds of genes which respond to selection on body size. As our approach is powerful and cost effective compared to existing approaches, we expect it to be a major component of diverse future efforts.

study on selected lineages of domesticated chickens [25]. Additionally, Burke *et al.* recently resequenced populations selected for divergent generation times [26]. Though Burke *et al.* report little evidence for canonical "selective sweeps" on newly arising or rare causal variants, they do not attempt to estimate the locations or number of causal alleles. Moreover, the history of the populations used potentially complicates these observations: before they were selected to have long and short generation time, the ancestors of all populations were selected to have long generation time, which may have biased later adaptive divergence towards alleles which have small enough effects to remain polymorphic during this initial period.

Here we further explore this approach, using populations of *D. melanogaster* derived from the outbred, lab-adapted LH_M population. This population was originally derived from a large collection of flies from California, and has been maintained under a precise and stable regime for over 400 generations [27]. Although selection related to environmental variation has been minimized in this population, individuals compete for a limited amount of food and mates each generation, and variation in many traits, including fitness, is abundant [28]. The factors that maintain variation in the face of drift and selection in this lab population are likely to be a subset of factors which maintain variation in populations in the wild.

Results

Selection for body size

To experimentally investigate the "Evolve and Resequence" (hereafter E&R) approach to genetic mapping, six populations were established: two were selected for large size, two for small size, and two were subjected to identical protocols, but not selected based on size (controls). A sieving apparatus was used to efficiently separate flies based on size: anesthetized flies were separated based on their ability to pass through a series of sequentially smaller sieves (see Methods). This allowed us to screen ~1800 flies per population, each generation, for over 100 generations. After 100 generations, the mean "sieve size" of the flies diverged substantially among the experimental populations (Figure 1, $F_{9,32} = 89.52$, $P = 0.0001$). Though a considerable response was seen in both directions, the response to selection was strongest among the small-selected lines. Indeed, by the end of the selection experiment many of the male flies (79% and 35%) in each small population passed through all 20 sequentially narrower sieves,

whereas no flies pass this far through the sieve system in the control populations. Anatomical measures of thorax, leg, and wing dimensions from each population verified considerable divergence in fly size (Figure 1 and Table S1). All anatomical measurements agree that populations selected for small size evolved substantially, while populations selected for large size changed more modestly and were significantly different from controls for only some traits. As large-selected populations evolved significantly in their ability to pass through the sieves, but have modest anatomical differences in the traits measured, this suggests that some of the response to selection is due to anatomical traits that were not directly measured, such as abdominal size.

Population-based resequencing of evolved populations

To simultaneously determine the locations and frequencies of genetic polymorphisms, we extracted DNA from 75 pooled females (2n = 150 chromosomes) for each population, and sequenced these populations with the Illumina Genome Analyzer. In total, we obtained 42.3 billion base pairs of sequence data, 99.8% of which aligned to the reference genome. After excluding the 23% of alignments with low mapping qualities, which includes non-unique alignments, each population had between 17-fold and 23-fold median coverage, with 87% to 93% of the genome having more than 10-fold coverage in each population (Table 1, Figure S1; this excludes the 4th (dot) chromosome, Y chromosome, mitochondria, centromeric heterochromatin and unmapped regions, which are shown in supplementary data but will not be considered further). Considering all 6 populations together (~120-fold genome coverage), over 27 million base pairs were found to be variable. However, the majority of these apparent polymorphisms are rare: 83.4% have overall frequencies less than 0.02. A considerable portion of these rare variants could be sequencing errors, which are difficult to completely exclude using pooled sequencing approaches. Mean error rates from the UNC-CH sequencing facility, where sequencing was conducted, are 0.5–3.0% depending on read position (C.D. Jones, pers. comm.), so apparent polymorphisms with >5% experiment-wide frequency should be true genetic variants. Even when only considering apparent polymorphisms with population frequencies >10% across the entire data set, 1.68 million bases are variable, verifying that there is considerable genetic variation in these populations. Although the large number of sequencing errors complicate some analyses by creating a large apparent excess of low frequency variation, these errors will be rare and randomly distributed, and are therefore not expected to be significantly differentiated between populations.

Polymorphism and differentiation

Differences in allele frequency between populations indicate that evolution has taken place, either due to stochastic forces (drift), selection, or both: this evolution is quantified in Figure 2. As expected, evolution occurred between the two control populations after they were separated from a common ancestor for over 100 generations. However, much more evolution has taken place between selection treatments than between control lines. In the two independent comparisons between a large- and small-selected line, 41,399 and 48,645 variants are >95% differentiated, compared to only 1,260 variants between controls (Figure 2). This considerable excess of highly differentiated variation indicates a substantial, genome-wide impact of artificial selection for body size. Additionally, of the 5587 variants that achieved this extreme level of differentiation in both comparisons, the vast majority (5537) changed frequency in the same direction, clearly implicating selection for body size.
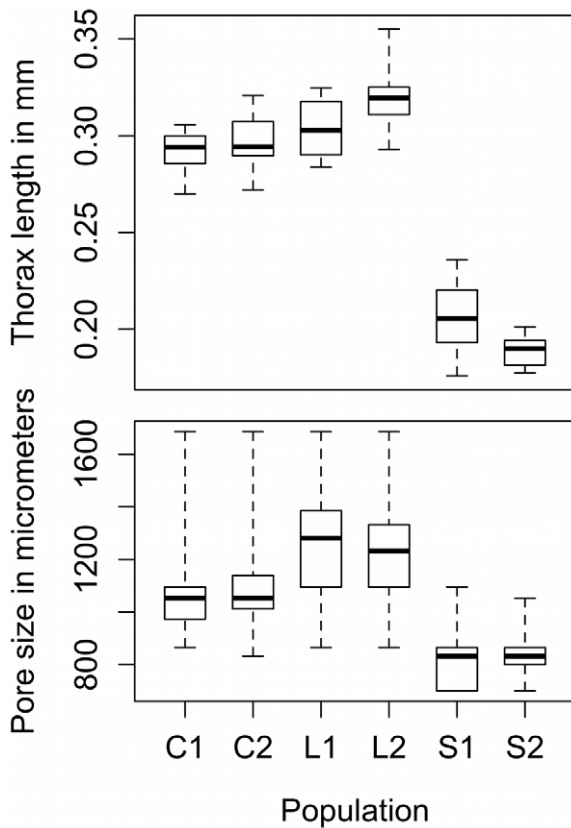
Figure 1. Response to selection. Above: Distribution of thorax length in each population after >100 generations of selection. Below: Distribution of sieve depth in each population. Median (line), 75% quartile (box), and range (whiskers) are shown. C1, C2 = controls, L1, L2 = large lines, S1, S2 = small lines.
doi:10.1371/journal.pgen.1001336.g001

To investigate the genome-wide impact of selection further, we determined the genomic distribution of heterozygosity in each population (only variants with an experiment-wide frequency
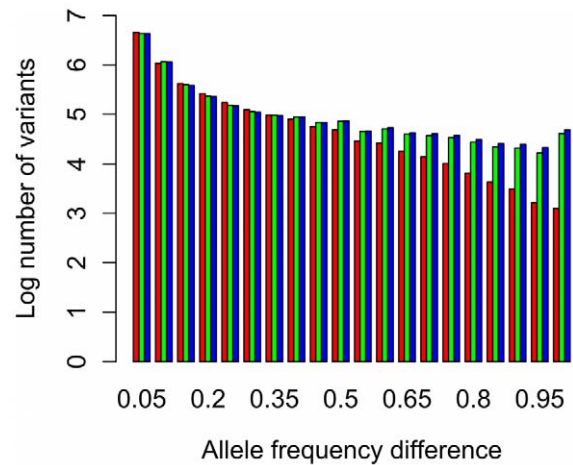
Table 1. Proportion of the euchromatic genome with a given fold coverage in each population.

	Population					
Fold coverage	**C1**	**C2**	**L1**	**L2**	**S1**	**S2**
0	0.05	0.05	0.05	0.05	0.05	0.05
1–5	0.02	0.01	0.01	0.01	0.01	0.01
6–10	0.08	0.02	0.02	0.04	0.06	0.05
11–15	0.18	0.08	0.07	0.12	0.16	0.15
15–20	0.20	0.16	0.15	0.19	0.22	0.22
21–25	0.17	0.19	0.18	0.19	0.19	0.19
26–30	0.12	0.17	0.16	0.15	0.13	0.13
31–35	0.07	0.12	0.12	0.10	0.08	0.08
36–40	0.05	0.08	0.09	0.06	0.05	0.05
41–45	0.03	0.05	0.06	0.04	0.03	0.03
46–50	0.02	0.03	0.04	0.02	0.02	0.01
51+	0.02	0.04	0.05	0.03	0.02	0.01

doi:10.1371/journal.pgen.1001336.t001

Figure 2. Frequency histogram of differentiation between populations, on a log scale. Pair-wise comparisons are shown between the two control populations, and between the two independent comparisons of a large- and small-selected population. Red = control population 1 versus control population 2; green = large population 1 versus small population 1; blue = large population 2 versus small population 2.
doi:10.1371/journal.pgen.1001336.g002

>0.10 were used in order to avoid sequencing errors). Median heterozygosity in 10-kb windows was 0.0031 and 0.0032 for control 1 and control 2 populations, respectively. Average heterozygosity in each of the two large populations was only slightly, but significantly, less than the average of the controls: 0.0031 (t-test $P = 0.048$) and 0.0030 ($P \ll 0.001$). The two small populations, however, have a 55% reduction in median heterozygosity compared to controls (0.00180 and 0.00175; $P \ll 0.001$). Fewer than 10 windows have near zero (<0.0001) heterozygosity in each of the large and control populations, while 60 and 85 windows are depleted of variation in each of the small populations. As this is still less than 1% of the genome, variation persists in most genomic regions. The number of breeding adults of all six populations were identical and remained constant throughout the 100 generations of the study, so this reduction is clearly due to the action of selection.

To locate a set of variants that have evolved due to selection for body size, we considered data from the two large- and two small-selected lines together. Only variants where the large-selected lines both had higher or lower allele frequencies than both the small-selected lines were considered high-confidence candidates for selection. This immediately excludes 66% of allele frequency changes due to drift or selection unrelated to experimental treatment, as well as eliminating variants that were affected by linked selection in inconsistent directions. For the 1,886,104 variants meeting this criterion, the minimum allele frequency difference (diffStat) between the four possible large-small comparisons was used as a composite statistic (Figure 3). Most (74%) variants had diffStats <0.10, with just over 4% (76,719) of variants having diffStats >0.50. To determine a set of loci that are likely to have evolved under selection for body size, we compared the distribution of this statistic to simulations of drift alone. The amount of allele frequency change expected due to drift depends on starting allele frequency, which we estimated for each variant as the average frequency of the two control populations (see Methods). For each allele frequency class, we then simulated drift for 110 generations, and simulated sampling error due to sequencing coverage (see Methods). Using these simulations, an
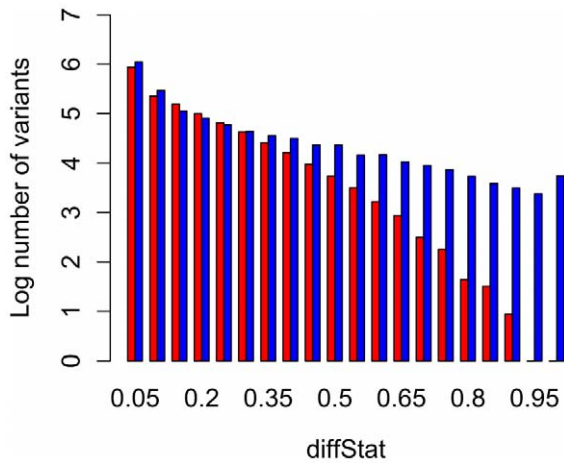
Figure 3. Frequency histogram of differentiation between treatments, on a log scale. The distribution of the diffStat is shown, when the two large- and two small-selected populations are considered together; this is compared to the expected distribution obtained via simulation. Blue = observed, red = simulated drift. Note the log scale.
doi:10.1371/journal.pgen.1001336.g003

expected distribution of diffStat was generated in the absence of selection. By comparing the observed and expected distributions of the statistic, we estimated false discovery rate (FDR) thresholds for evolution due to selection for body size (Tables S2, S3). This FDR calculation implicitly assumes that the number of effective loci in the genome is approximated by the number of polymorphisms. It does not assume that all polymorphisms are independent, however varying levels of linkage disequilibrium inflate the variance around

the mean FDR. For this and many other reasons, these calculations should be taken as approximations of the actual FDR. As experimental populations are quite different from control comparisons (Figure 2), the observed and expected distributions are radically different (Figure 3), and differentiation occurs at many loci across the genome (see below), this approximation seems acceptable.

Differentiated loci and gene annotations

Significantly differentiated variants are clearly distributed non-randomly across the genome: their distribution on chromosome arm 3L is shown in Figure 4, and other chromosomes are shown in Figure S2. Differentiation exhibits a distinct peaked pattern in many regions, with a wide range of peak sizes. Only a few variants are differentiated at some loci, providing precise hypotheses regarding functional targets (Figure 5). In other regions, especially surrounding the chromosome 2 centromere, significant differentiation is spread across megabases (Figure S2). A very large number of differentiated peaks are apparent when chromosomes are examined at a fine scale. Regardless of the precise number of differentiated peaks, it is difficult to estimate the number of alleles that were selected. For example, within each peak, it is possible that a single haplotype bearing the combination of multiple causal variants with the largest combined effect was selected. However, to determine a very conservative minimum estimate, we counted the number of regions containing significant variants (FDR<10%) that were separated from all other significant variants by a minimum distance of 10-kb. Based on this measure, 1236 distinct regions have evolved due to selection on body size—even if the minimum separation distance is extended to 50 kb, 304 distinct regions are observed. As the true number of selected variants is likely much higher, this indicates that the response to selection was
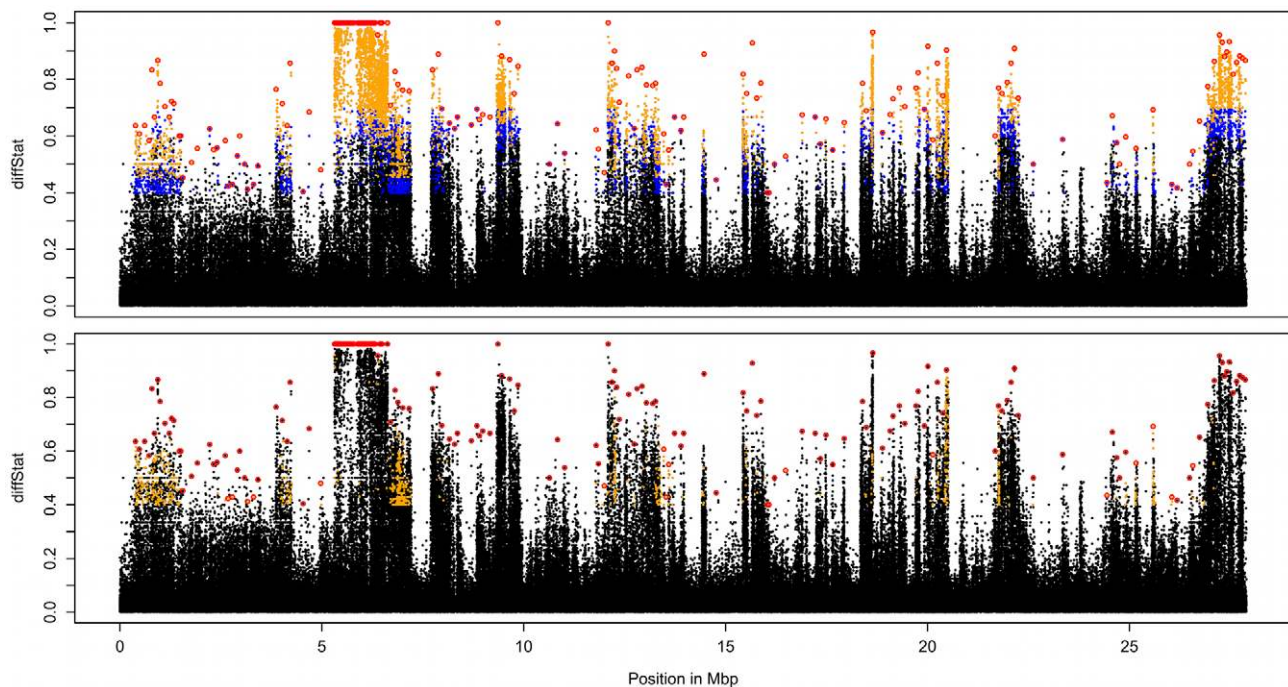
Figure 4. Differentiation on chromosome arm 3R. The diffStat is shown for each variant that had higher or lower allele frequencies in the large-selected lines compared to the small-selected lines. Above: Color coding indicates significance: black = nonsignificant variants, blue = significant variants at the permissive FDR threshold (FDR<10%); gold = significant variants at the restrictive FDR threshold (FDR<5%); red = peak variants. Below: Color coding indicates estimated starting allele frequency: black = all variants, gold = variants with an average control frequency <0.05; red circles indicate peak variants, as in A.
doi:10.1371/journal.pgen.1001336.g004

Figure 5. Fine-scale mapping of candidate causal variants. The diffStat is shown for each variant at each locus; black = nonsignificant variants, blue = FDR<10%; purple = FDR<5%; red = peak variants. Above: Eip63E; Below: dre4. For the candidate g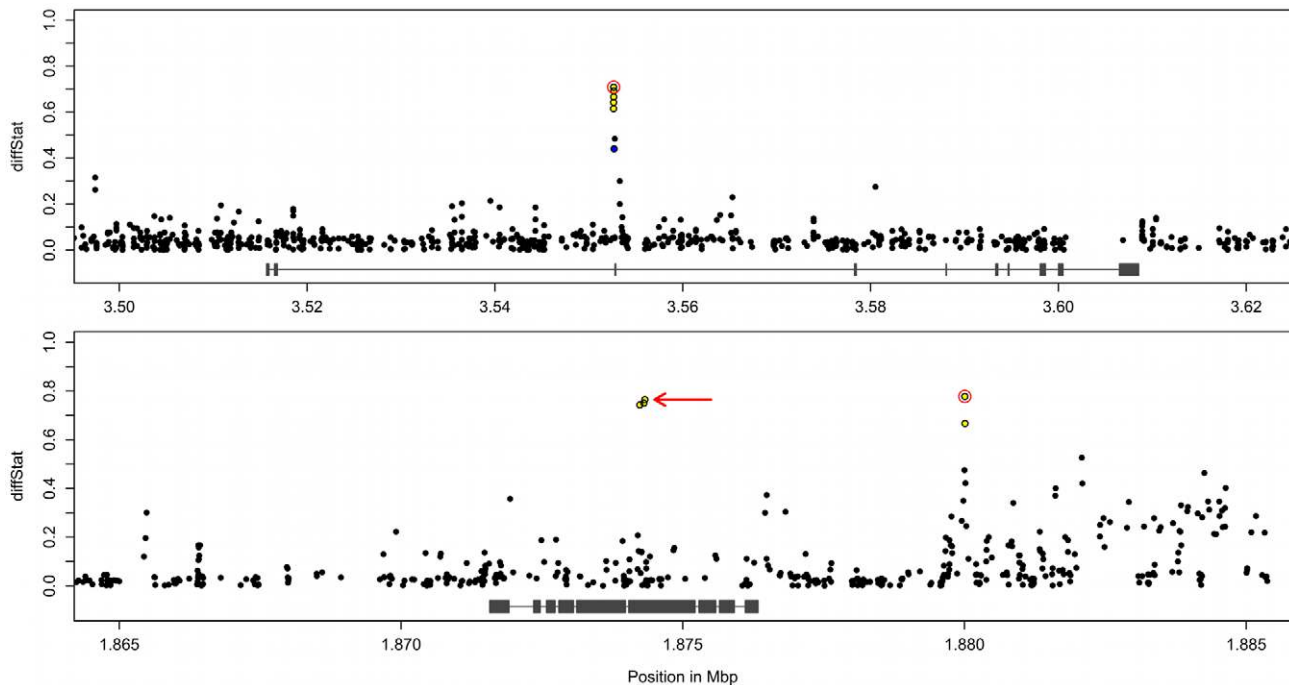ene at each locus, the exons are shown as linked grey boxes; only one transcript is shown for simplicity. The arrow in B indicates a serine-tryptophan replacement discussed in the text.
doi:10.1371/journal.pgen.1001336.g005

startlingly polygenic, with certainly hundreds, and likely thousands, of target loci.

Using the average frequency of each variant in the two control populations, we can obtain a rough estimate of the starting frequency of each differentiated polymorphism. As shown in Figure 4, the proportion of initially uncommon variants that became significantly differentiated varies greatly between peaks (other chromosomes are shown in Figure S3). In some peaks, not even one differentiated variant has an average frequency <0.05 in the control populations, while other peaks have many such variants.

When using any correlative analysis (e.g. genome-wide association studies, QTL mapping, or E&R), it is very challenging to differentiate causal alleles from linked variants. For example, simulations of genome-wide association studies indicate that non-causal alleles can be more significant than causal alleles when the non-causal alleles are in linkage disequilibrium with multiple causal variants [29]. Despite these caveats, we hypothesized that some variants with the local maximum diffStat values would be likely to either effect body size themselves, or be in close proximity to variants that do. To delimit a set of such variants, we centered a 100-kb window on each significant variant. As the structure of linkage disequilibrium is unknown in these populations, the choice of 100-kb is somewhat arbitrary, but is expected to be much larger than the normal extent of linkage disequilibrium across most of the genome, and is therefore conservative [30]. If the diffStat value of the variant in question was larger than or equal to the maximum within this window, it was considered a "peak variant" (that is, it was a local maximum; Figure 4). This method results in 5205 peak variants, 3572 of which lie in a 10-Mb region surrounding the chromosome 2 centromere (2L>18 Mb and 2R<5 Mb). Of the 1633 peak variants outside this region, less than 10% have estimated starting frequencies less than 0.05; in contrast, 41% of the 3572 variants in the region surrounding the centromere started

at frequencies below 0.05. Heterozygosity in this region is very low in the small-selected populations (median<0.0001), compared to the same region in the other four populations (0.0027–0.0030), or the rest of the chromosome in the small-selected populations (0.0024–0.0025). Together, these results implicate one or more major selective sweeps in this region in the small-selected populations, which fixed a large number of rare variants and eliminated variation surrounding the centromere.

In regions with distinctly differentiated peaks, we hypothesize that peak variants are near the direct targets of selection. As a partial test of this hypothesis, we assembled a list of genes at these loci. The 10-Mb region surrounding the chromosome 2 centromere was excluded due to the large number of fixed differences throughout this region. For the remaining 1633 peak variants, 632 genes either overlap the peak variant or are within 1-kb. Functional annotations of these loci were compared to the complete genome using annotations from FlyBase [31] and the Database for Annotation, Visualization, and Discovery (DAVID), which uses fuzzy clustering to group genes into functionally related classes based on the similarity of their annotations [32,33]. The most over-represented cluster of biological processes (GO terms) includes genes affecting post-embryonic development and metamorphosis, with post-embryonic development also the most significantly over-represented biological process individually ($P = 8.64\text{E}^{-7}$; Bonferroni-adjusted $P = 0.001$; see Datasets S1 and S2 for full results). As all anatomical features measured have changed between treatments, and the timing of metamorphosis is likely to alter adult size, these functions correspond precisely to phenotypic characterizations. This functional cluster includes genes such as ecdysone-induced proteins (*l(3)82Fd*, *Eip63E*, *Eip74EF*, *Eip75B*), many genes involved in anatomical development (*vein*, *plexus*, *headcase*, *blistery*, etc.) and others. The second most over-represented gene cluster was found to include the biological processes cell morphogenesis (cell size and shape): cell number and

cell size are both known to change with body size in *Drosophila* [7,34], so genes with this annotation are excellent candidates for harboring natural variation for these traits. Over 40 genes which are known to affect cell morphogenesis are near a peak variant ($P = 3.46E^{-6}$; Bonferroni-adjusted $P = 0.006$), including genes involved in epidermal growth factor signaling (including the *epidermal growth factor receptor*), the salvador/warts/hippo pathway (*salvador*, *crumbs*) [35–37], and many others including *E2F*, *dally*, *dally-like*, *knirps*, and *miniature*.

When the 3572 peak variants surrounding the chromosome 2 centromere are included, these categories remain the two most over-represented categories of biological processes. The large number of genes in this region precludes confident assessment of the targets of selection, but it is notable that the *ecdysone receptor* is found here, and contains fixed differences in multiple introns, exons, and both 5′ and 3′ UTRs. Finally, as an additional control, we determined if any of the biological process mentioned above were over-represented near variants which had >90% allele frequency difference between controls. Of the biological processes mentioned above, post-embryonic development was the most significant between control populations ($P = 0.002$), and was not significant after Bonferroni correction ($P = 0.74$). This is in stark contrast to the overrepresentation between treatments: cell morphogenesis, post-embryonic development and metamorphosis all show greater than 2-fold enrichment, and are all significant after Bonferroni correction. For a complete list of the genes near peak variants, and a list of the variants themselves, see Datasets S3 and S4.

Some genes are notable for their absence from this list: although insulin regulation is crucial for body size determination, many of the canonical genes in the insulin pathway are not near a peak variant. However, some of these genes, including *Tor*, *slimfast*, and *glut1* [38] overlap significant (FDR<10%) variants. Because these variants are within 50 kb of a more significant variant, they are not included in the peak variants list. This may indicate that natural variants at these loci have smaller effects than many other genes, and this hypothesis could be tested by including the significant variants in a follow-up association study. Even more remarkably, the adaptor protein *chico* and the *insulin receptor* itself are far from any polymorphisms which responded to selection in this regime.

Discussion

By resequencing experimentally evolved populations, the genomic impact of selection on allele frequency is shown to be considerable. Nearly all variants which are >95% differentiated between the two independent comparisons of a large- and small-selected population are differentiated in the same direction, supporting the assertion that these variants have changed due to selection on body size. Heterozygosity is reduced in the selected populations, especially in the small populations where the most phenotypic change has occurred, consistent with the expected effect of linked selection [39]. It is clear that at least hundreds, and likely thousands, of polymorphisms affect body size in this long-term laboratory population. Though some of these alleles may have arisen through mutation after the founding of this population from nature, most are expected to be variants that affect body size in natural populations. In any case, it is clear that the response to selection was due evolution at many loci throughout the genome. A recent meta-analysis of human height variation provides perspective on this number: by genotyping over 180,000 individuals, 180 loci affecting human height were located, but these loci together explain only 10% of the phenotypic variation [14]. Body size in *D. melanogaster* would appear to be similarly

polygenic, and our study demonstrates that these loci can be mapped with much less genotyping effort and expense than human GWA studies. Of course, mapping these loci provides only the initial step towards comprehensive characterization of this variation. For example, the relative effect sizes of these loci are not yet known, and it is possible that some loci are selected because they are compensating for the pleiotropic side effects of body size change. These challenges can be addressed either using association studies or further artificial selection, but in either case the loci mapped here can be used as *a priori* candidates, tremendously increasing power while decreasing the effort required.

The proportion of trait variation due to common versus rare variants is of great interest. Interestingly, differentiation of uncommon (<5% frequency) variants occurred at only some peaks (Figure 5, Figure S3), which may indicate that these peaks are caused by selective sweeps on rare (or at least uncommon) variation. At other peaks, no uncommon variation is differentiated, which is inconsistent with selective sweeps of haplotypes from low to high frequency. It is tempting to conclude that selection response in these regions was due to common variants, but this conclusion is tenuous. For example, if an uncommon combination of common mutations was selected in some regions, rare variants might hitchhike to high frequency on this uncommon haplotype. Conversely, peaks where only common mutations are differentiated could theoretically be created by slight changes at a large number of linked rare variants. We consider this second possibility to be unlikely, as the effective number of alleles per locus is likely modest in *D. melanogaster*, due to the short range of linkage disequilibrium in this species [30]. Consequently, these data are consistent with much of the standing variation in body size being due to common variants, but definitive conclusions regarding the relative importance of common and rare variants await follow-up studies determining if the most differentiated variant in each peak is indeed causal.

Though the most differentiated variant in each peak is not assured to be under direct selection, it is the most probable variant in each peak. Indeed, some peaks have only one or a few significant variants, allowing confident hypotheses to be formed about the specific causal variants, and in many others there is a single variant which is much more differentiated than all others nearby (Figure 5). The correlation between significantly differentiated "peak variants" and gene functions expected to effect body size is striking, and supports the assertion that these variants may be near the direct targets of selection. This correlation implies that many genes known to be involved in anatomical development, metamorphosis, cell number, and cell size harbor natural genetic variants affecting these same traits. However, we were also able to exclude the involvement of several loci, including *chico* and *insulin receptor* (*InR*), which have key roles in body size and development when experimentally manipulated [40,41]. Consequently, not all loci which posses the capacity to affect traits actually harbor functional variants in any given population.

As body size varies clinally with latitude, and variation at *InR* is also clinal [42], this gene is considered a candidate for adaptively affecting body size. This is potentially compatible with our results: if variation in *InR* is maintained by spatially varying selection, this variation may have been lost when this selection was removed by founding the stable LH$_M$ population. Other clinal genes, however, may have been selected in our experiment: recently, genotypic and expression variation at the *dca* gene (*Drosophila cold acclimation*, a.k.a. *Senescence marker protein-30*) was shown to associate with wing size in Australian populations [43]. In our experiment, several deletions in the 3′ UTR of *dca* changed under selection, providing a precise hypothesis for the location of functional variation at this locus. The

most differentiated of these deletions was present at frequencies of 1.0 in both large populations, and 0.0 and 0.5 in each small population, resulting in a diffStat of 0.50 (with an estimated FDR of 12.9%, this variant was not considered significant in the genomic analysis).

The greatest strength of the E&R approach may be the possibility to refine annotations at genes expected to influence a phenotype by identifying specific sub-genic functional elements, as illustrated by the *dca* example above. As an additional example, consider *Ecdysone-induced protein 63E*: this is a complex gene, with 13 alternative transcripts spanning nearly 95 kb. Deletions at this locus are generally lethal, but larvae that survive to pupation form very small pupae [44]. In response to selection for body size, only 4 SNPs and a 3-bp deletion became significantly differentiated at this locus, and all are within a 100-bp region in a single intron (FDR<0.006; Figure 5). Functional characterization of this small region may lead to insights regarding ecdysone-regulated size determination. Similarly, only 3 SNPs are differentiated in the gene *dre4* (FDR<0.00002; Figure 5). The product of this gene (also known as SPT16) forms a heterodimer known as FACT (with SSRP1) that is involved in chromatin remodeling in *Drosophila* and conserved throughout eukaryotes [45,46]. Loss-of-function mutations at this gene dramatically reduce ecdysteroid production at ecdysone regulated developmental stages, preventing molting: this gene is therefore an excellent candidate for altering critical size at metamorphosis through ecdysone signaling [47].

Finally, it should be noted that at many loci there is much less resolution to infer the causal mutations. For example, significant variants span ~25 kb at the *epidermal growth factor receptor*, and some differentiated regions are much larger and contain many genes (Figure S4). The set of significant variants at these loci is still a minute fraction of genomic variation, so these variants can now be used as *a priori* functional candidates in an association study. This will reduce the genotyping effort required, and greatly increase the statistical power. For this reason, we consider our approach to be largely complementary to more traditional genome-wide association studies, with mapping resolution at some loci small enough to proceed directly to functional characterization, while at others additional mapping will be required. Furthermore, this approach will increase the number of species where powerful genotype-phenotype mapping is possible, as it can be utilized in any species with a suitable life-history. We therefore expect the E&R approach to be a major component of future efforts to identify and characterize the molecular polymorphisms responsible for the tremendous phenotypic diversity observed within populations.

Methods

Selection on body size

To sort flies by size, approximately 1800 flies from each population were anesthetized with CO_2 and placed into a shaking column (Gilson Performer III, Gilson Company). For generations, 1–30, flies were separated using 6 U.S.A. Standard Test Sieves (ATSM E-11 specification; #10,12,14,16, 18 & 20), in which the diameter of the openings in each sieve was approximately 20% larger than the openings of the sieve below (average of top sieve openings = 2000 μm, average of bottom sieve openings = 850 μm). In order to create a finer scale selection gradient after generation 31, flies were sieved with a custom made set of 20 electroformed sieves, in which the diameter of the holes in each sieve were only 5% larger than the holes of the sieve below (average of top sieve holes = 1685 μm; average of bottom sieve holes = 800 μm). Each generation, the most extreme 160 males and 160 females in each selected line were allowed to reproduce. For control populations,

an equal number of random individuals was selected after they were sedated, sieved, and then mixed back together. Throughout, fly populations were maintained under a standard culturing process, designed to match the rearing conditions of the LH_M base population as closely as possible. Rearing conditions are described in detail in [27].

Phenotypic measurement

Leg, wing, and thorax measures were taken from each population to investigate the anatomical consequences of selection for sieve size. Length measures were made from images taken with a 3.3 Megapixel IC D integrated digital camera on a Leica M205C stereomicroscope using Image-Pro Analyzer 7.0 software. Fifteen individuals per sex were scored for twelve phenotypes. Thorax length was recorded as the distance between the posterior tip of the scutellum to the anterior most point of the prescutal suture. Distances between the posterior scutellar and upper humeral bristles and the posterior scutellar and anterior sternopleural bristles were recorded from the thorax as well. All thoracic measurements were done on the left side of the thorax after the legs had been disarticulated. Femur lengths for all legs were measured. The distances between nine landmarks on the wings were also evaluated following [48]. Lengths of each measure were recorded in triplicate from an individual and averages were used in analyses; no measurement was included if intra-individual variation was greater than ten percent. Average phenotype scores were assessed for thorax measures. For bilaterally symmetrical phenotypes, averages between left and right appendages were evaluated, thus phenotypes from individuals with body parts that sustained damage during the freezing/thawing/dissection process were not included. Tukey HSD tests were performed to asses statistical significance between small-selected, large-selected, and control populations, as presented in Table S1.

Resequencing and sequence analysis

DNA was isolated from bulked females for each population. For control populations, 25 females from generation 118 and 50 females from generation 120 were frozen, homogenized, and lysed together. Bulk lysate was then divided over multiple Qiagen DNeasy columns for purification, and the resulting elutions were pooled. Flies were collected from small and large populations concurrently with controls, but these females were from generations 108 and 110, as these populations were originally started 10 generations after controls. Library construction and genome sequencing were performed by the University of North Carolina at Chapel Hill High-Throughput Sequencing Facility according to standard Illumina protocols. A 35-bp unpaired library and a 75-bp paired-end library was created from each of the 6 populations, and each library was sequenced in one channel of the Illumina Genome Analyzer. Reads were aligned to the v. 5.26 *Drosophila melanogaster* genome using BWA [49]. For all downstream analyses, repetitive and low quality reads were excluded: BWA generates mapping qualities for each alignment: only reads with mapping qualities >15 ($P \approx 0.032$) were used; 91% of these reads had qualities ≥37 ($P \approx 0.0002$). BWA aligns reads to the reference genome permitting both base mismatches and insertions and deletions, each of which was considered a candidate polymorphism. For each base pair location with more than two states, the allele frequencies reported are for the common allele versus all other alleles. Heterozygosity was calculated at each base with 4× or more coverage in each population with the following formula: $\pi = 1 - ((A*A) + (G*G) + (C*C) + (T*T) + (D*D))$ where A, G, C, T, D are the frequencies of these bases at that site, where D is a deletion; for sites where the second most common allele was at <0.10

frequency, π was considered to be zero. These values of π were averaged across non-overlapping 10 kb windows to generate genomic distributions.

Significance testing

Population-based resequencing resulted in a large number of apparent genetic polymorphisms and an estimate of the frequency of each allele in each population. We were then interested in differentiating alleles that have been affected by selection (direct or linked selection) from those that have not. This requires accounting for two types of sampling error: the stochastic changes in allele frequency since these populations separated from a common ancestor (drift), and sampling error due to sequencing a small number of alleles from the larger population.

Observed allele frequency differences were quantified using a summary statistic: the pair-wise difference in allele frequency between each pair of divergently selected populations was computed, and the smallest difference between up-and down-selected populations (i.e., min[abs(up1-down1),abs(up1-down2),-abs(up2-down1),abs(up2-down2)]) was called the "diffStat" test statistic. To incorporate the consensus among comparisons, the test statistic is set to be zero unless all four comparisons have the same sign. Observed polymorphisms were binned by starting allele frequency, which was estimated using the average ending frequency of the two control populations. For each allele frequency bin, the observed distribution of diffStat was then compared to an expected distribution to generate a false discovery rate estimate. This expected distribution of differentiation without selection was determined by first simulating binomial sampling (drift) in 4 populations for 110 generations. Though the number of breeding adults in each experimental population was 320, differences in reproductive success depress the effective population size below this number. Estimates of variation in reproductive success for each sex are available for the LH_M population from which experimental lineages were derived (Pischedda *et al.* in prep): using these values and equations from [50], Ne in these populations was estimated at 0.699*N for the autosomes and 0.591*N for the X chromosome. We therefore used a simple binomial sampler to simulate drift for 110 generations in populations of 224 reproductive individuals (for the autosomes) and 189 reproductive individuals (for the X chromosome). For each of 10 starting allele frequency classes, we simulated drift in 500,000 replicate simulations to determine the expected distribution of diffStat across the genome. To incorporate sampling error due to sequencing a finite number of alleles, an additional sampling step was performed after the final generation of simulation. In each simulation run, for each population, a coverage value was sampled from the observed distribution of sequencing coverages for each experimental population, and this number of alleles was sampled from the simulated populations. For a given threshold diffStat value, the FDR is estimated as the expected number of polymorphisms/the observed number of polymorphisms: FDR values for each threshold and starting allele frequency are shown in Tables S2 and S3.

This analysis yields an estimate of the proportion of polymorphisms subject to either direct or linked selection above any given threshold diffStat value. However, it should be noted that the confidence intervals of allele frequency at any given polymorphic site are variable due to variable sequence coverage. An alternative approach would be to use Fisher's exact test to determine the statistical significance of allele counts, rather than using frequency data. To compare this approach to our method, we computed *p*-values from Fisher's exact test on all simulated polymorphisms and compared this distribution to the observed data. Results are

similar when using Fisher's exact tests. Using diffStat, 1236 regions are significant with 10-kb separation, and 304 regions are significant at 50-kb separation; the numbers from the Fisher's analysis are 1173 regions and 314 regions. When the peak variants are calculated from this Fisher's exact test data, and genes within 1-kb of these variants are determined, the same three functional categories (post-embryonic development, metamorphosis, and cell morphogenesis) are still quite significant after Bonferonni correction, but the >2-fold enrichment seen with diffStat is reduced. This is likely because, with the Fisher analysis, we consider the variant with the smallest p-value to be the best candidate for selection, whereas in the diffStat we consider the most differentiated polymorphism (or all of the most differentiated polys, if there is a tie). For example, if 10 polymorphisms at a locus are reciprocolly fixed between treatments, we consider them equally likely to be the target of selection in the diffStat analysis; in the FIsher's analysis, if one polymorphism has higher coverage, it will have the lowest *p-value* and be considered the best candidate. We consider the diffStat analysis to be a better way to balance type I and type II error, as plotting the *p*-values from Fisher's exact tests results in deceptively sharp peaks. The greater over-representation of expected functional categories may support the contention that the diffStat analysis as being more accurate at homing in on the selected locus.

Finally, to determine if estimating starting allele frequency from the controls introduces bias, we carried out several additional simulations. As above, we simulated 50,000 alleles from each of ten starting frequencies, this time using only two populations to simulate the two control populations. We then considered only those alleles with an average ending frequency between 0.40–0.60: Figure S5A shows that 90% of alleles with these ending frequencies started with frequencies between 0.20–0.80. Finally, given that the ending frequency was 0.40–0.60, we determined the distribution of differentiation between the two populations. As shown in Figure S5B, the expected distributions are very similar, when conditioned on the same average ending frequency, especially for alleles which started above 0.20. As alleles which contributed nearly all simulated values that started above 0.20, we conclude that estimating allele frequencies in this way introduces little error. In addition, FDR thresholds were very similar across allele frequency categories (Tables S2, S3), indicating that FDR thresholds are only slightly sensitive to starting frequency after 110 generations of evolution.

Supporting Information

Dataset S1 Functional clustering of GO terms generated by DAVID.
Found at: doi:10.1371/journal.pgen.1001336.s001 (0.46 MB XLS)

Dataset S2 Significant GO Terms, as generated by DAVID.
Found at: doi:10.1371/journal.pgen.1001336.s002 (0.20 MB XLS)

Dataset S3 Genes within 1 kb of a peak variant, excluding the Chr. 2 centromere.
Found at: doi:10.1371/journal.pgen.1001336.s003 (0.47 MB TDS)

Dataset S4 All peak variants.
Found at: doi:10.1371/journal.pgen.1001336.s004 (0.41 MB XLS)

Figure S1 Read coverage of genome partitions using reads with alignment qualities greater than 15. A: Chromosomes X, 2, and 3; B: centromeric regions, C: mitochondria, D: Y chromosome, E: U

and Uextra (unplaced regions). Females were sequenced, so Y coverage is expected to be near zero for unique alignments, though some male DNA could be present if females were mated and had stored sperm.

Found at: doi:10.1371/journal.pgen.1001336.s005 (0.73 MB DOC)

Figure S2 Differentiation of all chromosomes. The diffStat is shown for each variant that had higher or lower allele frequencies in the large-selected lines compared to the small-selected lines. Color coding indicates significance: black = nonsignificant variants, blue = significant variants at the permissive FDR threshold (FDR<10%); gold = significant variants at the restrictive FDR threshold (FDR<5%); red = peak variants. Left arms (L) are the to left of the centromere (centromere is after high x-axis values), and the right (R) arms are to the right of the centromere (centromere is before low x-axis values).

Found at: doi:10.1371/journal.pgen.1001336.s006 (6.21 MB DOC)

Figure S3 Differentiation of all chromosomes. The diffStat is shown for each variant that had higher or lower allele frequencies in the large-selected lines compared to the small-selected lines. Color coding indicates estimated starting allele frequency: black = all variants, gold = variants with an average control frequency <0.05; red = peak variants. Left arms (L) are the to left of the centromere (centromere is after high x-axis values), and the right (R) arms are to the right of the centromere (centromere is before low x-axis values).

Found at: doi:10.1371/journal.pgen.1001336.s007 (3.86 MB DOC)

Figure S4 Examples of significantly differentiated genes. The diffStat is shown for each variant at each locus; colors are as in figure S2. A: At EGFR several polymoprhisms are significantly differentiated across 25-kb, B: At dally, a large region of

differentiation encompassed multiple genes and many polymoprhisms. For the candidate gene at each locus, the exons are shown as linked grey boxes; only one transcript for simplicity.

Found at: doi:10.1371/journal.pgen.1001336.s008 (0.86 MB DOC)

Figure S5 Low bias introduced by estimating starting allele frequency. A: the proportion of variants with average frequency of 0.40–0.60, depending on starting frequency. B: Distribution of expected allele frequency differentiation for alleles which have a final average frequency between 0.45–0.50 in the two control populations; colors indicate starting allele frequencies, red = 0.50; orange = 0.45; yellow = 0.40; green = 0.35; blue = 0.30; purple = 0.25; fuchsia = 0.20. The distribution is very similar until starting allele frequency is 0.25 or less, and few alleles with this starting frequency end with an average frequency between 0.40–0.60, as shown in A.

Found at: doi:10.1371/journal.pgen.1001336.s009 (0.30 MB DOC)

Table S1 Significance of anatomical measurements.
Found at: doi:10.1371/journal.pgen.1001336.s010 (0.04 MB PDF)

Table S2 False discovery rates for each allele frequency class at various thresholds for autosomes.
Found at: doi:10.1371/journal.pgen.1001336.s011 (0.02 MB PDF)

Table S3 False discovery rates for each allele frequency class at various thresholds for the X chromosome.
Found at: doi:10.1371/journal.pgen.1001336.s012 (0.02 MB PDF)

Author Contributions

References

1. De Jong G, Bochdanovits Z (2003) Latitudinal clines in *Drosophila melanogaster*: body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. Journal of Genetics 82: 207–223.
2. Capy P, Pla E, David JR (1993) Phenotypic and genetic variability of morphometrical traits in natural populations of *Drosophila melanogaster* and *D. simulans*. I. Geographic variations. Genetics, Selection, Evolution 25: 517–536.
3. Gockel J, Kennington WJ, Hoffmann A, Goldstein DB, Partridge L (2001) Nonclinality of molecular variation implicates selection in maintaining a morphological cline of Drosophila melanogaster. Genetics 158: 319–323.
4. Partridge L, Barrie B, Fowler K, French V (1994) Evolution and development of body size and cell size in *Drosophila melanogaster* in response to temperature. Evolution 48: 1269–1276.
5. Prasad NG, Bedhomme S, Day T, Chippindale AK (2007) An evolutionary cost of separate genders revealed by male-limited evolution. American Naturalist 169: 29–37.
6. Robertson FW (1963) The ecological genetics of growth in *Drosophila melanogaster*. VI. The genetic correlation between the duration of the larval period and body size in relation to larval diet. Genetics Research 4: 74.
7. Partridge L, Langelan R, Folwler K, Zwaan B, French V (1999) Correlated responses to selection on body size in *Drosophila melanogaster*. Genetical Research 74: 43–54.
8. Partridge L, Fowler K (1992) Direct and correlated responses to selection on age at reproduction in *Drosophila melanogaster*. Evolution 46: 76–91.
9. Norry FM, Loeschcke V (2002) Temperature-induced shifts in associations of longevity with body size in *Drosophila melanogaster*. Evolution 56: 299–306.
10. Long TAF, Pischedda A, Stewart AD, Rice WR (2009) A cost of sexual attractiveness to high-fitness females. PLoS Biol 7: e1000254. doi:10.1371/journal.pbio.1000254.
11. Lefranc A, Bundgaard J (2000) The influence of male and female body size on copulation duration and fecundity in *Drosophila melanogaster*. Hereditas 132: 243–247.
12. Partridge L, Farquhar M (1983) Lifetime mating success of male fruitflies (*Drosophila melanogaster*) is related to their size. Animal Behaviour 31: 871–877.
13. Pitnick S, Garcia-Gonzalez F (2002) Harm to females increases with male body size in *Drosophila melanogaster*. Proceedings of the Royal Society B-Biological Sciences 269: 1821–1828.
14. Allen HL, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832–838.
15. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42: 565–569.
16. Park J-H, Wacholder S, Gail MH, Peters U, Jacobs KB, et al. (2010) Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nature Genetics 42: 570–575.
17. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, et al. (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature 465: 627–631.
18. Wright S (1977) Artificial Selection with Insects. Evolution and the Genetics of Populations Volume 3 Experimental Results and Evolutionary Directions. Chicago: University of Chicago Press.
19. Kohler RE (1994) Lords of the fly: Drosophila genetics and the experimental life. Chicago: University of Chicago Press. pp 23–33.
20. Robertson FW, Reeve ECR (1952) Studies in quantitative inheritance: !. The effects of selection of wing and thorax length in *Drosophila melanogaster*. Journal of Genetics 50: 414–448.
21. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, et al. (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461: 1243–U1274.
22. Wichman HA, Scott LA, Yarber CD, Bull JJ (2000) Experimental evolution recapitulates natural evolution. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 355: 1677–1684.
23. Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, et al. (2010) Antagonistic coevolution accelerates molecular evolution. Nature 464: 275–U154.
24. Nuzhdin SV, Harshman LG, Zhou M, Harmon K (2007) Genome-enabled hitchhiking mapping identifies QTLs for stress resistance in natural Drosophila. Heredity 99: 313–321.
25. Johansson AM, Pettersson ME, Siegel PB, Carlborg Ö (2010) Genome-wide effects of long-term divergent selection. PLoS Genet 6: e1001188. doi:10.1371/journal.pgen.1001188.
26. Burke MK, Dunham JP, Shahrestani P, Thornton KR, Rose MR, et al. (2010) Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature 467: 587–590.
27. Rice WR, Linder JE, Friberg U, Lew TA, Morrow EH, et al. (2005) Inter-locus antagonistic coevolution as an engine of speciation: Assessment with hemiclonal analysis. Proceedings of the National Academy of Sciences of the United States of America 102: 6527–6534.

28. Chippindale AK, Gibson JR, Rice WR (2001) Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. Proceedings of the National Academy of Sciences 98: 1671–1675.

29. Platt A, Vilhjálmsson BJ, Nordborg M (2010) Conditions under which genome-wide association studies will be positively misleading Genetics. Advance Online Publication: September 2, 2010, doi:2010.1534/genetics.2110.121665.

30. Andolfatto P, Wall JD (2003) Linkage disequilibrium patterns across a recombination gradient in African Drosophila melanogaster. Genetics 165: 1289–1305.

31. Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, et al. (2009) FlyBase: enhancing *Drosophila* Gene Ontology annotations. Nucleic Acids Research 37: D555–D559.

32. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protocols 4: 44–57.

33. Dennis GJ, Sherman BT, Hosack DA, Yang J, Gao W, et al. (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology 4: P3.

34. Calboli FCF, Gilchrist GW, Partridge L (2003) Different cell size and cell number contribution in two newly established and one ancient body size cline of *Drosophila subobscura*. Evolution 57: 566–573.

35. Hariharan IK (2006) Growth regulation: a beginning for the hippo pathway. Current Biology 16: R1037–R1039.

36. Grzeschik NA, Parsons LM, Allott ML, Harvey KF, Richardson HE (2010) Lgl, aPKC, and crumbs regulate the Salvador/Warts/Hippo pathway through two distinct mechanisms. Current Biology 20: 573–581.

37. Robinson BS, Huang J, Hong Y (2010) Crumbs regulates Salvador/Warts/Hippo signaling in *Drosophila* via the FERM-domain protein expanded. Current Biology 20: 582–590.

38. Mirth K, Riddiford LM (2007) Size assessment and growth control: how adult size is determined in insects. Bioessays 29: 344–355.

39. Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genetical Research 23: 23–35.

40. Bohni R, Riesgo-Escovar J, Oldham S, Brogiolo W, Stocker H, et al. (1999) Autonomous control of cell and organ size by CHICO, a *Drosophila* homolog of vertebrate IRS1-4. Cell 97: 865–875.

41. Chen C, Jack J, Garofalo RS (1996) The *Drosophila* insulin receptor is required for normal growth. Endocrinology 137: 846–856.

42. Paaby AB, Blacket M, Hoffmann AA, Schmidt PS (2010) Identification of a candidate adaptive polymorphism for *Drosophila* life history by parallel independent clines on two continents. Molecular Ecology 19: 760–774.

43. McKechnie SW, Blacket MJ, Song SV, Rako L, Carroll X, et al. (2010) A clinally varying promoter polymorphism associated with adaptive variation in wing size in Drosophila. Molecular Ecology 19: 775–784.

44. Stowers RS, Garza D, Rascle A, Hogness DS (2000) The L63 gene is necessary for the ecdysone-induced 63E late puff and encodes CDK proteins required for *Drosophila* development. Developmental Biology 221: 23–40.

45. Shimojima T, Okada M, Nakayama T, Ueda H, Okawa K, et al. (2003) *Drosophila* FACT contributes to Hox gene expression through physical and functional interactions with GAGA factor. Genes & Development 17: 1605–1613.

46. Tsunaka Y, Toga J, Yamaguchi H, Tate S, Hirose S, et al. (2009) Phosphorylated intrinsically disordered region of FACT masks Its nucleosomal DNA binding elements. Journal of Biological Chemistry 284: 24610–24621.

47. Sliter TJ, Gilbert LI (1992) Developmental arrest and ecdysteroid deficiency resulting from mutations at the *dre4* locus of *Drosophila*. Genetics 130: 555–568.

48. K. K, Thomsen H, Faurby S, Kjærsgaard A, Tejedo M, et al. (2009) Consequences of outbreeding on phenotypic plasticity in *Drosophila mercatorum* wings. Evolutionary Ecology 23: 403–415.

49. H. L, R. D (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.

50. Lande R, Barrowclough GF (1987) Effective population size, genetic variation, and their use in population management. In: Soulé ME, ed. Viable populations for conservation. Cambridge: Cambridge University Press. pp 87–123.