University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Lawrence G. Harshman Publications

Papers in the Biological Sciences

2007

Genome-Enabled Hitchhiking Mapping Identifies QTLs for Stress Resistance in Natural *Drosophila*

Sergey V. Nuzhdin
University of California, Davis, svnuzhdin@ucdavis.edu

Lawrence G. Harshman
University of Nebraska - Lincoln, lharshman1@unl.edu

M. Zhou
University of Nebraska - Lincoln

K. Harmon
University of California, Davis

Follow this and additional works at: http://digitalcommons.unl.edu/biosciharshman

Genome-Enabled Hitchhiking Mapping Identifies QTLs for Stress Resistance in Natural *Drosophila*

S. V. Nuzhdin,[1] L. G. Harshman,[2] M. Zhou,[2] and K. Harmon[1]

1. Department of Evolution and Ecology, University of California at Davis, Davis, California, USA
2. School of Biological Sciences, University of Nebraska, Lincoln, Nebraska, USA

Corresponding author – Dr. S. V. Nuzhdin, email svnuzhdin@ucdavis.edu

Abstract

Identification of genes underlying complex traits is an important problem. Quantitative trait loci (QTL) are mapped using marker-trait co-segregation in large panels of recombinant genotypes. Most frequently, recombinant inbred lines derived from two isogenic parents are used. Segregation patterns are also studied in pedigrees from multiple families. Great advances have been made through creative use of these techniques, but narrow sampling and inadequate power represent strong limitations. Here, we propose an approach combining the strengths of both techniques. We established a mapping population from a sample of natural genotypes and applied artificial selection for a complex character. Selection changed the frequencies of alleles in QTLs contributing to the selection response. We infer QTLs with dense genotyping microarrays by identifying blocks of linked markers undergoing selective changes in allele frequency. We demonstrated this approach with an experimental population composed from 20 isogenic strains. Selection for starvation survival was executed in three replicated populations with three control non-selected populations. Three individuals per population were genotyped using Affymetrix GeneChips. Two regions of the genome, one each on the left arms of the second and third chromosomes, showed significant divergence between control and selected populations. For the former region, we inferred allele frequencies in selected and control populations by pyrosequencing. We conclude that the allele frequency difference, averaging approximately 40% between selected and control lines, contributed to selection response. Our approach can

contribute to the fine scale decomposition of the genetics of direct and indirect selection responses and genotype by environment interactions.

Keywords: stress resistance, natural variation, complex traits

Introduction

The most common approaches for deciphering the architecture of natural genetic variation include quantitative trait loci (QTL) mapping (Lander and Botstein, 1989) and quantitative complementation tests (Geiger-Thornsberry and Mackay, 2004). An alternative strategy for mapping genes controlling complex characters is to follow frequency changes at marker loci in selected populations derived from a cross of parental lines. Selection changes the frequencies of the molecular markers because of hitchhiking with alleles of the selected trait (Thomson, 1977), allowing inference of the linkage between the markers and QTLs. While the idea for such an experiment was formulated long ago (Dumouchel and Anderson, 1968; Garnett and Falconer, 1975), its first implementation (Nuzhdin and Pasyukova, 1991) was later than conventional QTL mapping (Lander and Botstein, 1989). Several subsequent hitchhiking mapping studies (Keightley and Bulfield, 1993; Nuzhdin et al., 1993, 1998; Keightley et al., 1998) established that mapping QTLs using hitchhiking marker allele frequencies is a potentially powerful approach.

There has also been a solid theoretical foundation for the data analyses developed (Kim and Stephan, 1999). Lebowitz et al. (1987) analyzed the changes in marker allele frequencies under different intensities of selection, with varying sizes of the population under selection, number of generations of selection and with different numbers of the genotyped individuals. They concluded that for modest selection intensities, QTLs of relatively large effects can be detected, and the size of the selected population shall remain large to minimize the effects of drift. Gallais et al. (2007) developed traditional infinitesimal model to explicitly compare the powers of Marker-based (MB) and Trait-based (TB) analyses providing genotyping and phenotyping costs. They conclude that if the cost of phenotyping is low, but that of genotyping is high, the TB technique might be more efficient. Tenesa et al. (2005) use analytical approaches to conclude that MB and TB techniques have similar power, with one round of selection, when a single trait is analyzed, but MB is more efficient in multitrait analyses. They derive their inferences in association-mapping framework, flexible in relation to the initial QTL allele frequencies. Overall, TB mapping is well suited to studying fitness-related traits such as starvation stress resistance, since the selection can be implemented in large populations needed to minimize drift. The approach is powerful for low heritability traits and does not require the development and maintenance of inbred lines.

The potential of hitchhiking mapping is especially pronounced when facilitated by DNA-typing microarrays. Whole-genome polymorphism detection with short oligonucleotide chips was pioneered in yeast by Winzeler et al. (1998). They demonstrated that allelic variation in any isolates of a sequenced species can be scanned, mapped, and scored directly and efficiently by total DNA to chip hybridization. Winzeler et al. used this technique to compare 14 different yeast stocks using S98 Affymetrix arrays with 285,156 25b

features covering ~16% of the yeast genome. They found 11,115 single feature polymorphisms (*sfp*s, i.e., mishybridizations of oligonucleotide with DNA resulting from nucleotide changes, deletions, and insertions) or close to a thousand *sfp*s per genotype. This approach has been successfully extended to the genetic analysis of Arabidopsis (Borevitz et al., 2003) and Anopheles (Turner et al., 2005). We employed this same technology for genotyping in flies for the following reasons: (i) DNA of two unrelated flies is different in approximately 1% of nucleotides and most of the polymorphisms are present in low frequencies and (ii) the probability of a DNA mismatch lowering hybridization intensity at a detectable level is close to 50% depending on the proximity of the mismatch to the middle of the oligonucleotide printed on the microarray (Ronald et al., 2005). Accordingly, all of the founder genotypes are individually recognizable by thousands of *sfp*s (Gresham et al., 2006). When linkage blocks marked by these *sfp*s change in frequency, this change can be detected by numerous *sfp*s.

Here, we focus on the QTL analysis of natural genetic variation in starvation resistance. Natural populations of most species, including *Drosophila melanogaster*, are probably periodically subjected to food shortages to a degree that selects for starvation resistance. Thus, starvation is a particularly relevant form of stress, ecologically and evolutionarily, that warrants investigation in terms of its underlying molecular and genetic mechanisms. We select for resistance to acute starvation, not confounded by dehydration, following the procedures used by Harshman and Schmid (1998). Previously, selection for starvation resistance has resulted in extended longevity as an indirect response to selection (Rose et al., 1992), but this outcome was not consistent among similar selection experiments (Harshman and Hoffmann, 2000). Multiple stress resistance is another correlated response to selection (Harshman et al., 1999a,b) suggesting, as one possibility, involvement of one or a few genes that mediate a general form of stress resistance or, as another possibility, that many genes contribute to the indirect response to selection. Global hitchhiking mapping might be an ideal approach to test whether there are a limited number or large number of genes involved in the selection response.

Materials and methods

Laboratory selection for starvation resistance
Approximately 2 years before selection, we initiated a base population from 20 inbred lines of *D. melanogaster*. These inbred lines were derived from individual inseminated females collected in the field (Wolfskill Orchards maintained by the University of California at Davis, Yolo County, North Carolina, USA). The first generation of adults from laboratory culture was subjected to sib mating that continued for up to 40 generations. To establish the base population, we reciprocally crossed each of the inbred lines to a subset of other inbred lines in a balanced design based on the goal of equal representation of all lines. A standard number of progeny was harvested from each cross. We released thousands of these heterozygous progeny at the same time into a 91 × 60 × 31 cm population cage. Twenty bottles containing a standard *Drosophila* food (yeast, cornmeal, molasses, and agar) were in the cage with the adult flies. The next generation, the population of adult flies in the cage was over 10 000 individuals. Every week five fresh bottles of food were added to

the cage and five of the original food bottles were removed. After a month, the five oldest bottles were replaced by five fresh bottles every week. This overlapping generation regime was maintained for 26 months before the start of laboratory selection.

For laboratory selection we established three selected and three control populations. Each of these populations was initiated with progeny from 500 males and 500 individuals from the base population. Progeny males and females remained together for 3–4 days before separation by sex using brief exposure to ether. At approximately day 7 post-eclosion, 100 males or 100 females were placed in an empty bottle capped by a fiber plug saturated with water. We kept the plug wet throughout the time of selection at 25°C (12L:12D). At every 12-h interval, the bottles were checked until approximately 30% mortality and then checked every 8 h until approximately half the males and half the females were dead. Each selected line was assigned a corresponding control line that was treated in the same manner except for the period of acute starvation. Specifically, we matched selected line no. 1 to control line no. 1 in terms of the number of males and females used to initiate each generation.

We allowed one generation of relaxed selection after selection generation 13 and then collected adults for DNA extraction. For the generation used to extract DNA, 75 eggs were added to each of a series of rearing vials from each selected and control line. Adults reared from these vials were used to measure the direct response to selection in terms of survivorship tabulated at 12 h intervals in five replicate bottles for each sex for each selected and control line in the absence of food, but with water present and high humidity. The same age, density, and environmental conditions used for selection were used to measure the direct response to selection, but we continued this assay until all flies were dead. The average survival time within each replicate was fit to analysis of variance (ANOVA) model with selection regime, sex, selection by sex, population within selection, and sex by population within selection (see supplementary material file 1).

Genomic analysis
To ensure that DNA from more than one individual was not present, as would be the case for inseminated females, we isolated DNA from single males (7 days old). We homogenized a male using the A Dounce pestle in 800 μl of buffer (0.15M NaCl, 0.01M Tris–HCL pH 8.0, 0.005M EDTA, 0.2% NP-40). The homogenate was filtered through a small plug of glass wool in a Pasteur pipette. Nuclei were pelleted by centrifugation at 8000 rpm for 2 min and resuspended the pellet in 50 μl of Douncing buffer, then mixed with 150 μl of lysis buffer (0.3M NaCl, 0.05M Tris–HCL pH 8.0, 0.005M EDTA, 1% NaSarkosyl). We mixed the lysed nuclei with 200 μl of phenol:chloroform:isoamyl alcohol and followed with the addition of 1 μl of glycogen (20 μg/μl), 0.2 volume of 3M sodium acetate (pH 5.8) and 2–3 volumes of ethanol. We incubated the mixture at –20°C for over 1 h followed by centrifugation at 14 000 rpm for 15–20 min. After carefully removing the supernatant, we washed the pellet with 70% ethanol and centrifuged it at 14 000 rpm for 2–5 min. Then, the pellet was resuspended in 25 μl of TE buffer. The integrity and abundance of DNA were confirmed by visual inspection after electrophoresis of 2.5 μl of the DNA solution in 0.8% agarose.

To increase the quantity of DNA for use on microarrays we amplified it using Qiagen's Repli-g kit (Qiagen, Hilden, Germany: 59043 http://www1.qiagen.com/products/genomic dnastabilizationpurification/replig/repligkit.aspx?ShowInfo=1). We then purified each sample with phenol by combining the amplified DNA with 800 μl phenol:chloroform, centrifuging at 14 000 g for 5 min and collecting the aqueous layer. To precipitate the DNA we added 30 μl sodium acetate (pH 5.5) and 600 μl 100% cold ethanol to the aqueous layer, allowed the nucleic acids to precipitate on ice for 10 min and then pelleted the DNA by centrifugation for 5 min. After an additional wash with 80% ethanol, centrifuging for 3 min, and allowing the pellet to dry, we resuspended the DNA pellet in 50 μl of ddH2O.

We fragmented 10 μg of the amplified, purified DNA using Mike Zwick's resequencing array protocol (Cutler et al., 2001). To do this we prepared a fragmentation cocktail containing 80 μl One Phor All Buffer (Amersham Biosciences, Piscataway, NJ, USA: 27-0901-02), 3.66 μl DNAse1 (Promega: M610A), and 2.8 μl of acetylated bovine serum albumin (Invitrogen, Carlsbad, CA, USA: 15561-020) and added 4.3 μl of this buffer to 10 μg of the DNA. This mixture was incubated at 37°C for 16 min, heated to 99°C for 15 min and then cooled for 5 min at 12°C. Continuing with the Zwick's resequencing array protocol, we labeled the DNA by adding 1 μl RTdT enzyme (Promega, Madison, Wisconsin, USA: M1875) and 1 μl Biotin-N6-ddATP (Enzo, New York, NY, USA: 42809) to 1 μl of the fragmented DNA solution. We incubated the mixture at 37°C for 90 min, heated it to 99°C for 15 min to deactivate the enzyme and then allowed it to cool for 5 min at 12°C. The University of California at Davis School of Medicine Microarray Core Facility hybridized the microarrays using genomic DNA in place of cDNA in the standard Affymetrix protocol (Affymetrix, Santa Clara, California, USA). Raw hybridization intensities were extracted using the Bioconductor Affy package (Gautier et al., 2004; http://www.bioconductor.org) and slide mean normalized (see supplementary material file 2).

To discover the genomic regions divergent between selected and control populations, we used two approaches. First, we recorded the probes significantly different for hybridization intensity between selected and control lines using two-tailed *t*-test *P*-values less than 0.001 (PROC GLM, SAS Institute Inc., 1988). To evaluate the regions of differentiation, we calculated the numbers of significant probes within non-overlapping windows of 1Mb (each containing about 2000 oligonucleotides). The choice of the window size was due to the following considerations. The rate of recombination in females of flies is approximately 2 cM per 1 Mb. During the recombination breakdown phase, chromosomes pass through males with no recombination half of the time. Accordingly, initially fully linked *sfp*s situated 1Mb from each other are still associated in 50% of the chromosomes at the start of selection. Associations of selected alleles with *sfp*s in neighboring windows might decrease by as much as twofold. To test the significance of the regional divergence, we permuted the positions of oligonucleotides and recorded the highest number of *sfp*s within a window for each permutation. This analysis was overly simplistic in ignoring the relatedness of individuals within selected populations. Second, we reanalyzed the data with the ANOVA model incorporating the effects of selection and population nested within selection. Identical regions of genomic differentiation were detected with both analyses.

We selected eight Affymetrix probes of interest and designed PCR primer pairs using Primer3 (http:// frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) that would amplify

100–200 bp of the surrounding regions, which we obtained by BLASTing against the *Drosophila* genome. One primer of each pair we had 5' biotin labeled and high-performance liquid chromatography purified (Thermo Oligos, Uem, Germany). Additionally, we used Biotage's primer design SW program (http://techsupport.pyrosequencing.com/) to design a single pyrosequencing primer immediately upstream of the probe of interest on the strand complementary to the one amplified by the biotin-labeled PCR primer. We isolated DNA from the six populations, this time using pools of 100 flies. We used standard conditions to PCR amplify the DNA. The Veterinary Genetics Lab at UC Davis processed these PCR products on a PSQ 96MA Pyrosequencer, using their standard protocol. We mixed 3 μl beads and 37 μl binding buffer (Biotage 40-0033, Uppsala, Sweden) per PCR reaction and then added 40 μl of this mixture into each sample well in the PCR plate. While allowing the samples to incubate on a mixer for 10 min we combined 5 μl of sequencing primer (Thermo Oligos) with 35 μl annealing buffer (Biotage 40-0036) and added 40 μl to each well of an empty sequencing plate (Biotage 40-002). When the samples finished mixing, we used a vacuum tool to remove the liquid and leave the beads and PCR products bound to the probes. We then dipped the vacuum tool into troughs filled with 70% ethanol, denaturation solution (0.2M NaOH), and 1 × washing buffer (Biotage 40-0035 diluted 1:10) for 5 s each. After switching off the vacuum, we released the samples into the wells of the sequencing plate by inserting the probes of the vacuum tool into the bottom of the wells and rubbing in small circles. After heating the sequencing plate at 80°C for 2 min, we allowed it to cool to room temperature. As the samples cooled, we added enzyme mixture, substrate, and dNTPs in the volumes specified by supplier (Biotage 40-0045) to the cartridge of the pyrosequencer, inserted cartridge and sequencing plate into the pyrosequencer, and began the run. The frequencies of alleles were calculated from the relative signals (ordinates) corresponding to alternative nucleotides.

Results

The goal is to identify natural alleles that improve starvation resistance in selected lines, as compared to population mean. A sample of 20 *Drosophila* genomes from a natural population is expected to include such alleles (Madalena and Robertson, 1974). How do we find the genetic region where they are contained? To answer this question, we applied a hitchhiking mapping approach (fig. 1). First, we "mixed" a sample of genotypes into a population large enough to make drift of allele frequencies negligible. Second, to improve resolution of subsequent mapping, we decreased initial sampling linkage disequilibrium by allowing approximately 50 generations of recombination in the absence of artificial selection. Clearly, the genetic composition of this base population could have been affected by inadvertent selection and by drift during the linkage breakdown phase. To minimize the latter effect, the selected and control subpopulations were each initiated with large sample size (approximately 1000 individuals) from the base population of approximately 10,000–15,000 individuals. For subsequent hitchhiking mapping, drift has little importance as all control and selected populations were derived from a single founder population.
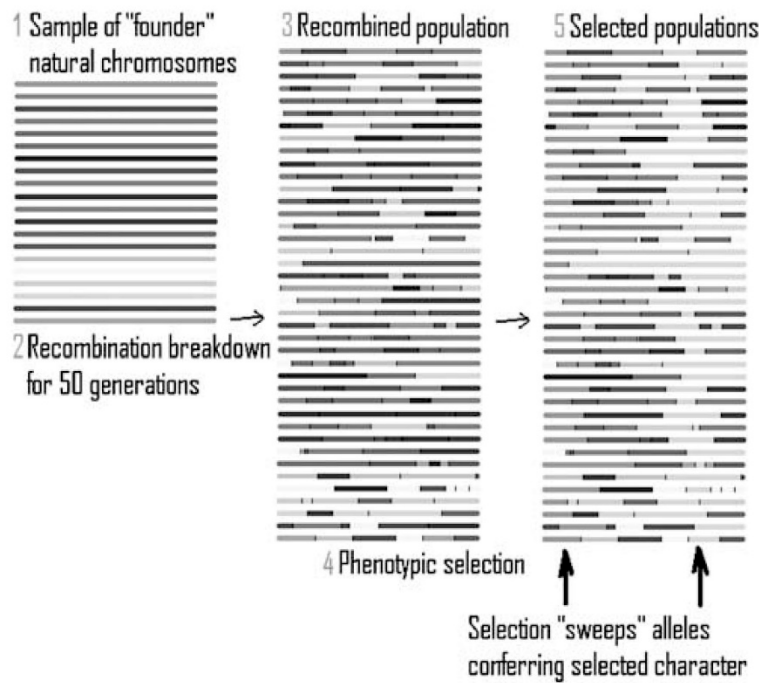
Figure 1. Steps of hitchhiking mapping: (1) sampling from nature establishes initial "sampling linkage disequilibrium," or "haplotypes" represented by different shades (a single chromosome per haplotype is presented); (2) recombination lowers "sampling linkage disequilibrium" to (3) chromosomal fragments; (4) phenotypic selection changes frequencies of alleles responsible for selection response; (5) which is established by whole-genome analysis of marker allele frequencies, to finally detect selection sweeps.

We intend to analyze the change in the whole-genome allele frequencies reached as a result of selection (symbolically represented in fig. 1). To implement selection, we placed a constant number of mated females or males in an otherwise empty bottle capped by a water-saturated fiber plug. We retained approximately 50% of the surviving males and females in each selected line as breeders for the next generation and the same number of randomly picked individuals was used for each matched control line. The minimum number of selected and control line flies was 250 males and 250 females each generation. After 13 generations of selection, there is a clear direct response of both sexes relative to the survival of control line flies. Mortality was monitored until all flies died; generating lifetime survival curves (fig. 2). As shown, both sexes are significantly more starvation resistant than the controls, as indicated by the grand average of all three selected lines compared to all three control lines. As the selected and control populations were clearly differentiated from each other as a result of selection ($P = 0.0001$, fig. 2), they should also be divergent for the frequencies of alleles conferring starvation resistance.
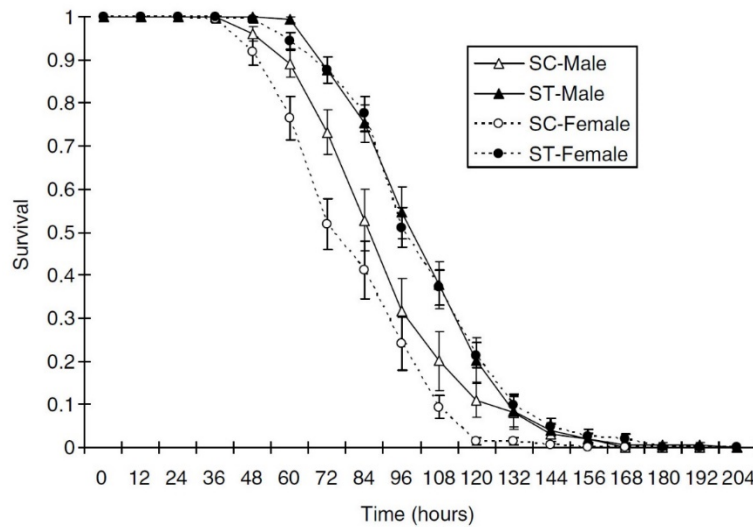
Figure 2. The direct response to selection is shown for selected lines (females or males) for comparison to control lines (females or males) 14 generations after the lines were initiated (13 generations of selection and 1 generation of relaxed selection). The grand mean number surviving and the standard error of the means for all three selected lines and all three control lines are shown at 12-h intervals during the starvation survival assay. We conducted the assay in replicate containers for each sex for each line in the acute absence of food, but with high humidity and access to water. Density and other environmental conditions were the same as used for the selection experiment except that in the assay we monitored mortality until all flies died. The selected lines (ST) clearly survive longer than the control lines (SC). The slope of survival appears to be similar for selected and control lines; the delay in the initial phase of mortality in the selected lines appears to make the major contribution to the response to selection. There is less variation among and within the selected lines than among and within the control lines that is not due to differences in population size (see Materials and methods). The relative uniformity of the selected lines might reflect a convergent response to selection whereas the control lines, not selected for starvation survival, are less constrained.

We isolated DNA for hitchhiking mapping and pyrosequencing after one generation of relaxed selection following 15 generations of selection for starvation resistance. We detected allele frequency differences among populations using Affimetrix GeneChip V.2. expression microarrays composed of 14 25b oligonucleotides per gene (with a few exceptions). If a natural allele carries a mismatch to a microarray oligonucleotide, it results in lowered hybridization signal. If the frequency of such an allele increases in selected populations, the average hybridization signal for this oligonucleotide will accordingly decrease. The difference in hybridization intensity between control and selected populations might then be significant if the changes in allele frequencies are large and consistent among replicated populations. We extracted DNA from three individual flies from each selected and control line and hybridized each of these samples of DNA to separate microarrays for a total of 18 arrays (three arrays per each selected and control line).

The methods used for analysis of the signals from total DNA–microarray DNA hybridizations are as follows. For every array feature, the mean normalized hybridization intensity (see supplementary material file 1) was analyzed in one way ANOVA to select oligonucleotides divergent between selected and control populations. The signal of hybridization from an individual oligonucleotide may be noisy and potentially misleading. Note, however, that in an earlier study we did confirm most *sfp*s when we checked them with direct resequencing (Turner et al., 2005). Accordingly, we based the analysis on the significance of haplotype changes rather than an individual *sfp*s. In our experimental scheme, polymorphisms tightly linked due to initial sampling remain in strong disequilibrium and represent a haplotype block. A block incurring a selective sweep is marked by numerous *sfp*s with significant divergence between control and selected populations. The composite signal of many divergent *sfp*s marking a haplotype block is unambiguous. Figure 3 represents the number of features within a 1Mb window which are significantly different ($P < 0.001$) between selected and control populations. We chose this window as it approximately corresponds to the size of linkage disequilibrium blocks after 50 generations of random mating. Several regions, most notably the left arm of the second chromosome, are greatly enriched by significantly divergent *sfp*s. To test whether such clumping is expected by chance alone, we permuted chromosomal locations of the *sfp*s 1000 times, each time recording the highest number of significant *sfp*s within a 1 Mb window. Not a single permutation resulted in a number of significances higher than detected with the original non-permuted data set. We conclude that our technique detects haplotype blocks strongly divergent between selected and control populations.
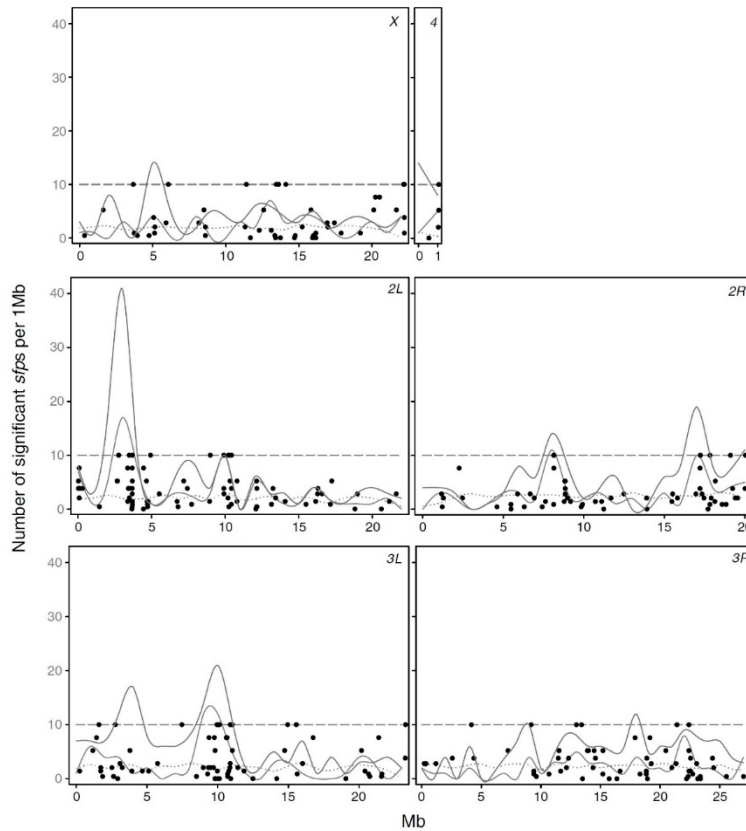
Figure 3. The features are shown that significantly differ ($P < 0.001$ is plotted at zero value, and those with $P < 0.0001$ at the value 10) between selected and control populations along chromosomes with *t*-test. The number of significant features per 1Mb is represented as a solid line. The number of features expected to show difference by chance (there are 2000 oligonucleotides per 1Mb on average) is represented as a dotted line. The dashed line is a permutation-based significance threshold. At least two chromosomal regions (on the left arms of the second and third chromosomes) appear responsible for selection response. In the most significant window, 17 *sfp*s instead of the 2 expected by chance are detected, and the numbers of significant *sfp*s in two neighboring windows are also elevated (5 and 8). Similarly, the third chromosomal differentiated region has a run of significant windows marked by 12, 12, and 5 *sfp*s. The same regions were detected when clustering of features significant at the level $0.001 < P < 0.01$ was considered (data not shown). The number of features significant at $P < 0.001$ level between selected and control populations with the selection and population nested within selection two-way ANOVA is represented by dot-dash line. ANOVA, analysis of variance.

    While microarrays establish the regions of divergence between selected and control lines, likely due to selection, they yield no information on actual allele frequency differences achieved due to selection. The next step of hitchhiking analysis should be to independently measure, with a different technique, the frequencies of *sfp*s in selected and control populations. The technique of choice was DNA sequencing and for this purpose

we designed primers immediately adjacent to significant *sfp*s from the second chromosome QTL. We used them for PCR amplifications from bulk samples of DNA prepared from 100 flies per population. We then analyzed the sequences of the oligonucleotides with a pyrosequencing technique that allows accurate estimation of allele frequency from bulk samples (Neve et al., 2002). Polymorphisms were evident for four oligonucleotides out of seven, consistent with the expectation that some *sfp*s are false discoveries (see footnotes for fig. 3). Polymorphic *sfp*s had different allele frequencies in selected populations compared to matched controls (table 1). The differences were the largest, approximately 40%, in the middle of the differentiated region. We conclude that the inference of regions of divergence between selected and control populations detected in microarray analysis is supported by an alternative technique.

Table 1. Allele frequencies of *sfp*s showing significant divergence in microarray hybridization signal between selected (s1, s2, and s3) and control (c1, c2, and c3) populations

Affymetrix nos.	Sequence/starting position	Pyrosequence	c1	c2	c3	s1	s2	s3
1629241	CGAGCACTCTAACCAGCT GTGGAGC	cgagcactcC	0.38	0.26	0.60	0.44	0.76	0.70
	3377983	cgagcactcT	0.40	0.42	0.52	0.41	0.73	0.70
1624958	TATCATCCCTTATTAAGCC ACCAAT	tatcaGcccttatTaC	0.37	0.38	0.44	0.47	0.80	0.73
	3522491	tatcaGcccttatCaA	0.35	0.34	0.41	0.40	0.63	0.61
1636605	GATTTAGTACCAAAGCTG AGTGCGA	gatttagT	0.19	0.35	0.63	0.83	0.77	0.73
	3703992	gatttagC	0.24	0.53	0.63	0.79	0.74	0.74
1640456	CGAAAGCCGCGTGTGGGTC GACCAC	cgaaagccG	0.42	0.12	0.13	0.41	0.58	0.56
	4449755	cgaaagccA	0.42	0.18	0.18	0.41	0.57	0.55

Abbreviation: *sfp*s, single feature polymorphisms

Affymetrix number and the sequence of the resequenced oligonucleotide are shown. The column "Pyrosequence" represents the sequence of the oligonucleotide up to a polymorphic site, the frequency of which (in two replicates) is reported in the columns to the right. For example, the expected sequence of the oligonucleotide 1629241 was CGAGCACTCTAACCAGCTGTGGAGC, while the pyrosequencer read two alleles: cgagcactcC and cgagcactcT. The sequence of the mismatching one is shown on top as its frequency was increased in the selected populations. Pyrosequencing-defined allele frequencies were highly reproducible between replicates (for instance 0.38 and 0.40 in the population c1, or 0.76 and 0.73 in the population s2).

Discussion

Inferring the architecture of genetic variation in complex traits remains one of the frontiers of research. While much progress has been achieved with QTL mapping, these experiments remain tedious, resource and labor intensive, and statistically challenging. Moreover, the QTL regions often span a broad region of the genome that often covers at least a 1000 genes making it difficult to identify candidate genes for further studies. A typical experiment starts with two genotypes, and the alleles differently contributing to the trait

value are mapped with a large panel of genotyped recombinant individuals. With few exceptions, successes are limited to model systems in which making controlled crosses and establishing isogenic lines are relatively easy (Mackay, 2004). Furthermore, it has recently been recognized that two segregating alleles might be too small a sample to derive meaningful conclusions. A single mapping population segregating for up to eight alleles was established (Peirce et al., 2004). Studying many alleles at once allows one to "polarize" their effects as compared to the population mean (Kopp et al., 2003). However, it also greatly increases statistical challenges.

Our study is intended to illustrate an alternative strategy that is more amenable to testing numerous natural alleles at once, does not require establishing panels of recombinant inbred lines and does not necessitate resource-consuming genotyping of hundreds of individuals. Instead, we focus on allele frequency differences achieved as a result of selection. Note that as all three starvation-selected populations are highly differentiated from control populations, the divergence must be largely due to selection as opposed to drift alone. With dense microarray-based genotyping, we mapped natural alleles of QTLs improving starvation resistance in natural populations. While chips are expensive, only a few of them are required because selection can be continued until a high degree of divergence is attained at which point hitchhiking mapping can be conducted. Note, that to fully interpret the data from this or future analogous experiments, more conceptual and modeling work is called for. As we do not know initial allele frequencies (one QTL allele might be linked to several different alleles in marker locus and other way around, and different QTL alleles in a locus may interfere during selection response), the distribution of their effects and the extent of their linkage disequilibrium with marker alleles, it is at present impossible to detail the model of selection response. These are important matters that need to be clarified for rectified interpretations.

While our primary goal was to test the efficacy of hitchhiking mapping in a multiple allele context, our results also help to guide future research on the genetic basis of starvation resistance. Genes that can confer resistance to starvation in *D. melanogaster* are known primarily from mutation studies of longevity. In addition, a study of global gene expression of adult flies under starvation conditions, in conjunction with the phenotypic effects of transposable element mutations (Harbison et al., 2005), also provided information about the impact of specific genes on starvation resistance. In that study, the response to starvation involved transcriptional alterations of nearly 25% of the coding genes. Upregulated genes were relatively highly represented in gene ontology categories of growth and maintenance. Protein biosynthesis and metabolism were strongly represented, as well as translation initiation and elongation factors, and hydrolases acting on acid anhydrides. Down-regulated genes fell disproportionately into the following categories: proteases, carrier activity, and defense-immunity responses. In general, we consider genes that affect intermediary metabolism to be worthy of consideration as candidates. For example, it is known that lipid accumulation is a typical response to selection for starvation resistance in the laboratory (Djawdan et al., 1998; Harshman et al., 1999a, b).

There are 117 genes in the interval 23F1-24F3. Candidate genes that play a role in metabolism include those involved in mitochondrial electron transport (*Pdsw*, CG15434) as well as genes involved in various metabolic processes including disaccharide metabolism

(*Tps1*). Prospectively important are the genes involved in pathogen defense and stress responses (*Thor*, CG33123, *Dot*, *Sr-CI*, *Sr-CIII*, *Traf1*). The *Thor* gene product inhibits translation upon nutrient deprivation, and it also plays a significant role in resistance to bacterial infection. The *Thor* promoter has an *NF-κâ* recognition sequence (Bernal and Kimbrell, 2000). Another gene involved in patho-*Traf1* is part of the *NF-κâ* cascade, which plays a major role in stress responses. *Thor* and *Traf1* might be the strongest candidate genes in the second chromosome QTL interval.

There are 162 genes in the interval 67B2-67E7. Although a case could be made for a various types of candidate genes, there are two especially compelling clusters of genes in this region of the genome. Laboratory selection for starvation resistance is associated with multiple stress resistance (Harshman et al., 1999a, b) and sometimes with longevity (Harshman and Hoffmann, 2000). Pertinently, there are eight heat shock protein genes (*Hsp67Bc*; *Hsp22*; *Hsp67Bb*, *Hsp26*; *Hsp67B*a; *Hsp23*; *Hsp*27, GC4461). Heat shock proteins are prime candidates that play a role in multiple stress resistance. For example, they could underlie the phenotypes conferred by the *D. melanogaster Methuselah* mutation which includes extended longevity as well as heat stress and starvation tolerance (Lin et al., 1998). Strikingly, five of seven of the *D. melanogaster* genes encoding insulin-like growth factors are clustered in this region. Specifically, *Ilp1–Ilp4* are contiguous and *Ilp-5* is closely linked. These genes have great potential to alter metabolism and play a role in the response to selection for starvation resistance. Clearly, in this interval on the third chromosome, the cluster of heat shock protein genes and the cluster of genes encoding insulin-like proteins are prime candidates for being involved in the response to selection for starvation resistance.

The next challenge is to narrow down the QTLs to the genes and nucleotides underlying quantitative variation. One way is to use an "association studies" paradigm (Mackay and Langley, 1990). Polymorphisms causing trait deviations (quantitative trait nucleotides or QTNs) are statistically inferred from associations between a phenotype and DNA variation in large panels of natural genotypes. The creative application of this approach has recently accounted for substantial breakthroughs (De Luca et al., 2003; Genissel et al., 2004), especially for traits with relatively simple genetic determination. However, further analysis revealed a potential of high false discovery rates—reliable detection of associations requires dense genotyping of thousands of individuals and replicating the experiments over several populations (Long and Langley, 1999). The prospects for this approach are not high because the phenotypic effect of a QTL is small (with some exceptions, 5–10% of genetic variation, Long et al., 2000; Remington et al., 2001; Robin et al., 2002; Shapiro et al., 2004), and the number of polymorphisms within a QTL region is enormous. In general, identification of QTNs responsible for small phenotypic differences between QTL alleles remains to be one of the greatest challenges in evolutionary genetics and genetics of complex diseases (Keightley, 1995; Long and Langley, 1999; Long et al., 2000; Genissel et al., 2004).

The power of detecting robust associations, while reducing the genotyping costs, might be accomplished by applying many generations of selection. To greatly increase the resolution, one might rely not on sampling but on population linkage disequilibrium—small haplotype blocks persisting in the populations due to limited population size (Wu and Zeng, 2001). In *Drosophila*, the natural linkage disequilibrium blocks might be as small as

200–500 bp (Haddrill et al., 2005). Very dense genotyping will be required to detect selection sweeps for such narrow chromosomal regions. However, this is not unthinkable. The next generation of tiling microarrays feature oligonucleotides spaced every 50 nucleotides and covering half of the genome (http:// www.affymetrix.com). Combining the power of tiling arrays and hitchhiking mapping might greatly clarify the mystery of the genetic nature of quantitative variation.

The potential of hitchhiking mapping supported by whole-genome approaches extends beyond increased power of analysis. One of the key limitations in improving plant and animal stocks remains in identifying factors contributing to selection response (Varshney et al., 2005). As a selected population is usually started from a few founders, or modest sized population, genome-enabled hitchhiking mapping might clarify the genetics of response with existing selected populations, and the approach is fiscally feasible. One of the key questions in evolution and ecology concerns the genetics of adaptation (Orr, 2005). Alleles are segregating in populations that are beneficial in marginal environments. Local subpopulations adapted to such environments should have increased frequencies of locally beneficial alleles. Whole-genome assessment of the genomic divergence between general and marginal populations should allow for inference of such divergence clarifying the genetics of adaptation. As oligonucleotide arrays are available for a large and fast growing assortment of species (http://www.affymetrix.com), the potential of future research directions appears impressive. There will be a complication in such studies—that the frequency of beneficial allele in a founder population is unknown. Moreover, in every selected region, a mixture of alleles with different effects might be present, and the dynamics of allele frequency changes are likely to be very complex. Theoretical analysis of the inferences will be a serious issue to address in the future. Our study has proven the potential of hitchhiking mapping with dense microarrays. In the future, it might be improved in multiple ways. For instance, microarray genotyping data are noisy, but the amount of noise may be reduced when DNA samples from multiple individuals are combined before hybridization. If hybridization technical error is not overwhelming, pooling individuals might lead to higher consistency of the data, and to more accurate inferences on allele frequency differences between samples. Another option is to use RNA samples which might enable simultaneous assessment of allele frequency divergence and transcript level divergence between selected lines. While here we only scan for the regions under selection, in the future it might be important to further study explicit genetics of selection response. For instance, one may design markers targeting separately each of alleles in the initial sample, follow their allele frequency changes during selection response and reconstruct the distributions of allelic effects in initial sample from such data.

References

Bernal A, Kimbrell DA (2000). *Drosophila* Thor participates in host immune defense and connects a translational regulator with innate immunity. *Proc Natl Acad Sci USA* 97: 6019–6024.

Borevitz JO, Liang D, Plouffe D, Chang H-S, Zhu T, Weigel D, et al. (2003). Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res* 13: 513–523.

Cutler DJ, Zwick ME, Carrasquillo MM, Yohn CT, Tobin KP, Kashuk C, et al. (2001). High-through-put variation detection and genotyping using microarrays. *Genome Res* 11: 1913–1925.

De Luca M, Roshina NV, Geiger-Thornsberry GL, Lyman RF, Pasyukova EG, Mackay TFC (2003). *Dopa decarboxylase* (*Ddc*) affects variation in *Drosophila* longevity. *Nat Genet* 34: 429–433.

Djawdan M, Chippindale AK, Rose MR, Bradley TJ (1998). Metabolic reserves and evolved stress resistance in *Drosophila melanogaster*. *Physiol Zool* 71: 584–594.

Dumouchel WH, Anderson WW (1968). The analysis of selection in experimental populations. *Genetics* 58: 435–449. Gallais A, Moreau L, Charcosset A (2007). Detection of marker-QTL associations by studying change in marker frequencies with selection. *Theor Appl Genet* 114: 669–681.

Garnett I, Falconer DS (1975). Protein variation in strains of mice differing in body size. *Genet Res* 25: 45–57.

Gautier L, Cope L, Bolstad BM, Irizarry RA (2004). Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20: 307–315.

Geiger-Thornsberry GL, Mackay TFC (2004). Quantitative trait loci affecting natural variation in *Drosophila* longevity. *Mech Ageing Dev* 125: 179–189.

Genissel A, Pastinen T, Dowell A, Mackay TFC, Long AD (2004). No evidence for an association between common nonsynonymous polymorphisms in *Delta* and bristle number variation in natural and laboratory populations of *Drosophila melanogaster*. *Genetics* 166: 291–306.

Gresham D, Ruderfer DM, Pratt SC, Schacherer J, Dunham MJ, Botstein D, et al. (2006). Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. *Science* 311: 1932–1936.

Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005). Multi-locus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res* 15: 790–799.

Harbison ST, Chang S, Kamdar KP, Mackay TFC (2005). Quantitative genomics of starvation stress resistance in *Drosophila*. *Genome Biol* 6: R36.

Harshman LG, Hoffmann AA, Clark AG (1999a). Selection for starvation resistance in *Drosophila melanogaster*: physiological correlates, enzyme activities and multiple stress responses. *J Evol Biol* 12: 370–379.

Harshman LG, Moore KM, Sty MA, Magwire MM (1999b). Stress resistance and longevity in selected lines of *Drosophila melanogaster*. *Neurobiol Aging* 20: 521–529.

Harshman LG, Hoffmann AA (2000). Laboratory selection experiments on life history and stress resistance traits in *Drosophila*; what do they really tell us? *Trends Ecol Evol* 52: 32–36.

Harshman LG, Schmid J (1998). Evolution of starvation resistance in *Drosophila melanogaster:* aspects of metabolism and counter-impact selection. *Evolution* 52: 1679–1685.

Keightley PD (1995). Quantitative genetics: loci with large effects. *Curr Biol* 5: 485–487.

Keightley PD, Bulfield G (1993). Detection of quantitative trait loci from frequency changes at marker loci under selection. *Genet Res* 62: 195–203.

Keightley PD, Morris KH, Ishikawa AV, Falconer M, Oliver F (1998). Test of candidate gene quantitative trait locus association applied to fatness in mice. *Heredity* 81: 630–637.

Kim YS, Stephan W(1999). Allele frequency changes in artificial selection experiments: statistical power and precision of QTL mapping. *Genet Res* 73: 177–184.

Kopp A, Graze RM, Xu S, Carroll SB, Nuzhdin SV (2003). Quantitative trait loci responsible for variation in sexually dimorphic traits in *Drosophila melanogaster*. *Genetics* 163: 771–787.

Lander ES, Botstein D (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185–199.

Lebowitz RJ, Soller M, Beckmann JS (1987). Trait-based analysis for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. *Theor Appl Genet* 73: 556–561.

Lin Y-J, Seroude L, Benzer S (1998). Extended life-span and stress resistance in the *Drosophila* mutant methuselah. *Science* 30: 943–946.

Long AD, Langley CH (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res* 9: 720–731.

Long AD, Lyman RF, Morgan AH, Langley CH, Mackay TFC (2000). Both naturally occurring insertions of transposable elements and intermediate frequency polymorphisms at the *achaete-scute* complex are associated with variation in bristle number in *Drosophila melanogaster*. *Genetics* 154: 1255–1269.

Mackay TFC (2004). The genetic architecture of quantitative traits: lessons from *Drosophila*. *Curr Opin Genet Dev* 14: 253–257.

Mackay TFC, Langley CH (1990). Molecular and phenotypic variation in the achaete-scute region of *Drosophila melanogaster*. *Nature* 348: 64–66.

Madalena FE, Robertson A (1974). Population structure in artificial selection: studies with *Drosophila melanogaster*. *Genetical Res* 24: 113–126.

Neve B, Froguel P, Corset L, Vaillant E, Vatin V, Boutin P (2002). Rapid SNP allele frequency determination in enomic DNA pools by Pyrosequencing (TM). *Biotecniques* 32: 1138–1142.

Nuzhdin SV, Keightley PD, Pasyukova EG (1993). The use of retrotransposons as markers for mapping genes responsible for fitness differences between related *Drosophila melanogaster* strains. *Genet Res* 62: 125–133.

Nuzhdin SV, Keightley PD, Pasyukova EG, Marozova EA (1998). Quantitative trait loci mapping in the course of divergent selection for sternopleural bristle number of *Drosophila melanogaster*. *Genet Res* 72: 79–91.

Nuzhdin SV, Pasyukova EG (1991). New approach to polygene mapping. Location of genes controlling fitness level in related *Drosophila melanogaster* stocks. *Genetika (Russia)* 27: 849–859.

Orr HA (2005). The genetic theory of adaptation: a brief history. *Nat Rev Genet* 6: 119–127.

Peirce JL, Lu L, Gu J, Silver LM, Williams RW (2004). A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genetics* 5: Art. No. 7.

Remington DL, Ungerer MC, Purugganan MD (2001). Map-based cloning of quantitative trait loci: progress and prospects. *Genet Res* 78: 213–218.

Robin C, Lyman RF, Long AD, Langley CH, Mackay TFC (2002). *hairy*: a quantitative trait locus for *Drosophila* sensory bristle number. *Genetics* 162: 155–164.

Ronald J, Akey JM, Whittle J, Smith EN, Yvert G, Kruglyak L (2005). Simultaneous genotyping, gene-expression measurement, and detection of allele-specific expression with oligonucleotide arrays. *Genome Res* 15: 284–291.

Rose MR, Vu LN, Park SU, Graves JL (1992). Selection on stress resistance increases longevity in *Drosophila melanogaster*. *Exp Gerontol* 27: 241–250.

SAS Institute Inc (1988). *SAS/STAT User's Guide. Release 6.03 Edition*. SAS Institute Inc.: Cary, NC. Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, et al. (2004). Genetic and developmental basis of evolutionary pelvic reduction in three-spine sticklebacks. *Nature* 428: 717–723.

Tenesa A, Visscher PM, Carothers AD, Knott SA (2005). Mapping quantitative trait loci using linkage disequilibrium: marker versus trait based methods. *Behav Genet* 35: 219–228.

Thomson C (1977). The effect of a selected locus on linked neutral loci. *Genetics* 285: 753–788.

Turner TL, Hahn M, Nuzhdin SV (2005). Genomic islands of speciation in Anopheles gambiae. *PLoS Biol* 3: 1572–1578.

Varshney RK, Graner A, Sorrells ME (2005). Genomics-assisted breeding for crop improvement. *Trends Plant Sci* 10: 621–630.

Winzeler EA, Richards DR, Conway AR, Goldstein AL, Kalman S, McCullough MJ, et al. (1998). Direct allelic variation scanning of the yeast genome. *Science* 281: 1194–1197.

Wu RL, Zeng ZB (2001). Joint linkage and linkage disequilibrium mapping in natural populations. *Genetics* 157: 899–909.