Investigating Natural Variation in *Drosophila* Courtship Song by the Evolve and Resequencing Approach

Thomas L. Turner¹ and Paige M. Miller

Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, California 93106

ABSTRACT A primary goal of population genetics is to determine the genetic basis of natural trait variation. We could significantly advance this goal by developing comprehensive genome-wide approaches to link genotype and phenotype in model organisms. Here we combine artificial selection with population-based resequencing to investigate the genetic basis of variation in the interpulse interval (IPI) of *Drosophila melanogaster* courtship song. We performed divergent selection on replicate populations for only 14 generations, but had considerable power to differentiate alleles that evolved due to selection from those that evolved stochastically. We identified a large number of variants that changed frequency in response to selection for this simple behavior, and they are highly underrepresented on the X chromosome. Though our power was adequate using this experimental technique, the ability to differentiate causal variants from those affected by linked selection requires further development.

RECENT technological advances have led to a rapid increase in the number of genotype–phenotype links in human genetics (McCarthy *et al.* 2008). Many of these links were assembled by genotyping a large number of individuals across the genome and conducting genome-wide association studies (GWAS). These genome-wide approaches can be conducted without *a priori* hypotheses regarding causal variants, thus maximizing their discovery potential. However, high-frequency variants generally have only mild to moderate phenotypic effects (Hindorff *et al.* 2009; Park *et al.* 2010), and as such require large numbers of genotyped individuals to be reliably mapped. Recent studies in humans leveraged data from hundreds of thousands of people but were still only able to map a fraction of the heritable variation for their traits of interest, illustrating the limitations of GWAS for linking genotype and phenotype (Allen *et al.* 2010; Yang *et al.* 2010, 2011; Makowsky *et al.* 2011).

Applications of genome-wide approaches in model systems hold tremendous promise (Manenti *et al.* 2009; Atwell

et al. 2010; Boyko *et al.* 2010; Huang *et al.* 2010). If we could efficiently characterize genotype–phenotype links in these experimental systems, we would develop our functional understanding of how biological systems tolerate or make use of genetic variation and also advance our evolutionary understanding of the maintenance of variation and causes of divergence. In the *Drosophila melanogaster* model system in particular, low levels of linkage disequilibrium and a wide array of tools for genetic manipulation allow for comprehensive follow-up to mapping studies (Mackay and Langley 1990; Odgers *et al.* 1995; Lyman *et al.* 1999; Inomata *et al.* 2004; Fiumera *et al.* 2007; Hill-Burns and Clark 2009). Mackay *et al.* (2008) proposed that completely sequencing ~200 inbred lines in this species should provide sufficient power for GWAS, because we can reduce the nonheritable component of phenotypic variation by repeatedly phenotyping many individuals of the same genotype. While this may be true for variation associated with a small number of large-effect alleles, a much larger number of lines may be necessary to attain sufficient statistical power when phenotypic variation is caused by many smaller-effect alleles (Long and Langley 1999). For this reason, we propose that population-based resequencing of artificially selected populations provides a powerful compliment to GWAS in *D. melanogaster* (Nuzhdin *et al.* 2007; Burke *et al.* 2010; Earley and Jones 2011; Turner *et al.* 2011; Zhou *et al.* 2011). This

Copyright © 2012 by the Genetics Society of America
doi: 10.1534/genetics.112.139337

Manuscript received February 6, 2012; accepted for publication March 18, 2012

Supporting information is available online at <http://www.genetics.org/content/suppl/2012/03/30/genetics.112.139337.DC1>.

¹Corresponding author: EEMB, Noble Hall, Room 2234, University of California, Santa Barbara, CA 93106. E-mail: tturner@lifesci.ucsb.edu

approach is not limited by the creation and maintenance of large sets of inbred lines and could therefore be leveraged in other species of *Drosophila*, other insects, annual plants, and any other system where large populations can be manipulated over multiple generations (Johansson *et al.* 2010; Parts *et al.* 2011). The ability to do these comparative studies will likely prove critical for gaining evolutionary insight from genotype–phenotype links. These advantages—additional statistical power and increased applicability—make genome-wide mapping approaches based on artificial selection a promising route to major advances in population genetics.

We recently resequenced replicate populations of *D. melanogaster* that were divergently selected for body size (Turner *et al.* 2011). After >100 generations of divergent selection, genomic divergence in excess of neutral expectations was apparent at hundreds of loci throughout the genome. Though this result supports the utility of the “evolve and resequence” (E&R) approach, the long-term selection conducted in this experiment may not be optimal for mapping purposes. If E&R requires large replicate populations to be selected for hundreds of generations, it will not be a feasible approach in many species. Such long-term experimental evolution projects can also lead to results that are difficult to interpret or less relevant to natural populations. For example, the fixation of alleles with additive effects may transform epistatic variation at other loci into additive variation (Falconer 1981). Additionally, long-term experimental evolution may lead to strong compensatory evolution at alleles not directly affecting the trait of interest. For example, if smaller flies have different optimal mating strategies than larger flies, long-term selection for small body size may lead to divergent selection on loci affecting courtship behavior.

Here, we applied the E&R approach to a smaller, shorter experiment to further investigate the power and promise of this method. We divergently selected male *D. melanogaster* for a model behavior of evolutionary interest: the interpulse interval (IPI) of the male courtship “song” (Figure 1). When male *D. melanogaster* court females, they produce an auditory signal with their wings that, together with pheromonal information, is believed to be one of the primary modes of information exchange between the sexes (Shorey 1962; Bennet-Clark and Ewing 1969; Spieth 1974; von Schilcher 1976; Gleason 2005). The primary auditory signal, the pulse song, consists of a series of song pulses and is sometimes preceded by a sine song. Though the function of the sine song is unknown, presence of the pulse song increases male courtship success considerably (Bennet-Clark and Ewing 1969; von Schilcher 1976; van Den Berg 1988; Ritchie *et al.* 1998, 1999; Talyn and Dowse 2004). The time between pulses in this song (the IPI) is a species-specific trait (Cowling and Burnet 1981; Wheeler *et al.* 1991; Gleason and Ritchie 1998; Blyth *et al.* 2008), with apparent intersexual coevolution between the IPI of the male song and female preference for a given IPI (Tomaru *et al.* 1998, 2000; Tomaru and Oguma 2000; Doi *et al.* 2001). As such,

Figure 1 *D. melanogaster* pulse song. Sample of pulse song showing four pulse trains: pulses appear as vertical lines at this scale (above). Magnification of two pulses from the third pulse train (below). Y-axis is amplitude in arbitrary units.

this trait is an evolutionarily relevant behavior that is likely under sexual selection within species and/or under selection for species recognition between species. Here we show considerable additive variation for this trait, with the IPI of divergently selected populations differing by several standard deviations after only 14 generations of selection. Resequencing these artificially selected populations at this time point yielded considerable power to distinguish alleles affected by selection from those that had become stochastically differentiated. However, E&R under these parameters resulted in poor mapping resolution, possibly due to: (i) selection on uncommon alleles with correspondingly large linked differentiation, (ii) selection on a very large number of causal loci, and/or (iii) inadequate haplotype diversity in the starting population. We discuss potential solutions to these problems to guide future efforts.

Methods

Starting population

All flies were maintained in 25 × 95 mm vials on molasses medium in standard *Drosophila* incubators, at 25° under a 12-hr light/dark cycle. To create a base population for E&R, we cross-mated flies from 173 inbred lines *en masse*. We used the Raleigh (RAL) collection of inbred lines created in the lab of Trudy Mackay (Ayroles *et al.* 2009; Mackay *et al.* 2012). These lines were collected by the Mackay lab in Raleigh, North Carolina, and each line underwent full-sib inbreeding for 20 generations to eliminate most genetic variation. This process basically extracts chromosomes from the wild population into stable lines with minimal selection. Though the inbreeding process will have subtle selective effects (such as loss of any recessive lethal mutations), mixing these lines back together approximately recreates a sample of the wild outbred population from Raleigh (referred to here as the “allRAL” population). These RAL lines are particularly useful as starting material for our E&R populations because they allow for comparisons of

E&R with GWAS on the same lines, and they provide the ability to repeatedly recreate the populations from the same lines.

To create our allRAL population, we mixed one to four virgin females and males from each of the 173 lines and stochastically allotted these flies into 110 culture vials, with five males and five females per vial. When the offspring of these parents eclosed (10–12 days later), we again mixed them together and created 100 new culture vials with five males and five females in each. This procedure was repeated until the F₅ generation. An important consideration is that the scale of linkage disequilibrium (LD) in this population is limited more by the diversity of haplotypes in the starting sample than by recombination during these few generations of mixing. For example, if a causal allele is present in 50% of the chromosomes in the natural Raleigh population, 50% of the 173 RAL lines are expected to carry haplotypes containing this variant, with similarity between haplotypes limited by the scale of LD in the starting population. Mixing 173 inbred lines is therefore approximately equivalent to starting the population from 86 diploid individuals from nature.

Evolve and resequence: population maintenance

In the F₅ generation of the allRAL population, we recorded courtship songs from 1,050 males (see *Song recording* below). Males with 20 or fewer recorded IPIs were discarded, leaving 861 males. We created two populations selected for short IPI by randomly distributing the males from the lower 20% tail of the distribution of IPI into two populations. We created two populations selected for long IPI in the same manner using males from the upper 20% tail of the distribution. For this first generation, all populations were established using virgin allRAL females from the F₅ generation. For subsequent generations we combined males with virgin females from their own populations. Each generation, we set up multiple culture vials by placing two selected males and six virgin females into each vial. After 48 hr of oviposition in these vials, we transferred the flies into vials containing fresh media, repeating this over a 6-day period to have a continuous supply of age-standardized males and virgin females for recording. We then froze parent flies for archiving. We collected virgin males and virgin females for phenotyping into unisex holding vials at low density (10 males or 20 females) using CO₂ anesthesia. Males were given a minimum of 48 hr to mature and recover from anesthesia before recording, while females were used as courtship objects for recording after 24 hr (1-day-old females elicit vigorous courtship but are reticent to mate in the recording interval). After recording, we placed each male individually into a test tube with a small amount of food. After recording was completed, we selected males from the appropriate test tubes and mixed them with virgin females from their population to create new culture vials. We maintained populations as a set of vials with only two selected males and six (randomly chosen) females per vial (rather than in large containers) to minimize competition between individuals and therefore minimize selection unrelated to treatment. Each

generation, we mixed offspring between all vials to maintain the set of vials as a panmictic population. Though our protocol may have inadvertently selected for other traits, such as fecundity or courtship effort, we would expect most of these factors to be selected in all populations and not lead to consistent allele frequency differentiation between treatments.

Evolve and resequence: selection protocol

After the first generation, in which the upper and lower 20% of males were selected, the precise number of males selected for each population varied according to the proportion that sang enough to be well quantified (>20 IPIs), and the time available for recording that generation. We quantified an average of 331 males per population each generation, and 15–25% of these males were selected to found the following generation (average = 20%). The precise numbers of males recorded and selected each generation are available in supporting information, [Table S1](#).

In addition to the ~20% of recorded males selected per population, we also transferred additional “migrant” males to each population every generation. This migration was intended to homogenize allele frequencies at neutral loci, reducing stochastic differentiation due to drift. These migrant males were chosen from populations selected in the opposite direction: no gene flow was allowed between populations selected in the same direction. Relatedness at neutral loci was therefore higher between populations selected in opposite directions (e.g., if gene flow is 5% between populations *a* and *b*, and 5% between populations *b* and *c*, then gene flow is on the order of $5\% \times 5\% = 0.25\%$ between populations *a* and *c*).

Males were initially chosen for migration if they had IPI values that *would have been* selected if they were sampled from populations selected in the opposite direction. For example, in the 4th generation of selection, an average 4.5% of males from each short-selected population had IPIs as long as the longest 20% of males in the long-selected populations (which were selected). These 4.5% migrated to the long-selected populations and were treated the same as selected individuals. In the 2nd, 3rd, and 4th generations of selection, migrants made up >20% of selected individuals (in the 1st generation of selection the four populations were derived from a common population). By the 5th generation of selection the migrant fraction averaged 4.8%. We continued migration at a low level (<3%) until the 11th generation of selection, even after the populations became divergent enough to no longer have any migrants that qualified (in round 9). We then continued selection for an additional 3 generations without migration, and halted it after 14 total generations of selection. Numbers of migrants for each population and each generation are provided in [Table S1](#).

Song recording

Hardware for song recording was adapted from a system developed by the Barry Dickson lab, Research Institute of Molecular Pathology, Vienna (von Philipsborn *et al.* 2011).

We recorded songs from 10 males at a time: each male was placed with a virgin female from his own population in an acrylic courtship chamber. Each chamber was attached to a custom-fitted 40PH array microphone (GRAS Sound and Vibration), and courtship occurred on plastic mesh immediately adjacent to this microphone. A 16-channel RogaDAQ universal serial bus (USB) data acquisition system (Roga Instruments) converted analog inputs into digital output for all 10 microphones in parallel. Resulting digital data were processed and displayed in real-time with Labview 2009 software (National Instruments). A bandpass filter was used to filter out all frequencies >1 kHz (as suggested by M. Ritchie, University of St. Andrews, Saint Andrews, Scotland), and remaining sound was stored as a text file. Courtship chambers with attached microphones were fitted inside a large wooden box lined with acoustic panels to reduce ambient noise. As all 10 microphones recorded remaining ambient sounds simultaneously, we removed this noise by subtracting the median of all recordings from each recording at each time point. This and other downstream analyses were performed with custom scripts written in Python. We then searched each recording for sound pulses that corresponded to fly-generated songs (Figure 1). The pulse song produced by males during courtship was distinguished from other sounds using a set of filtering criteria based on manual annotation of initial song recordings (pulse length <2 msec, cycles/pulse <10 , pulse amplitude >1.0 mV, 5 or more consecutive pulses per pulse train, and IPI between 15 and 100 msec). Recording took place in an environmentally controlled room, $25 \pm 0.5^\circ$. As IPI varies linearly with temperature (~ 1.0 msec per degree centigrade; Ritchie and Kyriacou 1994), we used the temperature during each recording interval to standardize all measurements within the one degree of variation present.

Some authors have reported that the IPI varies in a sinusoidal fashion during courtship (the Kyriacou and Hall cycle, or KH cycle; Kyriacou and Hall 1980), with a period of ~ 1 min. This phenotype may be difficult for female flies to perceive, and its importance is therefore controversial (Ritchie *et al.* 1999; Tauber and Eberl 2002). To minimize any possible effect of a KH cycle on measured IPI, we calculated the mean IPI per male from a 4-min courtship bout (average 247 IPI events). It has also been extensively reported that males sometimes generate a sine song of undetermined significance; this trait was not apparent in a subset of manually inspected recordings and was not considered further.

Sequencing

We pooled the 120 most extreme individuals from each population across generations 10–14 for DNA extraction. For example, from short-IPI population 1, 9 flies from generation 10, 8 from generation 11, 10 from generation 12, 32 from generation 13, and 61 from generation 14 were pooled to create a population of individuals with very extreme trait values (values for all populations are provided in Table S1).

These flies were pulverized together in a cold mortar and pestle, and we extracted DNA using the DNeasy Blood and Tissue kit (Qiagen). A total of 1080 μ l of buffer ATL and 120 μ l of proteinase K were added to the pulverized tissue and mixed; this mixture was incubated at 56° for 4 hr, mixed, and distributed across six DNeasy spin columns. The remainder of the extraction followed the Qiagen spin-column protocol, with elution in water rather than buffer. Mixing of the six eluted water samples then resulted in a DNA sample with approximately equal representation of the 240 chromosomes in the initial sample of 120 flies. An Illumina paired-end library was created from each of the four population samples of DNA by the DNA Technologies Core Facility at the University of California Davis Genome Center. Each library was then sequenced twice: first in a single lane of the Illumina Genome Analyzer, using the SBS Sequencing kit v.5 for 85 cycles, and a second time with the HiSeq Flowcell v.1 for 100 cycles. In both cases, images were analyzed with Illumina software (RTA 1.8). We aligned sequencing reads (from fastq files) to the v5.26 *D. melanogaster* reference genome using the Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009). Alignments with mapping qualities <15 were discarded ($\sim 30\%$ of data from each population), and remaining reads were combined into pileup files.

We tabulated apparent polymorphisms (small indels and single nucleotide polymorphisms) by simply counting the location and frequency of mismatches relative to the reference genome. Only the two most common alleles were considered at any variable site with three or more apparent alleles. Sites with fewer than five reads in any of the four populations were also discarded, as allele frequency estimates at these sites are extremely poor. Mode read coverage at the remaining variable base pairs was between 192 times and 202 times for each of the four populations, with $<0.17\%$ of sites having <20 times coverage in each population (the largely repetitive, low recombination “Het” regions, Y chromosome, mitochondria, and fourth dot chromosome were not considered).

This procedure resulted in a list of apparent polymorphisms that includes both true polymorphisms and sequencing errors. Indeed, as the coverage of most bases is >100 times and the expected error rate from Illumina sequencing is $>1\%$, we would expect most sites to be variable. It was observed that 94.5% of bases with sufficient coverage were apparently polymorphic. However, 95.0% of these “variable” sites had experiment-wide allele frequencies of $<5.0\%$. By discarding all apparent polymorphisms with $<5\%$ global frequency, we simultaneously eliminated most sequencing errors and true variants that are rare in all populations. The exclusion of true rare variants is not an issue for the analysis, as these variants would not be significantly associated with treatment in any case. The 5.49 million remaining apparent polymorphisms undoubtedly include some errors as well, but this is also not a major concern. These errors should be randomly distributed, not at high frequency in any given population, and therefore not significantly associated

with treatment. We consider an anticonservative approach to calling polymorphisms to be conservative with respect to our experimental question. Because we correct for multiple comparisons, including some sequencing errors in our analysis should be conservative, in contrast, discarding true polymorphisms at high frequency may lead to overconfidence in the mapping precision of our significant associations.

Results

Artificial selection on courtship song

The IPI in males from the allRAL population was 35.3 ms: very close to the “typical” values previously reported for *D. melanogaster* (Ritchie *et al.* 1994). There was considerable variation, however, with the 1% tails of the starting population having IPIs <31.1 ms and >41.8 ms (for comparison, typical values reported for the sympatric species *D. simulans* are ~45–55 ms (Kawanishi and Watanabe 1980; Cowling and Burnet 1981). Artificial selection on this character was successful immediately: all selected populations were significantly different from the starting population after a single generation of selection (one-tailed *t*-tests: $8.2E-3 < P < 1.3E-10$). Selection continued for an additional 13 generations, with the change in mean IPI showing a steady and remarkably consistent response to selection (Figure 2). By the time selection was halted, the mean of each population was more extreme than the 1% tails of the starting population (Figure 3). The average of the two short-IPI populations moved 2.25 and 2.38 standard deviations, while the two long-IPI populations moved even more: 2.99 and 3.47 standard deviations (where the standard deviation is taken from the starting population). The standard deviation also increased in populations selected for longer IPI relative to the starting population, while it decreased slightly in the short-IPI populations (starting SD = 2.4 ms, short-IPI SD = 2.0 and 2.1 ms, long-IPI SD = 4.5 and 5.2 ms). Because these distributions were significantly nonnormal on the basis of Shapiro–Wilk tests, we tested whether these standard deviations were significantly different from the starting population by bootstrapping each population 100,000 times: the short-IPI populations were not significantly different than the starting population, but the long-IPI populations had significant increases in variance ($P < 10E-5$ for both).

Sequencing of artificially selected individuals

To maximize our power to detect allele frequency differences at causal loci, we extracted DNA from the 120 individuals from each population with the most extreme differences in trait values. Average IPIs for these sequenced samples were 27.1 ms for both short-IPI populations and 52.0 and 52.6 for the long-IPI populations. This DNA was extracted *en masse* for each population and sequenced with the Illumina Genome Analyzer. Population-based resequencing of each of these samples allowed genetic variants to be located throughout the genome and the frequency of each variant to be estimated in each population (see *Methods*).

Figure 2 Change in mean interpulse interval (IPI) over time. Populations selected to have shorter IPIs are shown as circles, and populations selected to have longer IPIs are shown as triangles, with replicate populations shown as solid and open symbols.

After discarding all variants with overall frequencies <5% (which are mostly errors), 5.49 million variable sites remained. The average sequence coverage of each variable site was ~200-fold in each population, meaning that 200 chromosomes were sampled on average in each population (with replacement, from a sample of 240 chromosomes). With <0.2% of variable sites having <20-fold coverage in each population, error in estimating allele frequency should be modest at most sites.

Analysis of genomic differentiation

The primary effect of selection on phenotypes is to change the underlying allele frequency at causal variants. To locate these variants, differences in allele frequencies between treatments should be compared to an expected distribution of allele frequency differences between populations. This expected distribution results from genetic drift during the experiment, sampling error during sequencing, and possibly other factors. The amount of genetic drift depends on the effective population size, which is difficult to estimate precisely. We therefore used the allele frequency distribution between populations selected in the same direction as an empirical null distribution that simultaneously controls for all factors. The distribution of allele frequency differences across the genome between the two short-IPI populations and between the two long-IPI populations should estimate the cumulative effects of all factors except selection related to treatment. The average allele frequency difference at all 5.48 million variable sites is indeed slightly greater between vs. within treatments, indicating an apparent effect of selection (short-IPI 1 vs. short-IPI 2 = 7.16%, long-IPI 1 vs. long-IPI 2 = 7.37%, short-IPI 1 vs. long-IPI 1 = 9.47%, and short-IPI 2 vs. long-IPI 2 = 9.65%). This comparison should be conservative, as gene flow was allowed between populations selected in opposite directions only, resulting in a neutral expectation of slightly higher relatedness between treatments than within.

To quantitatively compare the distribution of allele frequency differences within and between treatments, we used a graphical approach (Figure 4). For each variable base

Figure 3 Distributions of interpulse interval (IPI) for populations selected for shorter values (red) and longer values (blue); the density trace for the starting population is shown for comparison. (A) Replicate 1; (B) replicate 2.

in the genome, we first plotted the amount of evolution between populations selected in the same direction. There are two such comparisons: allele frequency difference between the two short populations and the allele frequency difference between the two long populations. The blue points in Figure 4 show the distribution of differentiation in these two dimensions. Most of the variation lies in the center of the plot, with $\sim 75\%$ of the variation in the genome displaying $<10\%$ differentiation in either dimension and $>90\%$ of the variation displaying $<20\%$ differentiation. More important is the behavior of the outliers: though some variants are highly differentiated between the two replicate short-selected populations, these outliers are not more likely to be differentiated between the two long-selected populations. The blue points therefore form an approximately circular scatter plot in these two dimensions. However, when a similar plot is made for differentiation *between* treatments, the shape is quite different. Once again, the vast majority of the data are in the center of the plot, with over two-thirds of the variation displaying $<10\%$ differentiation, and $>85\%$ displaying $<20\%$ differentiation. The variants that are differentiated between one short vs. long comparison are, however, much more likely to be differentiated between the other short vs. long comparison than other variants: thus the circle has become distended along the diagonal. Assessing the probability that any one of these variants has been affected by selection is somewhat challenging due to variable allele frequency, sequencing coverage, and nonindependence between individual variants. We therefore used the blue points as an empirical null hypothesis to determine the proportion of differentiation expected from all factors other than treatment.

As the comparison of within-treatment frequencies should encompass all effects except selection (and linked selection) related to treatment, we can calculate the false discovery rate (FDR) for selection for any region of this plot. This is accomplished by simply dividing the number of variants in that region within treatments by the number of variants in that region between treatments. In Figure 4, points within the gray region are those where the sum of

the two axes is greater (or less than) 1 (-1). In this region there are 13,343 variants between treatments, but only 56 variants within treatments: FDR estimate is therefore $56/13,343 = 0.42\%$. The variants in this region were considered significant in later analyses. As an additional control, the corresponding regions in the other corners of the plot can be compared: these regions are highlighted in brown. Here, no excess of selection related to treatment is expected, unless the response to selection was on different alleles in the different replicates. The FDR estimate in these regions is $130/127$, or almost exactly 100% .

Figure 5 shows the distribution of differentiation along each chromosome. The y-axis statistic is the sum of allele frequency differences between the two replicate comparisons ($\text{abs}(\text{short-IPI population 1} - \text{long-IPI population 1} + \text{short-IPI population 2} - \text{long-IPI population 2})$). Casual visual inspection reveals that genetic variants significantly affected by selection are spread throughout much of the genome. There are, however, fewer significant variants on the X chromosome (85 or 0.008% of variants) than on any of the autosomal arms (2966–8286 or $0.26\text{--}0.75\%$ of variants). The proportion of significant variants in 10,000 bootstrap replicates was always an order of magnitude higher on the autosomes than the X, supporting the significance of this underrepresentation on the X.

As the starting population contained an approximately equal representation of the 173 RAL lines, we can use the released genome sequences from these lines to infer the starting frequency of individual variants. The Drosophila Population Genomic Project (DPGP) at University of California Davis has sequenced 36 of these RAL lines, each to ~ 10 times coverage. Of the 5.49 million polymorphisms with $>5\%$ global frequency in our current experiment, 55% were also found to be variable in the 36 v.1 DPGP RAL assemblies (downloaded from dpgp.org). The loci that were not variable in the DPGP release likely reflect some sequencing errors, but may also reflect: (i) sites where unique alignments were not possible using the unpaired 36-bp reads used in DPGP; (ii) indels (gapped alignments were not attempted in the DPGP release); or (iii) sites that

Figure 4 Sauron plot of genetic differentiation between populations. Each variable site in the genome is shown as a point, with the axes showing the allele frequency difference between two independent comparisons of the four populations. Blue points are comparisons between populations selected in the same direction, with the short-selected populations on the x-axis and the long-selected populations on the y-axis. Red points are comparisons between populations selected in opposite directions. Histograms are also shown, with blue histograms corresponding to within-treatment comparisons and red histograms to between-treatment comparisons. The gray regions of the plot are regions referred to in the text.

are not variable in the 36 genomes, but are variable in the 173 genomes. For the variants found in both data sets, we used the DPGP genomes to estimate the frequency of each variant in the starting population. As expected under selection, drift, or sampling error, variants with lower starting frequencies were less differentiated on average (Wilcox $P < 10E-15$; average pairwise differentiation of variants $<5\%$, $5-10\%$, and $>10\% = 0.09, 0.10, 0.14$).

Discussion

Phenotypes

Our strong phenotypic response to selection demonstrates that there is considerable additive variation in the interpulse interval of *D. melanogaster* courtship song. This is in contrast with an earlier study that was able to experimentally evolve populations with longer IPI, but not shorter IPI (Ritchie and Kyriacou 1996). Our ability to successfully select in both directions may be due to our larger population sizes and/or different starting populations compared to those used in the previous study. It is worth noting, however, that we saw a larger response to selection in the direction of longer IPI, which is partially consistent with previous results (Ritchie and Kyriacou 1996). We also found increased standard deviation in the long-IPI populations compared to the starting population. A simple explanation for this result is that, at the majority of causal sites, the long-IPI allele was less common than the short-IPI allele in the starting population. Under

this hypothesis, the amount of genetic variation for IPI would decrease in the short-selected populations as these long alleles became rare, but would increase in the long-IPI populations as they approached intermediate frequency. The response to selection between replicates in the same direction was remarkably consistent, which may be due in part to the gene flow permitted between replicates. Though gene flow was only permitted between replicates selected in opposite directions, alleles could have passed from short-IPI population 1 to short-IPI population 2 by first journeying through long-IPI populations.

Genotypes

We were successful in resequencing our artificially selected populations, and technological advances allowed an order of magnitude more genome coverage than previous experiments (Burke *et al.* 2010; Turner *et al.* 2010, 2011). By comparing the differentiation between populations selected in the same direction with those selected in opposite directions, we can determine loci affected by selection without estimating effective population size, sampling variance at individual alleles, or other complicating factors. Using this method, we show that the evolve and resequence approach has considerable power, without the need to select on a large number of experimental lines over many generations. With only four populations selected for 14 generations, the E&R approach was successful. The response to selection at a genetic level was also highly repeatable between replicates (Figure 4).

Though our power was good ($>13,000$ variants at a 0.5% FDR), there was little spatial clustering of significant variants along chromosomes (Figure 5). Highly significant variants were distributed across all four autosomal chromosome arms, with a smattering on the X chromosome as well. This result is in contrast to a previous experiment on body size variation in the same species (Turner *et al.* 2011). Though a large number of loci were also differentiated in the previous experiment, some appeared to form discrete “peaks” at many loci. Proving that these peaks were independently differentiated was difficult, but the considerable enrichment of candidate genes at differentiated loci supported this interpretation. In our current experiment, a corresponding gene ontology analysis is not feasible, both because of the diffuse nature of differentiation and the lack of candidate genes for IPI.

There are multiple possible explanations for the differences in mapping resolution between these two experiments. One possibility is that variation in IPI is caused by variants at a much larger number of loci than for body size. This would be surprising: we might expect body size to be a highly composite trait, potentially affected by a huge number of factors, while the IPI of courtship song might be a much more specific (if still quantitative) phenotype. This speculation is partially supported by the large number of genes known to affect body size (Carreira *et al.* 2009) and the corresponding lack of genes known to affect IPI (Gleason

Figure 5 Differentiation along chromosome arms. Each variable site in the genome is shown as a point, with colors corresponding to estimates of starting allele frequency (black, unknown; blue, 10–50%; purple, 5–10%; and red, <5%). Note that points were plotted starting with black and ending with red, with subsequent points obscuring many of the black, blue, and purple points. Y-axis is the association statistic: $\text{abs}(\text{short-IPI population 1} - \text{long-IPI population 1} + \text{short-IPI population 2} - \text{long-IPI population 2})$ with the FDR < 0.42% threshold, corresponding to the gray region in Figure 4, shown as a black line.

2005), though differences in research effort are difficult to quantify. Alternatively, if IPI is a sexually selected trait (Snook *et al.* 2005), and standing variation in sexually selected traits is caused by overall variation in condition, IPI may be extremely polygenic (Rowe and Houle 1996). A haplotype-based analysis might shed light on this question by shedding light on the number of loci in the genome that are independently differentiated, which is a minimum value for the number of causal alleles. The lack of haplotype information is one weakness of the pooled-sequencing approach we employed, but in this case haplotype information is available due to the nature of our starting population. The collection of RAL inbred lines used to create a base population has now been sequenced by the *Drosophila* Genetic Reference Panel (Mackay *et al.* 2012), and we are now attempting to use these individual genome projects to conduct haplotype-based analysis of our data.

An alternative possibility for the lack of mapping resolution is that selection was stronger in the IPI experiment, leading to lower mapping resolution. It is notable that few of the differentiated loci in the body size experiment were estimated to have been rare in the starting population. In contrast, many apparently rare variants are differentiated in large blocks in our current IPI experiment. Though the fraction of individuals selected was similar in each case (~17% per population in the body size experiment and ~20% per population in the IPI experiment), differences

in heritability and number of causal alleles could lead to stronger selection on individual variants in the IPI experiment. It may also be important that, in the current experiment, a very extreme subset of individuals were sequenced, rather than the entire population. This is equivalent to doing a single generation of very strong selection (taking a fine slice of the distribution) and may increase power to map uncommon variants. The ability to map uncommon alleles may be both an advantage and weakness of the E&R approach compared to GWAS. Though linkage disequilibrium is generally low in *D. melanogaster* (Mackay *et al.* 2012), selection on low frequency variants will cause differentiation of long stretches of linked DNA (hard sweeps; Maynard Smith and Haigh 1974), unlike selection on common variation (soft sweeps; Hermisson and Pennings 2005). Selection on a few uncommon alleles may lead to stochastic differentiation of many other loci and dramatically reduce the ability to resolve common alleles: selection on a rare allele near Mb 15 on 2L, for example, may be responsible for much of the differentiation on this chromosome arm (Figure 5). If this is the case, deeply sequencing populations selected under a weaker selection program for a short duration may yield more interpretable results. As there are many other differences between these two experiments (*e.g.*, haplotype diversity in the starting population, duration of selection, amount of resource competition during the experiment), precisely explaining the divergent results may require additional

experiments. These contrasting results also clearly indicate the need for theoretical development of this technique. We hope that the apparent success of the body size experiment, combined with the considerable power demonstrated in the much shorter song experiment, inspire and inform these future efforts.

The dramatic difference in the number of selected variants on the autosome vs. the X chromosome was surprising: only males were selected in our populations and the hemizygous nature of the male X chromosome should allow more power to select on recessive and codominant alleles if they are present. If most differentiation on autosomes is due to a few rare alleles, this result may be simply due to chance. However, it is worth noting that there is a deficit of genes with male-biased gene expression on the X chromosome in *D. melanogaster* (Parisi *et al.* 2003; Ranz *et al.* 2003). The reasons for this deficit are not yet entirely clear (Kaiser and Bachtrog 2010), but if it turns out that functional variation affecting male-specific sexually selected traits is underrepresented on the X, this may shed light on the causes of this deficit. Though drawing a firm conclusion is as yet premature, this result is a tantalizing example of insights that could be attained if methods of high-throughput mapping of genotype to phenotype are refined in model systems.

Acknowledgments

We are grateful for the help from the Barry Dickson lab at the Institute for Molecular Pathology; the advice of Michael Ritchie at the University of St. Andrews, St. Andrews, UK; and to the University of California, Santa Barbara, undergraduate research team of Veronica Cochrane, Raisa Mobin, Madeline Evancie, Joseph Sandoval, Jothisna Bodhanapati, and Zach Edwin who recorded many a fly singing. Funding was provided by UCSB, by the National Institutes of Health (R01 GM098614), and by National Science (NSF) Foundation support to the Center for Scientific Computing at UCSB (NSF MRSEC DMR-1121053 and NSF CNS-0960316).

Literature Cited

- Allen, H. L., K. Estrada, G. Lettre, S. I. Berndt, M. N. Weedon *et al.*, 2010 Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467: 832–838.
- Atwell, S., Y. S. Huang, B. J. Vilhjalmsen, G. Willems, M. Horton *et al.*, 2010 Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.
- Ayroles, J. F., M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman *et al.*, 2009 Systems genetics of complex traits in *Drosophila melanogaster*. *Nat. Genet.* 41: 299–307.
- Bennet-Clark, H. C., and A. W. Ewing, 1969 Pulse interval as a critical parameter in the courtship song of *Drosophila melanogaster*. *Anim. Behav.* 17: 755–759.
- Blyth, J. E., D. Lachaise, and M. G. Ritchie, 2008 Divergence in multiple courtship song traits between *Drosophila santomea* and *D. yakuba*. *Ethology* 114: 728–736.
- Boyko, A. R., P. Quignon, L. Li, J. J. Schoenebeck, J. D. Degenhardt *et al.*, 2010 A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* 8: e1000452.
- Burke, M. K., J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* 467: 587–590.
- Carreira, V. P., J. Mensch, and J. J. Fanara, 2009 Body size in *Drosophila*: genetic architecture, allometries and sexual dimorphism. *Heredity* 102: 246–256.
- Cowling, D. E., and B. Burnet, 1981 Courtship songs and genetic control of their acoustic characteristics in sibling species of the *Drosophila melanogaster* subgroup. *Anim. Behav.* 29: 924–935.
- Doi, M., M. Matsuda, M. Tomaru, H. Matsubayashi, and Y. Oguma, 2001 A locus for female discrimination behavior causing sexual isolation in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 98: 6714–6719.
- Earley, E. J., and C. D. Jones, 2011 Next-generation mapping of complex traits with phenotype-based selection and introgression. *Genetics* 189: 1203–1209.
- Falconer, D. S., 1981 *Introduction to Quantitative Genetics*, Ed. 2. Longman, New York.
- Fiumera, A. C., B. L. Dumont, and A. G. Clark, 2007 Associations between sperm competition and natural variation in male reproductive genes on the third chromosome of *Drosophila melanogaster*. *Genetics* 176: 1245–1260.
- Gleason, J. M., 2005 Mutations and natural genetic variation in the courtship song of *Drosophila*. *Behav. Genet.* 35: 265–277.
- Gleason, J. M., and M. G. Ritchie, 1998 Evolution of courtship song and reproductive isolation in the *Drosophila willistoni* species complex: Do sexual signals diverge the most quickly? *Evolution* 52: 1493–1500.
- Hermisson, J., and P. S. Pennings, 2005 Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169: 2335–2352.
- Hill-Burns, E. M., and A. G. Clark, 2009 X-Linked variation in immune response in *Drosophila melanogaster*. *Genetics* 183: 1477–1491.
- Hindorf, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta *et al.*, 2009 Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* 106: 9362–9367.
- Huang, X. H., X. H. Wei, T. Sang, Q. A. Zhao, Q. Feng *et al.*, 2010 Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42: 961–967.
- Inomata, N., H. Goto, M. Itoh, and K. Isono, 2004 A single amino acid change of the gustatory receptor gene, Gr5a, has a major effect on trehalose sensitivity in a natural population of *Drosophila melanogaster*. *Genetics* 167: 1749–1758.
- Johansson, A. M., M. E. Pettersson, P. B. Siegel, and Ö. Carlborg, 2010 Genome-wide effects of long-term divergent selection. *PLoS Genet.* 6: e1001188.
- Kaiser, V. B., and D. Bachtrog, 2010 Evolution of sex chromosomes in insects. *Annu. Rev. Genet.* 44: 91–112.
- Kawanishi, M., and T. K. Watanabe, 1980 Genetic variations of courtship song of *Drosophila melanogaster* and *Drosophila simulans*. *Jpn. J. Genet.* 55: 235–240.
- Kyriacou, C. P., and J. C. Hall, 1980 Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. *Proc. Natl. Acad. Sci. USA* 77: 6729–6733.
- Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 25: 1754–1760.
- Long, A. D., and C. H. Langley, 1999 The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Res.* 9: 720–731.
- Lyman, R. F., C. Lai, and T. F. Mackay, 1999 Linkage disequilibrium mapping of molecular polymorphisms at the scabrous locus associated with naturally occurring variation in bristle number in *Drosophila melanogaster*. *Genet. Res.* 74: 303–311.

- Mackay, T. F., and C. H. Langley, 1990 Molecular and phenotypic variation in the achaete-scute region of *Drosophila melanogaster*. *Nature* 348: 64–66.
- Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* genetic reference panel. *Nature* 482: 173–178.
- Makowsky, R., N. M. Pajewski, Y. C. Klimentidis, A. I. Vazquez, C. W. Duarte *et al.*, 2011 Beyond missing heritability: Prediction of complex traits. *PLoS Genet.* 7: e1002051.
- Manenti, G., A. Galvan, A. Pettinicchio, G. Trincucci, E. Spada *et al.*, 2009 Mouse genome-wide association mapping needs linkage analysis to avoid false-positive loci. *PLoS Genet.* 5: e1000331.
- Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. *Genet. Res.* 23: 23–35.
- McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little *et al.*, 2008 Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9: 356–369.
- Nuzhdin, S. V., L. G. Harshman, M. Zhou, and K. Harmon, 2007 Genome-enabled hitchhiking mapping identifies QTLs for stress resistance in natural *Drosophila*. *Heredity* 99: 313–321.
- Odgers, W. A., M. J. Healy, and J. G. Oakeshott, 1995 Nucleotide polymorphism in the 5' promoter region of esterase 6 in *Drosophila melanogaster* and its relationship to enzyme activity variation. *Genetics* 141: 215–222.
- Parisi, M., R. Nuttall, D. Naiman, G. Bouffard, J. Malley *et al.*, 2003 Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* 299: 697–700.
- Park, J. H., S. Wacholder, M. H. Gail, U. Peters, K. B. Jacobs *et al.*, 2010 Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat. Genet.* 42: 570–575.
- Parts, L., F. Cubillos, J. Warringer, K. Jain, F. Salinas *et al.*, 2011 Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21: 1131–1138.
- Ranz, J. M., C. I. Castillo-Davis, C. D. Meiklejohn, and D. L. Hartl, 2003 Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* 300: 1742–1745.
- Ritchie, M. G., and C. P. Kyriacou, 1994 Genetic variability of courtship song in a population of *Drosophila melanogaster*. *Anim. Behav.* 48: 425–434.
- Ritchie, M. G., and C. P. Kyriacou, 1996 Artificial selection for a courtship signal in *Drosophila melanogaster*. *Anim. Behav.* 52: 603–611.
- Ritchie, M. G., V. H. Yate, and C. P. Kyriacou, 1994 Genetic variability of the interpulse interval of courtship song among some European populations of *Drosophila melanogaster*. *Heredity* 72: 456–464.
- Ritchie, M. G., R. M. Townhill, and A. L. Hoikkala, 1998 Female preference for fly song: playback experiments confirm the targets of sexual selection. *Anim. Behav.* 56: 713–717.
- Ritchie, M. G., E. J. Halsey, and J. M. Gleason, 1999 *Drosophila* song as a species-specific mating signal, and the behavioural importance of Kyriacou & Hall cycles in *D. melanogaster* song. *Anim. Behav.* 58: 649–657.
- Rowe, L., and D. Houle, 1996 The lek paradox and the capture of genetic variance by condition dependent traits. *Proc. Biol. Sci.* 263: 1415–1421.
- Shorey, H. H., 1962 Nature of sound produced by *Drosophila melanogaster* during courtship. *Science* 137: 677–678.
- Snook, R. R., A. Robertson, H. S. Crudgington, and M. G. Ritchie, 2005 Experimental manipulation of sexual selection and the evolution of courtship song in *Drosophila pseudoobscura*. *Behav. Genet.* 35: 245–255.
- Spieth, H. T., 1974 Courtship behavior in *Drosophila*. *Annu. Rev. Entomol.* 19: 385–405.
- Talyn, B. C., and H. B. Dowse, 2004 The role of courtship song in sexual selection and species recognition by female *Drosophila melanogaster*. *Anim. Behav.* 68: 1165–1180.
- Tauber, E., and D. F. Eberl, 2002 The effect of male competition on the courtship song of *Drosophila melanogaster*. *J. Insect Behav.* 15: 109–120.
- Tomaru, M., and Y. Oguma, 2000 Mate choice in *Drosophila melanogaster* and *D. sechellia*: criteria and their variation depending on courtship song. *Anim. Behav.* 60: 797–804.
- Tomaru, M., H. Matsubayashi, and Y. Oguma, 1998 Effects of courtship song in interspecific crosses among the species of the *Drosophila auraria* complex. *J. Insect Behav.* 11: 383–398.
- Tomaru, M., M. Doi, H. Higuchi, and Y. Oguma, 2000 Courtship song recognition in the *Drosophila melanogaster* complex: heterospecific songs make females receptive in *D. melanogaster*, but not in *D. sechellia*. *Evolution* 54: 1286–1294.
- Turner, T. L., E. C. Bourne, E. J. von Wettberg, T. T. Hu, and S. Nuzhdin, 2010 Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat. Genet.* 42: 260–263.
- Turner, T. L., A. D. Stewart, A. T. Fields, W. R. Rice, and A. M. Tarone, 2011 Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in *Drosophila melanogaster*. *PLoS Genet.* 7: e1001336.
- van den Berg, M. J., 1988 *Assortative Mating in Drosophila melanogaster and Among Three Species of the melanogaster Subgroup*, pp. University of Groningen, The Netherlands.
- von Philipsborn, A. C., T. Liu, J. Y. Yu, C. Masser, S. S. Bidaye *et al.*, 2011 Neuronal control of *Drosophila* courtship song. *Neuron* 69: 509–522.
- von Schilcher, F., 1976 The role of auditory stimuli in the courtship of *Drosophila melanogaster*. *Anim. Behav.* 24: 18–26.
- Wheeler, D. A., C. P. Kyriacou, M. L. Greenacre, Q. Yu, J. E. Rutila *et al.*, 1991 Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. *Science* 251: 1082–1085.
- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42: 565–569.
- Yang, J., T. A. Manolio, L. R. Pasquale, E. Boerwinkle, N. Caporaso *et al.*, 2011 Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* 43: 519–525.
- Zhou, D., N. Udpa, M. Gersten, D. W. Visk, A. Bashir *et al.*, 2011 Experimental selection of hypoxia-tolerant *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 108: 2349–2354.

Communicating editor: F. F. Pardo Manuel de Villena

GENETICS

Supporting Information

<http://www.genetics.org/content/suppl/2012/03/30/genetics.112.139337.DC1>

Investigating Natural Variation in *Drosophila* Courtship Song by the Evolve and Resequencing Approach

Thomas L. Turner and Paige M. Miller

Table S1

Supporting Data

Table 1 is available for download at <http://www.genetics.org/content/suppl/2012/03/30/genetics.112.139337.DC1> as an Excel file.