

Composite Effects of Polymorphisms near Multiple Regulatory Elements Create a Major-Effect QTL

Ryan D. Bickel^{1,2*}, Artyom Kopp³, Sergey V. Nuzhdin²

1 School of Biological Science, University of Nebraska – Lincoln, Lincoln, Nebraska, United States of America, **2** Program in Molecular and Computational Biology, Department of Biological Sciences, University of Southern California, Los Angeles, California, United States of America, **3** Department of Evolution and Ecology, University of California Davis, Davis, California, United States of America

Abstract

Many agriculturally, evolutionarily, and medically important characters vary in a quantitative fashion. Unfortunately, the genes and sequence variants accounting for this variation remain largely unknown due to a variety of biological and technical challenges. *Drosophila melanogaster* contains high levels of sequence variation and low linkage disequilibrium, allowing us to dissect the effects of many causative variants within a single locus. Here, we take advantage of these features to identify and characterize the sequence polymorphisms that comprise major effect QTL alleles segregating at the *bric-a-brac* locus. We show that natural *bric-a-brac* alleles with large effects on cuticular pigmentation reflect a cumulative impact of polymorphisms that affect three functional regions: a promoter, a tissue-specific enhancer, and a Polycomb response element. Analysis of allele-specific expression at the *bric-a-brac* locus confirms that these polymorphisms modulate transcription at the *cis*-regulatory level. Our results establish that a single QTL can act through a confluence of multiple molecular mechanisms and that sequence variation in regions flanking experimentally validated functional elements can have significant quantitative effects on transcriptional activity and phenotype. These findings have important design and conceptual implications for basic and medical genomics.

Citation: Bickel RD, Kopp A, Nuzhdin SV (2011) Composite Effects of Polymorphisms near Multiple Regulatory Elements Create a Major-Effect QTL. PLoS Genet 7(1): e1001275. doi:10.1371/journal.pgen.1001275

Editor: Gregory S. Barsh, Stanford University, United States of America

Received: August 20, 2010; **Accepted:** December 13, 2010; **Published:** January 13, 2011

Copyright: © 2011 Bickel et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the NSF (DEB - 0548991) and NIH (RGM076643). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rbickel2@unl.edu

Introduction

Quantitative trait loci (QTLs) have traditionally been identified by linkage analysis in experimental crosses [1]. In a complementary approach, DNA polymorphisms causing trait variation can be inferred from statistical associations between sequence and phenotype in large panels of natural genotypes [2]. These methods have been advanced with creative applications [3,4], but in most cases QTL regions have only been narrowed to a single gene. Understanding the identity, distribution, and functional nature of nucleotide polymorphisms that make up QTL alleles remains a critical challenge in molecular and evolutionary genetics. In this report, we take a QTL study to its logical conclusion by describing the fine structure of functional variation that underlies large-effect QTL alleles segregating in a natural population.

To meet this aim, we sequenced the 148 kb *bric a brac* (*bab*) genomic region from 94 inbred lines of *D. melanogaster* sampled from a single location. The *bab* locus is composed of two anciently duplicated paralogs, *bab1* and *bab2*, which are involved in patterning the adult abdomen, legs, and ovaries [5]. In particular, *Bab* expression in the pupal abdominal epidermis is responsible for the sex-specific pigmentation of adult cuticle, and evolutionary changes in *bab* regulation have played a central role in the macroevolutionary diversification of *Drosophila* color patterns [6–8]. Moreover, segregation of *bab* alleles is responsible for ~60% of the genetic variation in female abdominal pigmentation within a natural population of *D. melanogaster* [9] (Figure 1). Linkage disequilibrium (LD) at the *bab* locus

extends for less than 1 kb, [10] providing an opportunity to dissect the causative nucleotide variation that contributes to the cumulative effect of QTL alleles. Previous analysis revealed evidence of directional selection in two regions within the *bab* locus, including the previously identified *cis*-regulatory element (CRE) that controls sex-specific abdominal pigmentation [8,10]. This confluence of population- and developmental-genetic evidence sets the stage for connecting specific DNA polymorphisms to variation in gene expression and, ultimately, in phenotypic traits. We have sequenced the *bab* locus from 96 *D. melanogaster* lines to identify the genetic variation responsible for pigmentation differences.

Results/Discussion

Causative polymorphisms cluster in three functional regions

We identified 6766 sequence variants, including 5566 single nucleotide polymorphisms (SNPs) and 1211 insertion/deletion (indel) polymorphisms, in the *bab* region. After removing polymorphisms with frequencies below 5%, which are difficult to associate with phenotypic variation, 3821 polymorphism (SNPs and indels) remained. We tested each polymorphism individually for an association with female pigmentation, and identified 266 polymorphisms that showed significant association at the false discovery rate (FDR) <0.4 ($p < .033$) (Spearman rank correlation, *q*-value correction) [11] (Figure 2A). All associated polymorphisms are located in non-coding DNA.

Author Summary

Genes responsible for quantitative variation have been identified for a diverse range of phenotypes. However, much remains to be learned about the distribution of causative genetic variation within a locus. In this study, we investigated a locus that contributes to natural variation in abdominal pigmentation in *Drosophila melanogaster*. We found that the large phenotypic effect of this locus results from the cumulative action of many small-effect polymorphisms that are concentrated in three distinct functional regions: a promoter, a tissue-specific enhancer, and a Polycomb response element (a region involved in chromatin remodeling). The same regions influence the adult phenotype and transcript abundance, indicating that the causative sequence variants act by modulating transcription. Interestingly, these polymorphisms cluster near, but not within, the functionally validated regulatory regions, suggesting that DNA sequences surrounding core functional elements may play a key role in quantitative variation.

To determine whether the significant polymorphisms were randomly distributed across the locus or clustered into functional regions, we looked at the proportion of associated polymorphisms in 5 kb sliding windows (Figure 2B) (see Materials and Methods for details). We found that most associated polymorphisms are located in three clusters: a 15 kb region in the first intron of *bab1*, an 8 kb region between the *bab1* and *bab2* coding regions, and a 9.5 kb region near the transcription start site of *bab2*.

These clusters of significantly associated points could be the result of a high number of causative polymorphisms or alternatively a single causative polymorphism with high levels of LD resulting in a cluster of associated polymorphisms. We examined the pairwise levels of LD for significant polymorphisms within each region (see Figure S1). Based on the LD pattern, we conservatively estimate that there are at least 28 associated polymorphisms, or groups of polymorphisms, assorting independently in region 1, 8 polymorphisms in region 2 and 14 polymorphisms in region 3. Thus, there are multiple polymorphisms within each region that are likely to be causative variation.

The first region resides within the large first intron of *bab1* and contains the *cis*-regulatory element that drives expression in the epidermis of the last three abdominal segments in females [8] (red bar in the blowup of region 1 in Figure 2A). This enhancer, which is the only identified sex-specific enhancer in the *bab* locus, is thought to control the sex-specific expression of both *bab1* and *bab2* [8]. This enhancer contains functionally characterized binding sites for the Abdominal-B (Abd-B) and doublesex (*dsx*) transcription factors [8]. None of the significantly associated



Figure 1. Three representative females (lines W3-123, W3-112, and W3-135) showing the range of pigmentation variation. The arrow indicates A6, the segment analyzed.
doi:10.1371/journal.pgen.1001275.g001

polymorphisms affect these binding sites, and indeed there are no common polymorphisms in the 624 bp core enhancer. However, the variation surrounding this location is strongly associated with pigmentation variation (Figure 2A, 2B). Macroevolutionary changes in *bab* expression are largely due to sequence changes within the core enhancer but outside of the Abd-B and *Dsx* binding sites, indicating that CREs present large targets for functionally relevant mutations [8]. Our results reveal an even larger mutational target by demonstrating that sequence variation surrounding the core CRE can result in significant phenotypic variation without disrupting the primary function of the enhancer. This region also shows an excess of high-frequency derived alleles [10], suggesting that genetic variation in the region flanking the sex-specific abdominal CRE has been shaped by directional selection.

The second identified region is located in the intergenic region between *bab1* and *bab2* and contains a predicted and experimentally verified Polycomb response element (PRE) [12–14] (red bar in the blowup of region 2, Figure 2A). PREs regulate transcriptional activity over multiple cell generations by affecting chromatin state. As in the CRE region, the distribution of significantly associated polymorphisms is considerably wider than the functionally characterized PRE, with most effects located outside of the core region. These results suggest that PREs, as well as tissue-specific enhancers, present a large target for functionally and evolutionarily important mutations.

The third region showing a high concentration of significantly associated polymorphisms spans the promoter and transcription start site of *bab2* (red bar in the blowup of region 3, Figure 2A). While promoters have been traditionally thought of as passive players in transcription, recent research suggests that promoters are often active in transcriptional regulation [15]. Several studies have shown that polymerases remain stalled on developmentally important genes, poised to begin transcription [16,17]. This pattern is observed on the *bab2* promoter during embryonic development and in S2 cells [18], while the pupal stages are yet to be investigated. Our results suggest that not only are promoter regions involved in transcriptional regulation, but that sequence variation in these regions can have a significant impact on gene expression and phenotype. It is intriguing that no similar concentration of associated SNPs is observed around the *bab1* promoter (Figure 2A, 2B). We believe that the explanation for this discrepancy may lie in the differences between the transcriptional activity of *bab1* and *bab2* at the pupal stage (see below).

In addition to these three regions, there appears to be an excess of associated polymorphisms near the downstream exons of *bab2*. However, this observation can be attributed to cryptic population structure. The lines that comprise our sample were collected in two different years. The region of highly significant F_{ST} between the two collections [10] coincides with the region of apparent association near the 3' exons of *bab2* (Figure 2B). When the samples from different years are examined separately, no association is observed in this region (Figure 3A), suggesting that it contains no causative polymorphisms.

The existence of females with completely pigmented sixth abdominal segment distinguishes *D. melanogaster* from its close relatives *D. simulans*, *D. mauritiana*, and *D. yakuba*. To test whether derived SNP alleles are associated with darker phenotypes, we polarized DNA polymorphisms in the *bab* region using the genome of *D. simulans*. The 107 SNPs associated with phenotypic variation at $FDR < .2$ (Figure 3A) show a marginal bias towards derived alleles having a darker phenotype (53 of 76, 2-tailed Fisher's exact test $P = 0.034$; 31 polymorphisms not polarized).

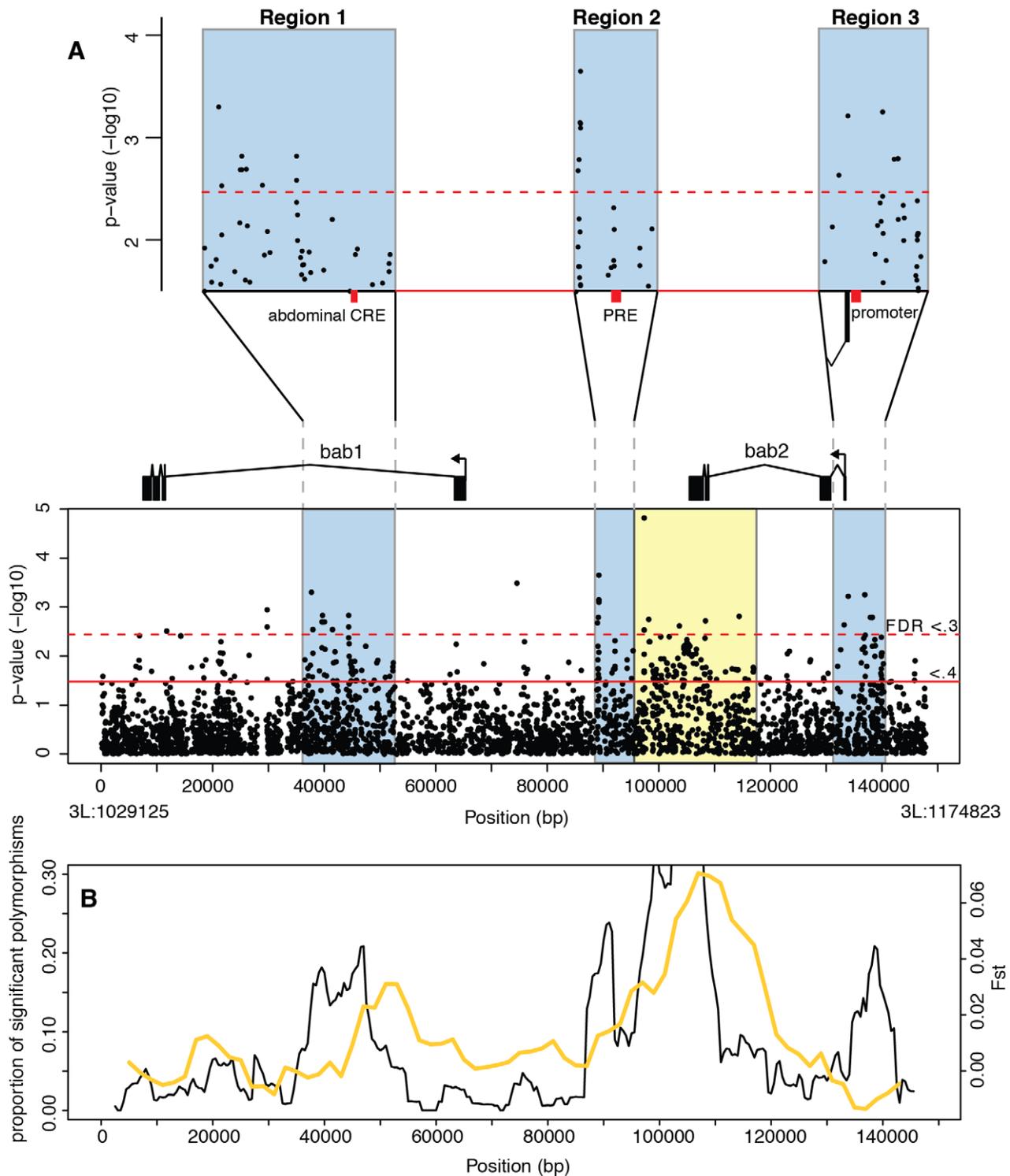


Figure 2. Association mapping of female A6 pigmentation at the *bab* locus. (A) P-values from a Spearman rank correlation for each polymorphism identified in the *bab* locus. The dotted and solid red lines indicate FDR values of 0.3 and 0.4 respectively. 3 functional regions containing most associated polymorphisms (shaded blue rectangles) are shown in greater detail above. The sex-specific abdominal CRE (in region 1), Polycomb response element (region 2), and the *bab2* promoter (region 3) are indicated by red bars. The yellow region contains an excess of linked polymorphisms, but is attributed to cryptic structure in our samples. (B) Black line shows the proportion of all polymorphisms in 5 kb sliding windows that are significantly associated with variation in pigmentation (FDR < .4), plotted against the midpoint of the window. Overlaid in yellow is F_{st} between samples collected in separate years (see text for details). doi:10.1371/journal.pgen.1001275.g002

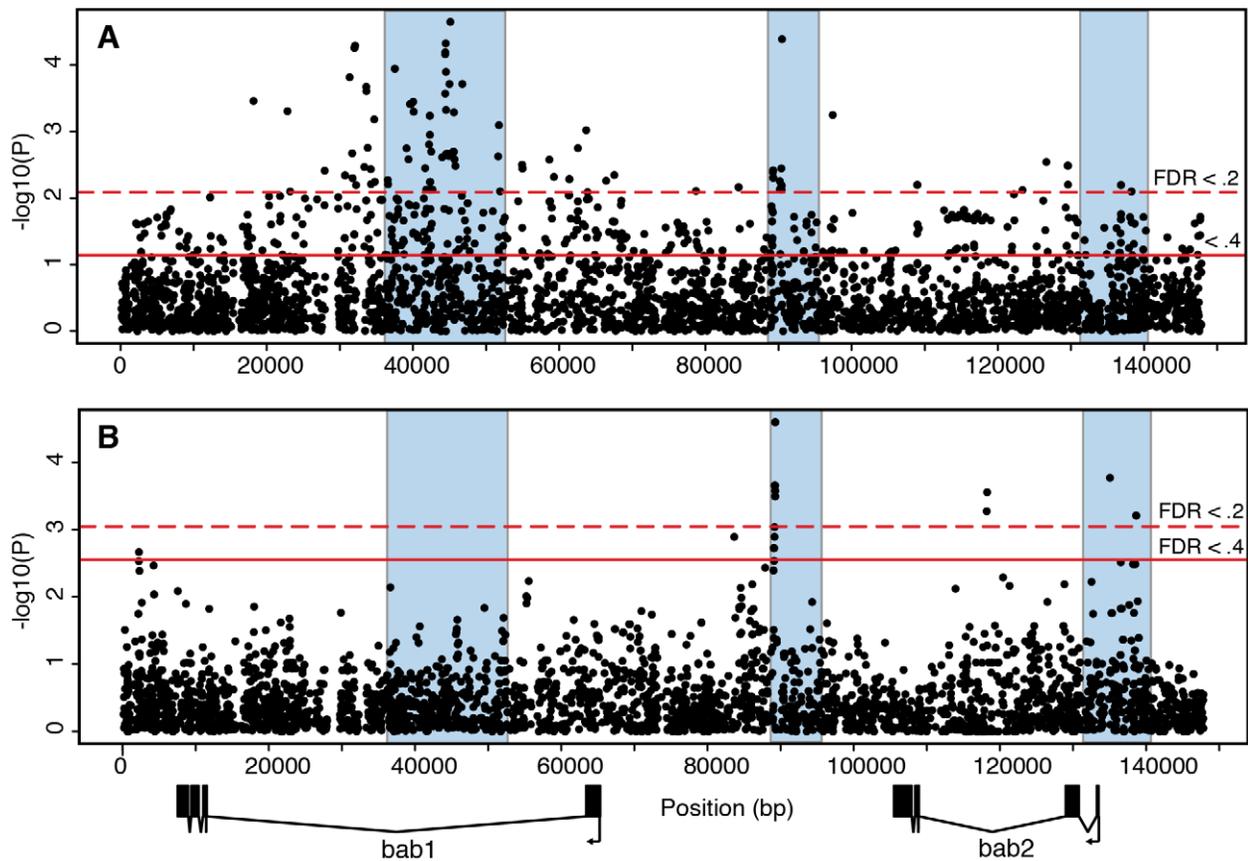


Figure 3. Association mapping of adult pigmentation and transcriptional variation in the subset of lines used to measure *bab* expression. (A) Association between polymorphisms and pigmentation. (B) Association between polymorphisms and *bab2* transcript levels. The p-values are from a Spearman rank correlation test for all polymorphisms. Regions shaded in blue are the same as in Figure 2. doi:10.1371/journal.pgen.1001275.g003

Causative polymorphisms have small individual effect sizes

The effect size of each of these associated polymorphisms is small, but within each region, these polymorphisms form haplotypes that have a large effect on pigmentation. To examine these effects, we assigned each individual a haplotype score, which indicates the relative number of dark and light alleles they contain at the significant polymorphisms (see Material and Methods for details). A negative haplotype score indicates more light alleles than dark alleles, a positive score more dark than light and a score of zero indicates an equal number of dark and light alleles. There is a strong correlation between the haplotype score and pigmentation for all three regions (Figure 4) ($p < .001$).

Within both regions 1 and 2 each polymorphism increases pigmentation by about 1% percent of the total variation (see Table 1). Some of these associated polymorphisms are likely not causative due to LD and false positives, but the effect size of each polymorphism is still small.

Despite each polymorphism only having a small effect on phenotype, together the polymorphisms result in haplotypes with large potential effects. Each region accounts for between 20% and 50% of the pigmentation variation. When we perform this analysis using haplotypes from all 3 regions, a similar pattern emerges; there is a small effect of each individual polymorphism, but a large overall effect (Table 1) (Figure 4). Together these three regions explain 84% of the pigmentation variation among the studied genotypes. The sum of regional effects appears to be greater than

the total effect of polymorphisms in all three regions because our sample had a number of lines with completely pigmented A6 segments. Since pigmentation score cannot be higher than 100, such lines are saturated with dark alleles. In those haplotypes, each individual dark allele appears to have a smaller effect size.

Polymorphisms act in *cis* to influence transcript abundance

The location of associated polymorphisms in three non-coding regions suggests that QTL effects are due to differences in transcriptional regulation rather than protein function. Variation in gene expression has become a recent frontier in molecular and evolutionary genetics [19]. Local (*cis*-acting) regulatory variation may consist of expression quantitative trait nucleotides (eQTNs) in sequences responsible for controlling transcription. In yeast, for instance, transcriptional variation is frequently associated with eQTNs located close to promoter regions [20]. Other possible mechanisms include variation in splicing efficiency [21] and conformation [22]. A natural allele can incorporate multiple eQTNs with a complex interplay of these mechanisms [23], and *cis*-acting variants can also interact with distant (*trans*) regulatory factors in complex ways [24].

We examined *cis*-acting transcriptional variation in *bab1* and *bab2* directly using allele specific expression (ASE) [25,26]. We selected 51 of our lines that were all collected in a single year. Analysis of association between sequence variation and pigmentation in this sub-panel reveals the concentration of significant

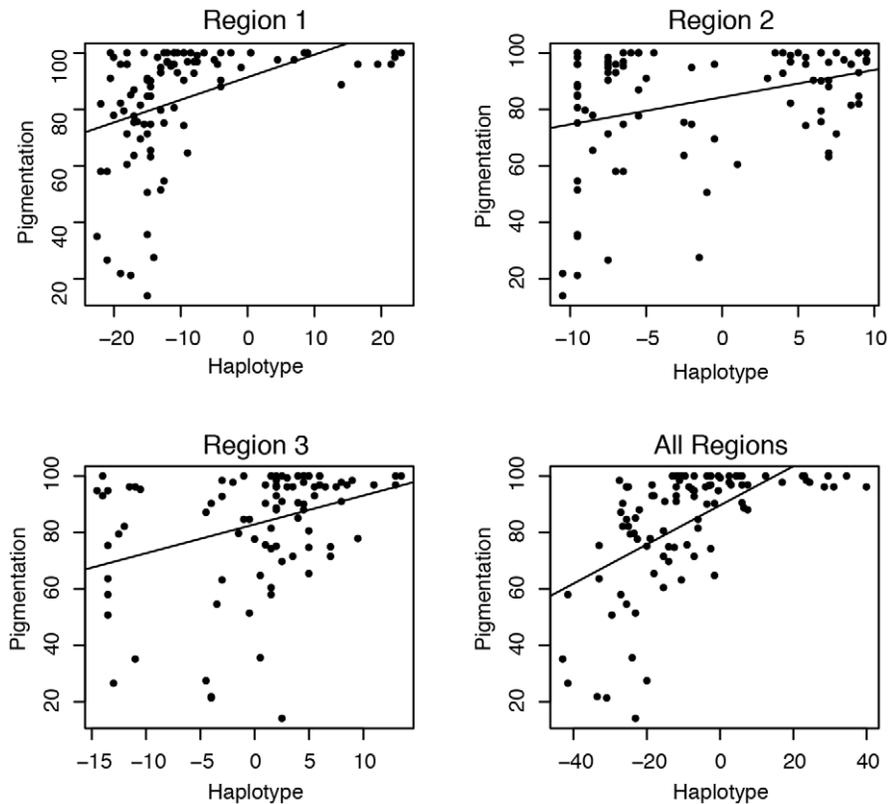


Figure 4. Pigmentation and the haplotype of associated polymorphisms. The haplotype represents the relative number of dark and light alleles at the associated polymorphisms within the region. Individuals with negative scores have more light alleles than dark alleles and positive scores have more dark alleles than light alleles (see Materials and Methods for details). The line represents the best fit of the pigmentation as a function of haplotype score. Each of the three regions is displayed individually and all three together.
doi:10.1371/journal.pgen.1001275.g004

polymorphisms in the same three non-coding regions (Figure 3A). We crossed each of these lines to two tester strains that contained low frequency SNPs in the coding regions of *bab1* and *bab2*, allowing us to differentiate the two alleles. The use of two different tester strains allowed us to identify consistent *cis* effects and minimize complex interactions between *cis* and *trans* factors. Heterozygous female abdomens were collected at 48–56 hours after pupariation, when the *bab* genes are expressed in the abdominal epidermis [6]. In these heterozygotes, we measured the relative transcript levels of the *bab1* and *bab2* alleles from the line of interest and the tester strain, and corrected for differences in the tester strains (see Materials and Methods for details). In this experimental design, any differences in expression between the

two parental alleles are due to *cis*-acting polymorphisms, and all 51 lines can be compared directly since they are measured relative to the same testers.

We used these transcription measurements to identify polymorphisms affecting transcription at the *bab* locus. We found a number of SNPs that are associated with variation in *bab2* transcript abundance (Figure 3B). Of the three previously identified regions affecting pigmentation, the regions flanking the PRE and the *bab2* transcription start site contain polymorphisms with the strongest effects on *bab2* transcript levels ($p < .001$; $FDR < .2$). This overlap suggests that at least some of the variation in sex-specific pigmentation is the result of differential regulation of *bab2* in different QTL alleles. We did not detect any effect of the region

Table 1. Estimation of polymorphism and region effect size.

	Number of sites	Average effect size of each polymorphism	Proportion of total variation explained by average polymorphism	Total estimated effect of region	Proportion of total variation explained
Region 1	55	0.80 (.19)	0.009	44	0.512
Region 2	21	0.97 (.28)	0.011	20.37	0.237
Region 3	29	1.02 (.30)	0.012	29.58	0.344
All 3 regions	105	0.69 (.10)	0.008	72.45	0.842

doi:10.1371/journal.pgen.1001275.t001

straddling the sexually dimorphic abdominal CRE (region 1 in Figure 2A) on transcript abundance. This may be due to the fact that, in contrast to the PRE and the promoter, this CRE is active only in the epidermis of the last three abdominal segments, whereas the RNA samples used in the analysis of transcriptional variation comprised multiple other tissues where the *bab* genes are expressed, including the anterior abdominal segments, the ovaries, the oenocytes, muscles, and the genitalia. Alternatively, sequence variation in this CRE may influence transcription during a time period that was not sampled in our experiments.

Although both *bab1* and *bab2* are expressed in the abdominal epidermis and contribute to the patterning of pigmentation [6,8], no sequence variants at the *bab* locus affect *bab1* transcript levels (FDR > 5) (data not shown). Chromatin immunoprecipitation data show extensive binding of RNA polymerase II (polIII) to the region surrounding the transcription start site of *bab2*, but not *bab1*, at the pupal stage [27]. In contrast, both promoter regions show similar polIII binding during embryonic stages. Thus, it appears that sequence variation at the *bab* locus affects adult pigmentation primarily by modulating the transcription of *bab2*. It is possible, of course, that *bab1* also contributes to phenotypic variation but at a different developmental stage that falls outside of the sampled period. Since pigmentation pattern is determined in a specific cell type at precise times during development [5,6,28], a more detailed understanding of how sequence variation affects transcription and translates into phenotypic differences will require quantitative analysis of gene expression with greater spatial and temporal resolution.

Previous research has identified two direct transcriptional regulators of *bab*, *Abd-B* and *dsx* [8]. Although the CREs responsible for *bab* expression in the abdominal epidermis are well defined [8], it is conceivable that additional regions modulate *bab* transcription in a redundant manner [29,30]. A search of the entire *bab* locus reveals 1412 potential Abd-B core binding sites (TTTAY) and 73 potential Dsx binding sites (ACWAWGT), which is higher than expected based on sequence composition (485 and 47 sites, respectively). However, none of the associated polymorphisms were found in potential Dsx binding sites, and only a few in potential Abd-B binding sites, in any of our analyses. This observation confirms that most variation in transcript levels and adult phenotype is due to sequence variation outside of known transcription factor binding sites.

Understanding the causative variation at the *bab* locus

The *bab* locus was identified as a large effect QTL responsible for most of the natural variation in female abdominal pigmentation [9]. By investigating the function of *bab* at two different levels, we were able to elucidate the fine structure of QTL alleles and characterize DNA sequence variation that affects transcription and phenotypic differences. Three major lessons emerge from this analysis. First, large-effect QTL alleles can reflect the cumulative effects of many nucleotide polymorphisms. This finding is in contrast to several cases where large QTLs are due to single coding or regulatory mutations with major phenotypic effects [31–33]. On the other hand, detailed analysis of a single gene responsible for 100% of a phenotypic difference between two *Drosophila* species revealed contributions from several independent regulatory regions [34]. Within *Drosophila melanogaster*, multiple polymorphisms in a single *cis* regulatory element of the *ebony* gene contribute to pigmentation differences between populations [35]. In addition many studies have shown the contribution of multiple polymorphisms to variation in protein function (e.g. [36–39]). These results suggest that accumulation of multiple mutations at

the same locus may be a relatively common mechanism of micro- and macroevolutionary change.

Second, the causative polymorphisms are not distributed randomly throughout the locus, but are located in multiple, distinct functional regions. In the case of *bab*, the vast majority of polymorphisms associated with transcriptional and phenotypic variation are concentrated around three regions: the core promoter, tissue-specific enhancer, and Polycomb response element – the elements expected to contribute the most to transcriptional regulation. Thus, functional annotation of the human and model system genomes, such as that generated by the ENCODE and modENCODE projects [27,40], may go some way towards predicting the loci of evolutionary variation. Conversely, quantitative-genetic analyses of natural variation may help identify and annotate functional genomic regions.

The third major lesson is that the functional elements present larger targets than previously thought for phenotypically and evolutionarily important mutations. The regions detected in our analysis are much larger than the core elements defined previously by functional assays. Over 90% of significant polymorphisms are found around, but not precisely inside, the core CRE, PRE, and promoter. This pattern indicates that DNA sequence variation in regions surrounding the core elements has major phenotypic consequences. These sequence variants may act by modulating the regulatory activity of the functional elements, perhaps through subtle changes in chromatin structure or transcription factor recruitment, without disrupting their primary functions. As a result, important evolutionary changes may occur through mutations in DNA sequences that lack readily identifiable molecular functions. The rapidly growing body of genome-wide association studies will put this proposition to the test. However, the final lesson from the *Drosophila bab* locus is that fine-scale dissection of functional variation is only possible in genes with little linkage disequilibrium, underscoring the importance of model systems in framing the analysis of human genetic variation.

Materials and Methods

Lines selection and sequencing

96 inbred lines of *D. melanogaster* were chosen for sequencing the *bab* locus. 86 lines were chosen at random from the W1 and W3 inbred line collections maintained in the Nuzhdin lab. Progenitors of these lines were originally collected from Wolfskill Orchard in Winters, CA, and inbred by full sib mating for a minimum of 20 generations. 10 additional lines with low pigmentation scores were selected, including 8 lines from Winters and 2 from North Carolina. Sequencing was performed as described [10].

Pigmentation phenotyping and association mapping

3–5 day old females were scored for pigmentation as described [9]. All statistical tests were performed using the R Statistical package (<http://www.r-project.org/>). Spearman's correlation test and associated p-values were calculated using “cor.test” command in the base Stats package for each polymorphism individually. False Discovery rates were calculated using the q-value package.

Identification of enriched regions and estimating effects size

To identify regions with an excess of associated points, we scanned the *bab* locus in 5 kb sliding windows, shifting the window in 500 bp increments. Regions where more than 0.12 of polymorphisms were significant were tentatively identified as enriched regions. We manually excluded the region that contained the cryptic population structure and then manually manipulated

each remaining region to only include the significant points (i.e. we narrowed the region if the high score was the result of all of the polymorphisms being clustered on one side of the window).

To understand the effect size for each region, we created a region haplotype score for each individual. This haplotype score is the difference between the number of dark alleles and light alleles for significant polymorphisms divided by 2. ((Number of dark alleles – number of light alleles)/2). Thus an increase in the haplotype score by one indicates a change of a single light allele to a dark allele. A score of zero would indicate an equal number of dark and light alleles for that region. This formula was necessary because not all individuals had a genotype assigned for all polymorphism.

We then fit a linear model with the formula: *pigmentation score* = *haplotype*. The coefficient of the *haplotype* term is the average effect of each allele on the pigmentation score. To calculate the average effect size of the entire region we multiplied the average effect of each allele by the number of associated polymorphisms in that region.

Quantification of *bab* transcripts

We measured allele-specific expression of both *bab1* and *bab2* in 51 lines out of the 96 used for pigmentation mapping. Two lines that carried rare SNPs in the coding regions of *bab1* and *bab2* were chosen as the tester strains. Each of these tester strains was crossed to the 51 experimental lines in several replicates, creating F1 genotypes that were heterozygous for the tester and experimental alleles.

We collected 48–56 hour whole pupal abdomens from the female F1s. Individuals from replicate crosses were combined into

pools of 10, and we attempted to obtain at least 2 pools from each cross. From each of these samples, a 218 bp fragment of *bab1* and a 142 bp fragment of *bab2*, spanning the SNPs of interest, were amplified. We used Illumina high throughput sequencing of barcoded samples to measure the relative amounts of each allele as described [26]. To normalize the samples from the two tester lines, we took the residuals from a fixed cofactor ANOVA. We then mapped the expression variation across the *bab* region using the same procedures as for the adult phenotype.

Binding site predictions were made using custom R and Perl scripts for the sequences ACWAWGT (*Dsx*) and TTTAY (*Abd-B*).

Supporting Information

Figure S1 Pairwise r^2 LD values for significantly associated points in the 3 designated regions. Each square represents the pairwise r^2 value. Darker squares indicate higher r^2 values. Found at: doi:10.1371/journal.pgen.1001275.s001 (0.64 MB TIF)

Acknowledgments

We thank the Joint Genome Institute community sequencing program, Wendy S. Schackwitz, and Len A. Pennacchio for help with the sequencing. We thank Jennifer Brisson and three anonymous reviewers for helpful comments on the manuscript.

Author Contributions

Conceived and designed the experiments: RDB AK SVN. Performed the experiments: RDB. Analyzed the data: RDB AK SVN. Wrote the paper: RDB AK SVN.

References

- Mackay TF (2001) Quantitative trait loci in *Drosophila*. *Nat Rev Genet* 2: 11–20.
- Reich DE, Lander ES (2001) On the allelic spectrum of human disease. *Trends Genet* 17: 502–510.
- Genissel A, Pastinen T, Dowell A, Mackay TF, 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.
- De Luca M, Roshina NV, Geiger-Thornsberry GL, Lyman RF, Pasyukova EG, et al. (2003) Dopa decarboxylase (*Ddc*) affects variation in *Drosophila* longevity. *Nat Genet* 34: 429–433.
- Couderc JL, Godt D, Zollman S, Chen J, Li M, et al. (2002) The bric a brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in *Drosophila*. *Development* 129: 2419–2433.
- Kopp A, Duncan I, Carroll SB (2000) Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* 408: 553–559.
- Gompel N, Carroll SB (2003) Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. *Nature* 424: 931–935.
- Williams TM, Selegue JE, Werner T, Gompel N, Kopp A, et al. (2008) The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* 134: 610–623.
- Kopp A, Graze RM, Xu SZ, Carroll SB, Nuzhdin SV (2003) Quantitative trait loci responsible for variation in sexually dimorphic traits in *Drosophila melanogaster*. *Genetics* 163: 771–787.
- Bickel RD, Schackwitz WS, Pennacchio LA, Nuzhdin SV, Kopp A (2009) Contrasting patterns of sequence evolution at the functionally redundant bric a brac paralogs in *Drosophila melanogaster*. *J Mol Evol* 69: 194–202.
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100: 9440–9445.
- Ringrose L, Rehmsmeier M, Dura JM, Paro R (2003) Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila melanogaster*. *Dev Cell* 5: 759–771.
- Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, et al. (2006) Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat Genet* 38: 700–705.
- Kwong C, Adryan B, Bell I, Meadows L, Russell S, et al. (2008) Stability and dynamics of polycomb target sites in *Drosophila* development. *PLoS Genet* 4: e1000178. doi:10.1371/journal.pgen.1000178.
- Goodrich JA, Tjian R (2010) Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet* 11: 549–558.
- Wu JQ, Snyder M (2008) RNA polymerase II stalling: loading at the start prepares genes for a sprint. *Genome Biol* 9: 220.
- Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, et al. (2007) RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* 39: 1512–1516.
- Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, et al. (2007) RNA polymerase is poised for activation across the genome. *Nat Genet* 39: 1507–1511.
- Skelly DA, Ronald J, Akey JM (2009) Inherited variation in gene expression. *Annu Rev Genomics Hum Genet* 10: 313–332.
- Ronald J, Brem RB, Whittle J, Kruglyak L (2005) Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS Genet* 1: e25. doi:10.1371/journal.pgen.0010025.
- Telonis-Scott M, Kopp A, Wayne ML, Nuzhdin SV, McIntyre LM (2009) Sex-specific splicing in *Drosophila*: widespread occurrence, tissue specificity and evolutionary conservation. *Genetics* 181: 421–434.
- Duan S, Huang RS, Zhang W, Bleibel WK, Roe CA, et al. (2008) Genetic architecture of transcript-level variation in humans. *Am J Hum Genet* 82: 1101–1113.
- Stam LF, Laurie CC (1996) Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* 144: 1559–1564.
- Wang HY, Fu Y, McPeck MS, Lu X, Nuzhdin S, et al. (2008) Complex genetic interactions underlying expression differences between *Drosophila* races: analysis of chromosome substitutions. *Proc Natl Acad Sci U S A* 105: 6362–6367.
- Wittkopp PJ, Haerum BK, Clark AG (2004) Evolutionary changes in cis and trans gene regulation. *Nature* 430: 85–88.
- Main BJ, Bickel RD, McIntyre LM, Graze RM, Calabrese PP, et al. (2009) Allele-specific expression assays using Solexa. *BMC Genomics* 10: 422.
- Celniker SE, Dillon LA, Gerstein MB, Gunsalus KC, Henikoff S, et al. (2009) Unlocking the secrets of the genome. *Nature* 459: 927–930.
- Wittkopp PJ, Carroll SB, Kopp A (2003) Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends Genet* 19: 495–504.
- Frankel N, Davis GK, Vargas D, Wang S, Payre F, et al. (2010) Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*.
- Hong JW, Hendrix DA, Levine MS (2008) Shadow enhancers as a source of evolutionary novelty. *Science* 321: 1314.
- Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP (2006) A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313: 101–104.

32. Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, et al. (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet* 38: 107–111.
33. Chan YF, Marks ME, Jones FC, Villarreal G, Jr., Shapiro MD, et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* 327: 302–305.
34. McGregor AP, Orgogozo V, Delon I, Zanet J, Srinivasan DG, et al. (2007) Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* 448: 587–590.
35. Rebeiz M, Pool JE, Kassner VA, Aquadro CF, Carroll SB (2009) Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* 326: 1663–1667.
36. Babbitt CC, Silverman JS, Haygood R, Reininga JM, Rockman MV, et al. (2010) Multiple Functional Variants in cis Modulate PDYN Expression. *Mol Biol Evol* 27: 465–479.
37. Horan M, Millar DS, Hedderich J, Lewis G, Newsway V, et al. (2003) Human growth hormone 1 (GH1) gene expression: complex haplotype-dependent influence of polymorphic variation in the proximal promoter and locus control region. *Hum Mutat* 21: 408–423.
38. Sun C, Southard C, Witonsky DB, Olopade OI, Di Rienzo A (2010) Allelic imbalance (AI) identifies novel tissue-specific cis-regulatory variation for human UGT2B15. *Hum Mutat* 31: 99–107.
39. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* 39: 31–40.
40. Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447: 799–816.