Linked genetic variation and not genome structure causes widespread differential expression associated with chromosomal inversions

Iskander Said[a,b], Ashley Byrne[c], Victoria Serrano[b], Charis Cardeno[d], Christopher Vollmers[a,b], and Russell Corbett-Detig[a,b,1]

[a]Genomics Institute, University of California, Santa Cruz, CA 95064; [b]Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064; [c]Department of Molecular, Cellular and Developmental Biology, University of California, Santa Cruz, CA 95064; and [d]Department of Evolution and Ecology, University of California, Davis, CA 95616

Chromosomal inversions are widely thought to be favored by natural selection because they suppress recombination between alleles that have higher fitness on the same genetic background or in similar environments. Nonetheless, few selected alleles have been characterized at the molecular level. Gene expression profiling provides a powerful way to identify functionally important variation associated with inversions and suggests candidate phenotypes. However, altered genome structure itself might also impact gene expression by influencing expression profiles of the genes proximal to inversion breakpoint regions or by modifying expression patterns genome-wide due to rearranging large regulatory domains. In natural inversions, genetic differentiation and genome structure are inextricably linked. Here, we characterize differential expression patterns associated with two chromosomal inversions found in natural *Drosophila melanogaster* populations. To isolate the impacts of genome structure, we engineered synthetic chromosomal inversions on controlled genetic backgrounds with breakpoints that closely match each natural inversion. We find that synthetic inversions have negligible effects on gene expression. Nonetheless, natural inversions have broad-reaching regulatory impacts in *cis* and *trans*. Furthermore, we find that differentially expressed genes associated with both natural inversions are enriched for loci associated with immune response to bacterial pathogens. Our results support the idea that inversions in *D. melanogaster* experience natural selection to maintain associations between functionally related alleles to produce complex phenotypic outcomes.

chromosomal inversions | differential expression | genome structure

Chromosomal inversions—reversed regions in the linear map order of a chromosome—are central to a myriad of evolutionary processes and are ubiquitous in natural populations (1, 2). Because chromosomal inversions strongly suppress recombination in heterozygotes, they can maintain associations between mutations that would otherwise be broken down by recombination. If the mutations captured by a chromosomal inversion are more fit in similar genetic or environmental contexts, natural selection may favor the novel arrangement. Indeed, this observation is the basis of numerous theoretical models that aim to explain the initial rise in frequency and maintenance of polymorphic inversions in natural populations. Inversions are thought to be involved in sex-chromosome evolution (3), local adaptation (4, 5), meiotic drive (6), and complex behavioral traits (7). Although the importance of chromosomal inversions in evolution is increasingly widely appreciated (1, 2), we are only beginning to understand the sources of natural selection that affect inversions at the molecular and phenotypic levels.

Gene expression profiling offers an appealing avenue for investigating the molecular impacts of chromosomal inversions (8) and for identifying candidate phenotypes that are favored by natural selection (9). Previous works in *Drosophila* found expression differences associated with natural inversions on the same chromosome

(9–11) as well as for loci distributed genome-wide (10, 11). These results have been interpreted to be a consequence of linked allelic variation and suggest a role of natural selection in maintaining linkage among functional variants. However, engineered chromosomal inversions in yeast and *Drosophila* can affect expression at hundreds to thousands of genes genome-wide (8, 12), indicating that genome structure itself can profoundly impact gene expression even in the absence of linked allelic differentiation. For this reason, the relative impacts of genome structure and linked allelic variation remain largely unknown, and distinguishing among them is central to advancing our understanding of inversion biology.

There are at least three nonmutually exclusive ways in which inversions could influence gene expression. Specifically, (*i*) inversions can affect gene expression patterns genome-wide through reorganizing large regulatory domains (8, 12). (*ii*) Inversions also sometimes impact gene expression locally through the modification of the genetic regions or epigenetic environment adjacent to their breakpoints (11, 13). (*iii*) Finally, inversions can maintain linkage with *cis* or *trans* acting regulatory elements located within or near to the inverted region due to

---

Significance

Chromosomal inversions are among the primary drivers of genome structure evolution and are thought to be favored by natural selection because they suppress recombination between co-adapted genes. However, dramatically reorganizing the genome could also have its own functional implications. In natural inversions, genome structure and allelic content are inextricably linked, and quantifying their relative contributions is crucial for understanding genome structure evolution. Here, we use genome engineering tools to construct synthetic inversions whose structures mimic natural inversions. We find that synthetic inversions do not influence gene expression, while natural inversions influence expression genome-wide. Our results indicate that genetic variation associated with inversions has widespread *cis* and *trans* regulatory effects and support evolutionary models wherein natural selection maintains co-adapted variation.

---

suppressed recombination in heterozygotes (9, 11). Comprehensively evaluating the gene expression changes of naturally occurring inversions and distinguishing between these hypotheses is therefore a fundamental step toward understanding the sources of natural selection that influence the distribution of chromosomal inversions in natural populations.

Here, we analyze differential gene expression associated with two high-frequency inversions of *Drosophila melanogaster*. Both inversions show evidence of natural selection and are minimally genetically differentiated from the standard arrangement in the species' ancestral range (14–17). These samples are therefore ideally suited to evaluating the effects of natural selection and its resulting consequences at the level of gene expression. To independently investigate the effect of genome structure, we use the flippase recognition target/flippase (FRT/FLP) recombination system (18, 19) to engineer synthetic inversions whose breakpoints closely mimic the positions of the natural inversions. Because each FRT element was inserted on a completely homozygous genetic background, comparisons with natural inversions enable us to distinguish between the impacts of genome structure from those of linked allelic variation with which natural inversions are associated.

We find that whereas synthetic inversions influence transcript abundance at a handful of loci, natural inversions have much broader impacts on gene expression both locally and genome-wide, suggesting both *cis* and *trans* effects are contributed by alleles associated with natural inversions. Strikingly, differentially expressed genes for both natural inversions are enriched for defense responses to pathogens, suggesting that these structural variants experience selection to maintain complexes of functionally related alleles with similar phenotypic effects.

Results and Discussion

Experimental Crosses, Sequencing, and Primary Analysis. We first identified isofemale lines that were homozygous for each natural chromosomal inversion or for the standard arrangement using PCR (20). All lines were from a single collection from a putatively ancestral population sampled in Siavonga, Zambia (21). We produced crosses among homozygotes of each arrangement, as well between homozygotes of alternative arrangements ($n = 10$–15; Table S1). In this way, we sought to ensure that all samples for natural inversions were outbred and their genetic backgrounds largely randomized with the exception of the inversions. From these crosses, we retained sets of 20 males, aged to 4 d old, for total RNA extraction and sequencing.

Concurrently, we produced and confirmed synthetic inversions to closely mimic natural inversions using the DrosDel collection (22), which is based on the FRT/FLP recombinase system (18, 19) ($n = 2$–3; Tables S1 and S2). Because this collection was created using P-element vectors, which insert randomly in the genome, the breakpoints of synthetic inversions are not identical to the base pair with those of natural inversions. We therefore designed synthetic breakpoints to be as similar as possible. The mean distance between natural and synthetic breakpoints is 92 Kb, and three of four synthetic breakpoints are in the same topologically associated domain as the natural inversion breakpoints (determined using the map of ref. 23 and Table S2). As with the natural isolates, we studied homozygotes and heterozygotes of these arrangements on this controlled genetic background.

After sequencing, we obtained an average of 8.8 million read pairs per library (Table S1). All differential expression (DE) analyses simultaneously compared the three possible genotypes—inversion homozygote, heterozygote, and standard homozygote—for each natural and synthetic inversion using a generalized linear modeling framework to detect differences in expression levels among arrangements (24). Because we assayed the three possible genotypic combinations, our crossing scheme enabled us to classify expression patterns across differentially expressed loci as a consequence of inversion dosage. Specifically, we defined

three expression classes: (*i*) additive, for loci in which expression levels in heterozygous individuals were intermediate to either homozygote; (*ii*) overdominant, for loci in which the heterozygote's expression exceeded either homozygote; and (*iii*) underdominant, for loci where the heterozygote's expression was less than either homozygote. As in ref. 9, we note that expression categories are not necessarily related to the fitness of these arrangements but describe only relative patterns of gene expression across arrangements.

Natural Inversions Affect Expression at Substantially More Loci than Synthetic Inversions. We identified a moderate number of genes in which expression was affected relative to the standard arrangement either as a heterozygote or as a homozygote in natural inversions [117 and 87 for *In(2L)t* and *In(3R)K*, respectively; Fig. 1 and Dataset S1]. This result contrasts sharply with those from our two synthetic inversions that we designed to have very similar breakpoints to the natural inversions. For the two synthetic inversions, we found that the effects on gene expression is modest [10 and 1 genes for the synthetic inversions *In(2L)st* and *In(3R)sK*, respectively; Dataset S1]. To confirm that these differences do not reflect differences in our power to detect DE in natural and synthetic inversions, we subsampled the natural inversion data to identical sample sizes and repeated our analyses. In the sub-sampled data, we identified 90 and 76 differentially expressed loci between arrangements for *In(2L)t* and *In(3R)K*. This implies that these large-scale differences in expression patterns between natural and synthetic inversions are not consequences of reduced statistical power in the smaller sample sizes of synthetic inversions. More generally, since genome structure appears to have little effect on gene expression patterns genome-wide, much of the observed DE in natural inversions is likely due to differences in their linked allelic content.

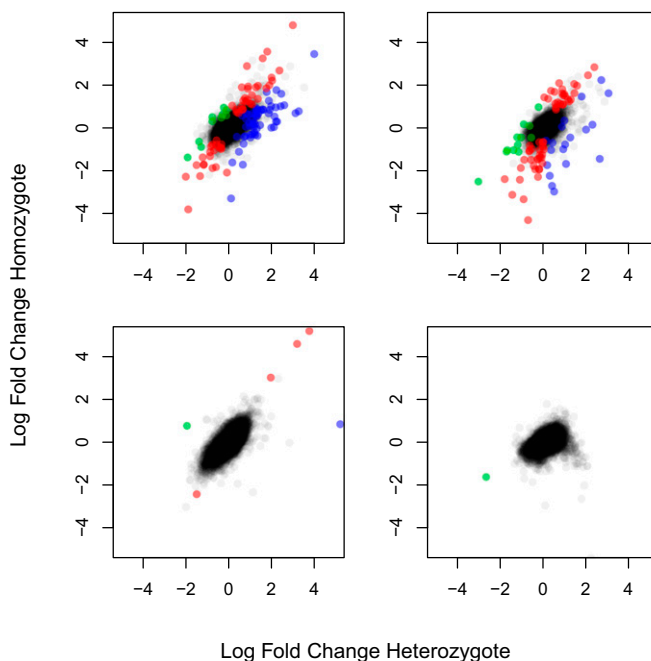**Fig. 1.** Relative expression impacts of chromosomal inversions. Log fold-change in inversion heterozygotes and homozygotes relative to standard arrangement homozygotes for *In(2L)t* (*Left*) and *In(3R)K* (*Right*) with natural inversions (*Top*) and synthetic inversions (*Bottom*). Expression differences that are significant at the q < 0.2 level are shown as red (additive), blue (overdominant), and green (underdominant).

EVOLUTION

No Evidence of Breakpoint Effects on Expression Patterns. In addition to genome-wide expression changes, it is possible that inversions influence transcription at loci proximal to their breakpoints (9, 11, 25). This might occur either because inversions directly interrupt genic sequences or because inversions alter proximities to local regulatory elements. However, for none of the natural or synthetic inversions do any loci adjacent to a breakpoint show significant expression differences. Furthermore, we find only two differentially expressed genes within 100 Kb of any inversion breakpoint, which is inconsistent with an impact of inversion breakpoints on patterns of DE at proximal loci [$P = 1$ and $P = 0.4362$, Fisher's exact test for $In(2L)t$ and $In(3R)K$, respectively]. We also obtained similar nonsignificant results for enrichment of DE at 20 Kb and 50 Kb from inversion breakpoints ($P > 0.1$ for each inversion and distance). In combination with the weak impacts of the synthetic chromosomal inversions on gene expression genome-wide, our results from breakpoint proximal genes suggest that the changes in genome structure associated with $In(2L)t$ and $In(3R)K$ are not the primary drivers of the expression changes associated with the natural inversions. Nonetheless, synthetic inversions have been shown to exhibit variable effects across the *Drosophila* genome, sometimes influencing expression at similar numbers of loci as we found for natural arrangements (12), and inversions have even more widespread impacts in yeast (8). Therefore, when feasible, genome structure comparisons should be used for investigations of gene expression as well as for studies of other molecular phenotypes of naturally occurring chromosomal inversions.

Enrichment of Chromosome-Specific Additive Gene Expression Effects. Because natural inversions suppress recombination in heterozygotes, they often maintain strong associations with linked alleles (15). Consistent with this, we find an excess of differentially expressed loci on the same chromosome arm as each inversion [$P = 2.77e-5$ and $P = 2.45e-3$, Fisher's exact test for $In(2L)t$ and $In(3R)K$, respectively; Fig. 2]. Furthermore, among DE loci, those located on the same chromosome arm as each inversion are enriched for additive expression patterns, where the heterozygote is intermediate relative to the two homozygotes [$P = 8.16e-4$ and $P = 5.99e-3$, Fisher's exact test for $In(2L)t$ and $In(3R)K$, respectively; Fig. 2]. Because *cis* acting regulatory elements affect expression of only one gene copy, we expect *cis* mutations to contribute primarily additive expression differences, consistent with previous results in this species (26). Therefore, a reasonable interpretation of the enrichment of additive loci on the same chromosome arm is that each natural inversion is associated with a group of alleles that have *cis* acting effects on gene expression. Many of these DE genes are distant from inversion breakpoints in genomic regions where strong genetic associations between inversions and linked variation break down (15). Hence, the excess of linked regulatory variation associated with chromosomal inversions is consistent with a role of natural selection in maintaining linkage among functional variants.

***Trans* Regulatory Effects Are Widespread.** Given the large impacts on gene expression genome-wide (Fig. 2), our results also suggest that loci that are linked to natural inversions have significant *trans* acting regulatory effects. One important consideration is that if alleles on other chromosome arms were strongly associated with inversions, these genome-wide effects might also be attributed to linked *cis* acting elements (10, 11). However, when we searched the genomes of each arrangement for alleles that are strongly associated with inversions but that are not physically linked, we found few. Indeed, F_{ST} between arrangements, a measure of genetic differentiation, is reduced on chromosome arms that do not contain chromosomal inversions (Fig. 3). Whereas there is a significant association between mean F_{ST} and DE on inverted chromosome arms [$P = 0.0007$ and $P = 0.0027$,
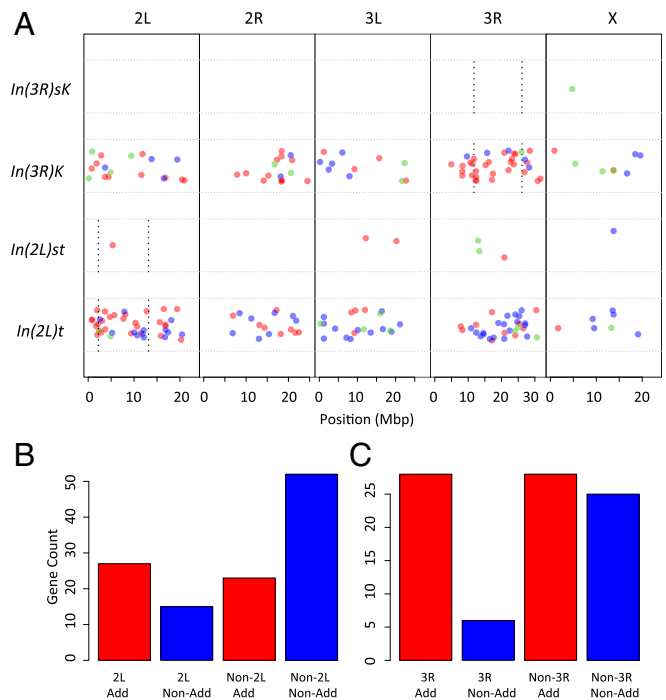

Fig. 2. The genomic distributions of DE genes. (*A*) Differentially expressed genes with additive (red), overdominant (blue), and underdominant (green) expression patterns for each natural and synthetic inversion across the five major chromosome arms of *D. melanogaster*. In order from *Left* to *Right*, chromosome arms displayed in each panel are 2L, 2R, 3L, 3R, and ×. From *Top* to *Bottom*, inversions shown are $In(3R)sK$, $In(3R)K$, $In(2L)st$, and $In(2L)t$. Positions of inversion breakpoints are shown for each as dashed vertical lines. Each point was jittered vertically to improve visualization. (*B*) The number of additive (red) and nonadditive (blue) genes on chromosome arm 2L and off for $In(2L)t$. (*C*) The number of additive (red) and nonadditive (blue) genes on chromosome arm 3R and in the rest of the genome for $In(3R)K$.

permutation test for $In(2L)t$ and $In(3R)K$, respectively], we find no association between mean F_{ST} and the presence of differentially expressed loci on collinear chromosome arms [$P = 0.94$ and $P = 0.41$ permutation test for $In(2L)t$ and $In(3R)K$, respectively]. A previous analysis of whole-genome resequencing data in samples from the species' ancestral range also found few alleles on unlinked chromosome arms that are in strong linkage with natural inversions (15), consistent with our results. These data therefore lend additional support to the idea that DE loci at unlinked genes is a consequence of *trans* regulatory effects of loci that are closely associated with chromosomal inversions.

We next sought to quantify the relative impact of *trans* regulation. To do this, we estimated the proportion of loci on collinear arms that are DE. Because there are few strongly linked sites on collinear arms (Fig. 3), DE genes found in these regions should be largely attributable to *trans* effects. Then, by multiplying this rate by the number of genes on the same chromosome arm, we can estimate the number of linked genes whose expression was impacted by *trans* effects. Finally, by subtracting the expected proportion of *trans*-impacted loci from the set of DE genes on the same chromosome arm, we obtain an estimate of the number of loci whose expression is modulated by *cis* impacts (i.e., following the approach of ref. 26). Using this simple approximation, we estimate that expression patterns of 79% and 81% of differentially expressed genes, for $In(2L)t$ and $In(3R)K$, respectively, result from *trans* regulatory effects. Therefore, although the evolution of chromosomal inversions is often interpreted with specific reference to their effects on linked sequence
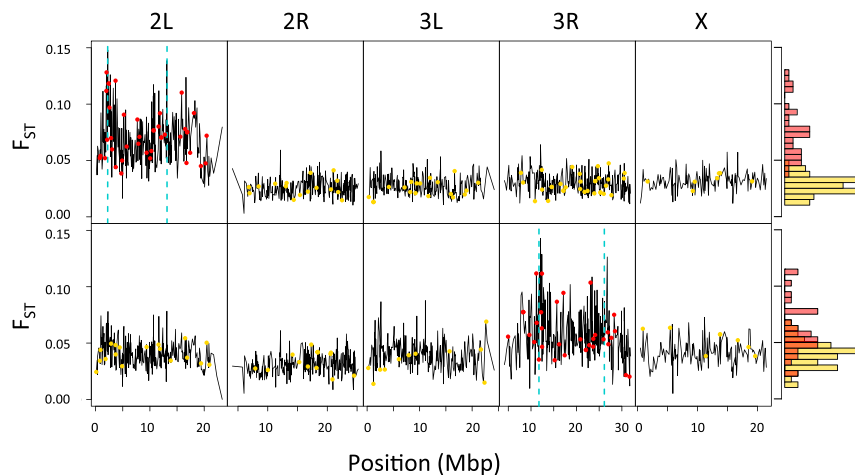
Fig. 3. Genetic differentiation is correlated with DE genes on inverted chromosome arms. F_{ST}, a measure of genetic differentiation, between standard arrangement homozygotes versus inverted arrangement homozygotes for *In(2L)t* (*Top*) and *In(3R)K* (*Bottom*). DE genes are marked along each panel as red (on the inverted chromosome) and gold (on collinear chromosomes); each panel corresponds to one of the major chromosome arms of this species: 2L, 2R, 3L, 3R, and X from *Left* to *Right*. Inversion breakpoint positions are marked with dashed vertical lines. On the right, histograms show the distributions of F_{ST} for windows containing DE genes on collinear chromosome arms (gold) and DE genes on inverted chromosome arms (red).

variation, the majority of the gene expression impacts extend much farther—potential affecting the entire genome.

There are at least two plausible mechanisms through which inversion-associated alleles could contribute *trans* regulatory effects. One obvious mechanism is through expression differences among transcription factors, which could then influence expression at other loci (11). Consistent with this hypothesis, we note that there are five and two transcription factors in the list of differentially expressed genes on the inverted chromosome arm for *In(2L)t* and *In(3R)K*, respectively (Dataset S1). Expression of these loci may therefore contribute to the genome-wide expression differences. However, because expression networks are sometimes regulated by protein and transcript abundances, it is also possible that the *trans* effects that we observe are a consequence of compensatory effects due to expression differences caused by *cis* regulatory alleles that perturb expression networks. More detailed analyses are necessary to conclusively identify the specific molecular causes underlying the broad expression differences we observe associated with the allelic content of natural inversion polymorphisms.

DE Patterns Suggest a Role in Immune Response. Strong biogeographic evidence supports a role for natural selection in influencing the distribution and abundance of these chromosomal inversions in natural populations (e.g., refs. 14–17, 27, and 28). It is therefore valuable to ask whether there are commonalities in the functions of genes that are differentially expressed between arrangements, which could provide insights into the functional impacts of chromosomal inversions in natural populations. For each natural inversion, we identified numerous gene ontology terms that are significantly enriched within the set of differentially expressed genes (Dataset S2). One striking feature is that differentially expressed loci for both natural inversions are enriched for defense responses to bacteria (Table 1), suggesting that both of these inversions may be favored due to similar phenotypic effects.

One possible reason that similar gene ontology terms were identified in the two natural inversion comparisons is that the inversions influence expression of the same genes. We found a significant overlap in the sets of differentially expressed loci that are DE in both natural inversions (19 genes, $P \le$ 1e-3, permutation test). However, the bacterial defense-related GO terms are attributed to enrichment in nonoverlapping sets of loci between the two inversion comparisons (Dataset S2). This indicates that the similarity of functional category enrichment is not driven entirely by overlap in DE genes and instead suggests that each inversion influences variation in similar phenotypes through independent genetic mechanisms.

Genetic variation in immune response can be maintained by frequency-dependent balancing selection (29, 30). Under such a model, we expect to find chromosomal inversions at intermediate frequencies across diverse populations. Despite their recent

Table 1. Enriched GO terms related to immune response to bacterial infections associated with DE genes in each natural inversion

Inversion	GO term	Genes	P value	Fold enrichment	Benjamini
In(2L)t	Defense response to Gram-positive bacterium	5	3.35E-04	14.57	0.0209
	Antimicrobial	4	6.41E-04	22.66	0.0253
	Bacteriolytic enzyme	3	0.0017	45.33	0.0339
	Defense response to Gram-negative bacterium	5	0.0084	6.13	0.1907
In(3R)K	Response to bacterium	8	3.66E-10	42.56	6.08E-08
	Innate immune response	8	5.36E-07	15.62	4.45E-05
	Defense response	6	8.77E-07	31.92	4.85E-05
	Immunity	6	2.03E-05	17.49	3.56E-04
	Innate immunity	6	1.62E-05	18.30	3.78E-04
	Antibacterial humoral response	3	0.0025	38.57	0.0986

evolutionary origins (15, 31), both of these inversions have been observed at intermediate frequencies across the species' ancestral range in Sub-Saharan Africa but are only rarely found as fixations within populations (14, 15, 32, 33). A role in impacting diverse immune responses is therefore consistent with the patterns of chromosomal inversion frequency variation in natural populations of *D. melanogaster*. However, we caution that this is far from a definitive list of the possible inversion-related phenotypes. Because we assayed expression only in whole adult males and because these inversions each span ~10% of the genome, it is likely that both *In(2L)t* and *In(3R)K* influence additional complex phenotypes. Nonetheless, a role in immune response could be consistent with the biogeographic distributions of these common inversions and offers evolutionarily important and testable hypotheses that could shed light on the functional impacts of chromosomal inversions in natural populations.

Conclusion

By contrasting differential gene expression patterns associated with two natural chromosomal inversions with synthetic inversions whose structures closely mirror them, we distinguished between the impacts of structural and linked allelic variation on gene expression patterns. Our results suggest that natural inversions influence gene expression as a consequence of linked allelic variation maintained within regions of suppressed recombination associated with natural inversions and are therefore consistent with previous findings in *Drosophila*. Both natural inversions that we studied impacted expression genome-wide, suggesting pronounced *trans* acting effects on gene expression. Furthermore, these differentially expressed genes are functionally related and lend further support to evolutionary models wherein inversions are favored by natural selection because they suppress recombination between alleles that are favored in similar contexts. Therefore, collectively, our data lend strong support to the idea that natural selection on chromosomal inversions operates to maintain combinations of alleles that act in concert to produce diverse phenotypic outcomes in natural populations.

Methods

Constructing Synthetic Inversions. For both natural inversions, we created synthetic inversions with similar breakpoints on a controlled genetic background [*In(2L)st* and *In(3R)sK*; Table S2]. Specifically, we selected pairs of FRT-bearing stocks from the DrosDel collection (22) for which the FRT elements were inserted in opposite orientation and as near to the natural inversions' breakpoints as was feasible (Table S2). Where possible, we additionally selected inversion breakpoints that were in the same topologically associated domains (TADs) (using the TAD map of ref. 23) as the natural inversion breakpoints. We induced inversions as described in ref. 18 and confirmed each rearrangement using PCR. To do this, we used NEB Long Amp Taq Polymerase for each PCR longer than 1,500 bp and NEB Taq Polymerase for all confirmation reactions shorter than this length. For both, we followed the manufacturer's reaction recommendations. DrosDel stocks and the PCR primers and conditions used in each confirmation reaction are listed in Table S2.

Experimental Crosses. We selected lines from the putatively ancestral population of *D. melanogaster* that we collected in Siavonga, Zambia (21, 33). We used the PCR primers of ref. 20 to identify isofemale lines that are homozygous for *In(2L)t* or *In(3R)K* or that are homozygous for the standard arrangement. We produced crosses between inversion homozygotes, inversion heterozygotes, and standard arrangement lines and between standard arrangement lines. This scheme created all possible genotypes, and by outcrossing lines, we mitigated the potentially important confounding effects of inbreeding within isofemale lines on natural gene expression phenotypes. Table S1 contains a complete list of lines and crosses used in this study.

RNA Extractions, Library Construction, and Sequencing. We collected males within 12 h of emergence and aged them for 4 d in male-only vials for each cross. Then, we flash-froze males in liquid nitrogen and stored them at −80°. We used TRIzol to extract whole RNA from groups of 20 males in each cross following the Invitrogen RNA extraction protocol (tools.thermofisher.

com/content/sfs/manuals/trizol_reagent.pdf). Finally, for each sample, we quantified RNA degradation by visual inspection of bioanalyzer traces produced using RNA nano bioanalyzer chips following the manufacturer's instructions. We produced sequencing libraries for each cross using the Smartseq2 protocol (34), which uses oligo-dT primers to reverse-transcribe poly-A–tailed mRNA transcripts in whole RNA extractions. We sequenced all libraries on five lanes of a HiSEq. 2500 using 50 bp paired-end reads. All sequencing was performed at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley.

In designing our plate layout for RNA extractions and library preparation, we randomized all samples across genotypic classes to mitigate the possibility that we would recover expression differences as an indirect consequence of block effects. Furthermore, all libraries in a given DE analyses were prepared in the same plate on the same day and pooled only once and therefore are sequenced at similar ratios on each sequencing lane and sequenced on the same sequencing lanes.

RNA-Seq Alignment and Transcript Quantification. We aligned all RNA-seq data to version 6.13 of the *D. melanogaster* reference genome using the STAR aligner version 020201 (35). We used the program's default settings and provided the appropriate genome annotation file for genome version 6.13. We sorted alignments based on coordinate positions using Samtools v1.3.1 (36). Finally, we obtained transcript counts for each annotated gene in the *D. melanogaster* genome using HTseq (37) and using the options "-s no" and "-r pos" but otherwise default program parameters.

Inversion Genotyping and Validation. To confirm inversion genotypes of each cross, we used two approaches. First, we coextracted DNA during RNA extractions following the Invitrogen TRIzol extraction protocol. We genotyped all samples using PCR following ref. 20. Second, we obtained genotype calls for each sample in the regions within 100 Kb of each inversion breakpoint. Breakpoint coordinates were extracted from their original publication (15, 20) and updated coordinates from *D. melanogaster* genome release version 5 to version 6 using the flybase coordinate conversion tool (flybase.org/static_pages/downloads/COORD.html). We generated genotype calls for all samples in these genomic regions using the Genome Analysis Toolkit v3.4–46-gbc02625 (38) following the best practices guidelines for genotyping RNA-seq data. We then performed a principal component analysis using Plink v1.9 (39), where we fit only a single principal component. We determined genotype concordance based on visual inspection of the eigenvalue distributions for each genotype class (Figs. S1 and S2), which indicated strong clustering within genotype classes consistent with the unique origins and decreased genetic variability within inversion-bearing chromosomes (13, 15, 31).

Transcript Filtering and Normalization. We performed DE analyses using the EdgeR package (24), where we first filtered genes for which the estimated counts per million transcripts was below 1 for more than half of the samples in a given comparison. Subsequently, we performed sampled normalization within the EdgeR package using the trimmed mean of M values (TMM) transcript normalization procedure (40). Prenormalization and postnormalization log counts per million are displayed as Figs. S3 and S4 and suggest that normalization was largely successful. We also evaluated the utility of the normalization method of DESeq2 (41) and found concordant results (i.e., we obtained strong overlap in DE transcripts across both natural inversion comparisons). In this work, we report the results obtained using the TMM method. Pre- and postnormalization log counts per million are shown in Figs. S3–S5 for natural and synthetic inversion comparisons.

DE. For each inversion comparison, we estimated the common, trended, and tagwise dispersion using the estimateDisp() function. We then fit a negative binomial generalized linear model to the normalized count data for each transcript, where we tested for an effect of arrangement genotype on expression levels. We retained genes that remained significantly differentially expressed in each comparison after applying a 20% false discovery rate correction. We performed all of these analyses within the EdgeR package. The full matrices and output for all genes included in this analysis are presented within Dataset S1.

Expression Patterns. We categorized each differentially expressed gene based on its expression pattern as additive, overdominant, or underdominant. If the inversion homozygote and inversion heterozygote both displayed higher or both displayed lower expression than the standard arrangement homozygote but the expression change associated with the inversion homozygote was the greater of the two, we termed the gene expression additive. If the

inversion heterozygote had a higher expression pattern than either arrangement homozygote, we assigned it to the class overdominant. Conversely, if the inversion heterozygote displayed a lower expression level than either arrangement homozygote, we defined the gene as underdominant. We note that these categories are not necessarily related to fitness, or even protein abundances, but describe only the expression patterns associated with a given gene across arrangements.

F_{ST} Analyses. Using the RNA-seq–based genotyping approach (above), we sought to identify genetically differentiated loci across the genomes of inversion-bearing individuals relative to the standard arrangement. We then computed F_{ST} (following ref. 42) for each SNP that had genotype data for at least 50 individuals, and we aggregated these data into nonoverlapping 25 SNP windows. Then, to determine if DE genes on different or the same chromosome arms as each common inversion are significantly associated with genetically differentiated regions, we performed permutation tests. Specifically, from among the set of genes that met our inclusion criteria for DE analysis, we selected at random the same number as we found to be DE on the same chromosome arms, and we compute the mean F_{ST} within this set. We then determined if this was greater than the value for the full set of truly DE genes. We performed the same test using data for the rest of the genome to determine if inadvertent linkage to *cis* elements genome-wide could also contribute to these DE effects. We performed all permutations 10,000 times.

GO Analysis. To identify biological commonalities among differentially expressed genes, we used the DAVID analysis framework (43) and applied it to the subset of differentially expressed genes that were identified in each natural inversion comparison. We provided the background set as the subset of the total genes whose read counts met our criterion for inclusion in our DE analysis (above; Dataset S1). All terms that remained significant at the 20% false discovery rate based on a Benjamini–Hochberg correction were retained as significant. A full list of the significant GO terms for each inversion is presented in Dataset S2.

Permutation Tests. To determine if there is an excess of overlap in the sets of differentially expressed loci shared between natural inversions, we performed permutation tests. Specifically, among the set of genes that were expressed at sufficient levels to meet our filtering criteria in both inversions, we randomly resampled an equal number of genes as we found to be differentially expressed within this set. We then asked in what proportion of 1,000 permutations an equal or greater number of genes were selected from each list.

Data and Reagent Availability. All synthetic inversion stocks created as a part of this study are available from the corresponding author upon request. All short read data are deposited within the Sequence Read Archive under accession no. PRJNA434443.

1. Hoffmann AA, Rieseberg LH (2008) Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol Syst* 39:21–42.
2. Kirkpatrick M (2010) How and why chromosome inversions evolve. *PLoS Biol* 8: e1000501–e1000505.
3. Charlesworth D, Charlesworth B, Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. *Heredity (Edinb)* 95:118–128.
4. Lee C-R, et al. (2017) Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat Ecol Evol* 1:119.
5. Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics* 173:419–434.
6. Sandler L, Hiraizumi Y, Sandler I (1959) Meiotic drive in natural populations of *Drosophila melanogaster*. I. The cytogenic basis of segregation-distortion. *Genetics* 44: 233–250.
7. Wang J, et al. (2013) A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* 493:664–668.
8. Naseeb S, et al. (2016) Widespread impact of chromosomal inversions on gene expression uncovers robustness via phenotypic buffering. *Mol Biol Evol* 33:1679–1696.
9. Fuller ZL, Haynes GD, Richards S, Schaeffer SW (2016) Genomics of natural populations: How differentially expressed genes shape the evolution of chromosomal inversions in Drosophila pseudoobscura. *Genetics* 204:287–301.
10. Huang W, et al. (2015) Genetic basis of transcriptome diversity in Drosophila melanogaster. *Proc Natl Acad Sci USA* 112:E6010–E6019.
11. Lavington E, Kern AD (2017) The effect of common inversion polymorphisms In(2L)t and In(3R)Mo on patterns of transcriptional variation in Drosophila melanogaster. *G3 (Bethesda)* 7:3659–3668.
12. Meadows LA, Chan YS, Roote J, Russell S (2010) Neighbourhood continuity is not required for correct testis gene expression in Drosophila. *PLoS Biol* 8:e1000552–12.
13. Wesley CS, Eanes WF (1994) Isolation and analysis of the breakpoint sequences of chromosome inversion In(3L)Payne in Drosophila melanogaster. *Proc Natl Acad Sci USA* 91:3132–3136.
14. Aulard S, David JR, Lemeunier F (2002) Chromosomal inversion polymorphism in Afrotropical populations of *Drosophila melanogaster*. *Genet Res* 79:49–63.
15. Corbett-Detig RB, Hartl DL (2012) Population genomics of inversion polymorphisms in Drosophila melanogaster. *PLoS Genet* 8:e1003056.
16. Mettler LE, Voelker RA, Mukai T (1977) Inversion clines in populations of Drosophila melanogaster. *Genetics* 87:169–176.
17. Knibb WR (1982) Chromosome inversion polymorphisms in Drosophila melanogaster II. Geographic clines and climatic associations in Australasia, North America and Asia. *Genetica* 58:213–221.
18. Golic KG, Golic MM (1996) Engineering the Drosophila genome: Chromosome rearrangements by design. *Genetics* 144:1693–1711.
19. Golic KG, Lindquist S (1989) The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome. *Cell* 59:499–509.
20. Corbett-Detig RB, Cardeno C, Langley CH (2012) Sequence-based detection and breakpoint assembly of polymorphic inversions. *Genetics* 192:131–137.
21. Pool JE, et al. (2012) Population genomics of Sub-Saharan Drosophila melanogaster: African diversity and non-African admixture. *PLoS Genet* 8:e1003080.
22. Ryder E, et al. (2004) The DrosDel collection: A set of P-element insertions for generating custom chromosomal aberrations in Drosophila melanogaster. *Genetics* 167: 797–813.
23. Eagen KP, Hartl TA, Kornberg RD (2015) Stable chromosome condensation revealed by chromosome conformation capture. *Cell* 163:934–946.
24. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
25. Tadin-Strapps M, et al. (2004) Cloning of the breakpoints of a de novo inversion of chromosome 8, inv (8)(p11.2q23.1) in a patient with Ambras syndrome. *Cytogenet Genome Res* 107:68–76.
26. Lemos B, Araripe LO, Fontanillas P, Hartl DL (2008) Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc Natl Acad Sci USA* 105: 14471–14476.
27. Krimbas CB, Powell JR (1992) *Drosophila Inversion Polymorphism* (CRC Press, Boca Raton, FL).
28. Kapun M, Fabian DK, Goudet J, Flatt T (2016) Genomic evidence for adaptive inversion clines in Drosophila melanogaster. *Mol Biol Evol* 33:1317–1336.
29. Ghosh R, Andersen EC, Shapiro JA, Gerke JP, Kruglyak L (2012) Natural variation in a chloride channel subunit confers avermectin resistance in C. elegans. *Science* 335: 574–578.
30. Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. *Nature* 400:667–671.
31. Andolfatto P, Wall JD, Kreitman M (1999) Unusual haplotype structure at the proximal breakpoint of In(2L)t in a natural population of *Drosophila melanogaster*. *Genetics* 153:1297–1311.
32. Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE (2016) A thousand fly genomes: An expanded Drosophila genome nexus. *Mol Biol Evol* 33:3308–3313.
33. Lack JB, et al. (2015) The Drosophila genome nexus: A population genomic resource of 623 Drosophila melanogaster genomes, including 197 from a single ancestral range population. *Genetics* 199:1229–1241.
34. Picelli S, et al. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* 10:1096–1098.
35. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.
36. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
37. Anders S, Pyl PT, Huber W (2015) HTSeq–A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.
38. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498.
39. Chang CC, et al. (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4:7–16.
40. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11:R25.
41. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
42. Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics* 132:583–589.
43. Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1–13.

EVOLUTION