Multiparent Recombinant Inbred lines crossed to a tester provide novel insights into sources of *cis* and *trans* regulation of gene expression

Fabio Marroni[1*], Alison M Morse[2,3], Adalena V Nanni[2,3], Nadja Nolte[4,5], Patricka Williams-Simon[6], Luis Leon Novelo[7], Rita M Graze[8], Paul Schmidt[9], Elizabeth King[10], Lauren M McIntyre[2,3*]

[1]Department of Agricultural, Food, Environmental and Animal Sciences, University of Udine, Udine, 33100, Italy

[2]Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL 32611, USA

[3]University of Florida Genetics Institute, University of Florida, Gainesville, FL 32611, USA

[4]Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana 1000, Slovenia

[5]Jožef Stefan International Postgraduate School, Ljubljana 1000, Slovenia

[6]Department of Biology, University of Pennsylvania, 433 S University Ave., 226 Leidy Laboratories, Philadelphia, PA 19104, USA

[7]Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston-School of Public Health, 1200 Pressler St. Suite E809, Houston, TX 77030, USA

[8]Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA

[9]University of Pennsylvania, Department of Biology, 433 S University Avenue, Philadelphia, PA, 19104

[10]Division of Biology, University of Missouri, 401 Tucker Hall, Columbia, MO 65211, USA

[*]Corresponding author: Department of Agricultural, Food, Environmental and Animal Sciences, University of Udine, Udine, 33100, Italy. Email: fabio.marroni@uniud.it

[*]Corresponding author: Department of Molecular Genetics and Microbiology and University of Florida Genetics Institute, University of Florida, Gainesville, FL 32611, USA. Email: mcintyre@ufl.edu

Abstract

We propose crossing multi-parent recombinant-inbred-lines (RILs) to a common tester and measuring allele specific gene expression in the offspring. With a few RIL x Tester crosses, genetic variation in the regulation of gene expression between RIL alleles can be identified. Testing whether allelic imbalance between two RIL x Tester crosses is equal, is a test of *cis* or *trans* depending on the RIL alleles compared. We demonstrate this approach in a long-read RNA-seq experiment in female abdominal tissue at two time points in *Drosophila melanogaster*. Among the 40% of all loci that show evidence of genetic variation in *cis, trans* effects due to the environment are detectable in 31% of loci and *trans* effects due to genetic background are detectable in 19% of loci with little overlap in sources of *trans* variation. The loci identified in this study are associated with loci previously reported to exhibit genetic variation in gene expression in a range of tissues and large population samples, suggesting that there is standing variation for genetic regulation of gene expression suggesting a potential source of cryptic genetic variation by which pleiotropic effects respond to changes in the environment. For loci in a QTL for thermotolerance, previously shown to differ in expression based on temperature; eleven of these loci had evidence for regulation of gene expression regardless of the environment, including Cpr67B, a cuticular protein. This study provides a blueprint for efficiently identifying alleles in multi-parent populations that contribute to regulatory variation in gene expression.

Introduction:

Spatial and temporal control of gene expression is known to be 'remarkably complex' (Arnone and Davidson 1997). The evolution of these complex components of gene expression regulation has been a subject of lively discussion since we have been able to empirically examine these questions (McGinnis et al. 1984; Scott and Weiner 1984; Krumlauf 1994). This involves unpacking the relative contributions of evolutionary changes in different components, for example contrasting protein coding versus *cis*-regulatory regions (Stern and Orgogozo 2008; Durkin et al. 2024) or *trans* regulatory factors; *cis* regulatory variation in promoters and enhancers, or *cis* by *trans* interactions (Mattioli et al. 2020; Ballinger et al. 2023; Hansen et al. 2024).

Allelic imbalance as a method to empirically estimate genetic variation in *cis* and *trans* regulation was laid out in a seminal paper in 2004 (Wittkopp et al. 2004). Measurements of allelic expression are affected by technical considerations such as map bias (Degner et al. 2009) and reference bias (Chen et al. 2021). In addition, tests for whether the source of differences in allelic expression are due to *cis* effects are usually tests between alleles within an individual, while to detect *trans* effects an allele's expression is typically compared between individuals (e.g. Wittkopp et al. 2004; Fear et al. 2016; Osada et al. 2017). *Trans* tests between individuals have lower power than the within individual *cis* effect tests for allelic imbalance and can be further confounded by differences in overall expression (León-Novelo et al. 2014; Sherbina et al. 2021). Differences in allelic imbalance in genetically identical heterozygotes compared between environments indicates *trans* variation from the environment must be the source of the difference in allelic imbalance. However, *cis* must be present in at least one of the two environments, implying that the *trans* due to the environment is likely interacting with the *cis* regulatory polymorphisms (Fear et. al. 2016, León-Novelo et al. 2018).

We propose a genetic design with crosses between multi-parent recombinant inbred lines (RILs) and an inbred tester. The tester should have a different genetic background from the RIL parents. This cross produces genotypes that are heterozygous at most loci. When there is allelic imbalance in expression within the individual there are *cis* regulatory polymorphisms, and potential for variation in *trans* to interact with the regulatory polymorphisms in *cis*. The test of the null hypothesis that the allelic imbalance between two RIL x Tester crosses is equal compares individuals to determine if the imbalance differs between them. Given the presence of *cis* variation, this test detects additional *cis* or *trans* sources of regulatory variation depending on the genotypes of the RILs at the locus (**Figure 1**). The test for differences in allelic imbalance

has been shown to efficiently identify regulatory variation in crosses among inbred lines, since the number of alleles to be evaluated (**n**) can be screened in **n**-crosses instead of the $(n^2-n)/2$ crosses in the diallel design which evaluates all alleles pairwise (Miller et al. 2021). Here we show that RIL lines, in addition to the efficiency of evaluating the parental alleles when crossed to a common tester, provide additional novel insights into the relative sources of *cis* and *trans* regulatory variation.

The design proposed here, a multiparent RIL crossed to a tester, leverages the community's extensive investment in multi-parent genetic resources. Multi-parent genetic resource panels have been developed for many plant breeding applications in many different systems including maize (Dell'Acqua et al. 2015), rice (Raghavan et al. 2017), wheat (Mackay et al. 2014), sorghum (Ongom and Ejeta 2018), barley (Sannemann et al. 2015), soybean (Hashemi et al. 2022), strawberry (Mangandi et al. 2017), and okra (Sandeep et al. 2025). Multi-parent populations have also been developed for QTL mapping in many model organisms such as Arabidopsis (Kover et al. 2009), Drosophila (King et al. 2012), *C. elegans* (Noble et al. 2017), mouse (Srivastava et al. 2017), and yeast (Cubillos et al. 2013).

Since the first results in QTL mapping (Lander and Botstein 1989), the field has grappled with how to move from QTL regions to direct links between genes and phenotypes, although there are caveats about whether it is useful to focus on loci of large effect (Rockman 2012). There have also been proposals to combine QTL/GWAS, and eQTL to uncover regulatory networks underpinning complex traits (e.g. Jansen et al. 2009; Zhu et al. 2016) and differences in molecular phenotypes like gene expression may assist in prioritizing candidate loci (Wayne and McIntyre 2002).

In this study, we provide a proof of principle for efficiently identifying alleles that differ in regulatory variation for gene expression. By analyzing differences in allelic imbalance at each locus we identify genetic drivers of regulatory variation in gene expression. This is different from mapping eQTL for overall expression which instead identify 'local' and 'distal' sources of variation in expression (Rockman and Kruglyak 2006). eQTL studies, like QTL studies require hundreds of samples to map variation in expression phenotypes whereas studies of allele imbalance can be conducted within an individual, and *n* alleles can be assayed in *n* crosses.

We conducted a replicated long read RNA-seq experiment in *Drosophila melanogaster*, where we measure allelic imbalance. We selected four multi-parent RIL lines to cross to a common tester, and by comparing pairs of RIL x Tester crosses we identify genetic sources of regulatory

variation in *cis* and *trans.* A comparison between two time points identifies environmental sources of *trans* effects. We used a Bayesian model that accounts for difference in overall expression, as well as potential differential map bias between alleles (León-Novelo et al. 2018). We link results to a thermotolerance QTL (Williams-Simon et al. 2024).

Results

In order to test for allelic imbalance, there needs to be polymorphism in the exonic regions of the gene, enabling the assignment of reads to alleles. Almost all of the genes in *D. melanogaster* 6.50 (Öztürk-Çolak et al. 2024) had at least 1 SNP in either the tester or the RIL. There were only 538, 614, 498, 548 loci in dm12279, dm12272, dm11255, dm11037 respectively that could not be evaluated for allelic imbalance due to lack of polymorphisms between the tester and RIL haplotype.

Long read sequencing

We generated 128,001,025 million long reads. The median number of reads per sample was 4,671,327 with ~11,000 genes detected. (**Table S1**). All but one of the samples were found to be of similar and good quality. The dm12279 replicate 1 sample at 0 to 1 hour had a lower proportion of mapped reads compared to the other samples (**Table S1)**. The long reads enabled a high proportion of the reads to be assigned to an allele, and a median of 3910 genes were tested for allelic imbalance (**Table S2**).

We test allelic imbalance between two RIL X Tester crosses (**Figure 1)**. A pre-condition for the comparison is the significant effect of allelic imbalance in at least one cross. The inference about the source of the difference in allelic imbalance, *cis* or *trans*, depends on the comparison being made. If reciprocal crosses are compared at autosomal loci, the alleles are the same between the reciprocals, but they are inherited from two different parents. The null is that the parent of origin has no effect on allelic imbalance. The parent of origin effect is a *trans* effect that impacts the *cis* regulation (Miller et al. 2021, Puixeu et. al. 2023). For example, if the methylation status of one of the alleles is altered affecting expression, the source of that alteration is a *trans* factor.

***cis* and *trans* sources of regulatory variation.** The proportion of genes showing allelic imbalance in the test H_{01} (within individual cross) is 28% and across all 4 RIL x Tester crosses in both environments we detect *cis* effects in ~40% of genes (**Table S4**). The RIL x Tester

design is efficient because every test of H_{01} is a test of *cis* and every pairwise test of H_{03} between RIL x Tester crosses is a test of the source of regulatory variation.

We condition the tests between individuals (H_{03}) on the test of H_{01}. When the RIL alleles differ between crosses, the test between individuals is a test of *cis* differences between the RIL alleles, and *trans* effects from the different backgrounds that interact with *cis* differences among the three alleles. We see evidence for *cis* differences between RIL alleles at 63% of the loci tested. When the RIL haplotype is shared between individuals, the *cis* effects are shared between the two crosses, as in a reciprocal cross. Differences in allelic imbalance between RIL x Tester crosses with the same genotype at the expressed locus are due to *trans* sources from the genetic background. We find evidence for a genetic effect in *trans* for 19% of the loci tested. The test statistic used in the test of *cis* and *trans*, is the same and differences are unlikely to be the result of a difference in power. Interestingly, all but 49 of the genes with evidence for *trans* also have evidence for *cis*, suggesting that allelic interactions between *cis* and *trans* sources of variation are prevalent*. (***Figure 2***)*

We examined sources of *cis* and *trans* for each pair of RIL x Tester crosses. We see a consistent pattern (more *cis* than *trans*) for each genotype and in both environments (**Figure 3**). We also see consistent estimates of the relative proportion of *cis* and *trans* for both the higher coverage in 3-8 days and lower coverage 0-1 hours, although consistent with expectations higher coverage improves detection of both sources of regulatory variation.

The environment is a *trans* effect. The difference in presence/absence of gene expression is also expected to be due to *trans* factors attributable to the environment. There were 16,834 genes in the annotation. Of these 13,804 were detected in either environment (0-1 hours) or (3-8 days). In order to account for the difference in coverage between the 0-1- hour and the 3-8 day samples, we used the genes detected in the 0-1 hour that are absent in the 3-8 day samples as an estimate of *trans* presence/absence (*n=355*). Assuming the effect is symmetric between the two time points, this would suggest that ~700 genes or 5% of genes differ in presence/absence *trans* effects. This relatively small difference in presence/absence compared to many tissue comparisons, is likely due to the similarity of the two environments. For loci expressed at both time points with evidence of H_{01}, we estimate that (320) 31% of the loci show *trans* effects due to the environment. There is little overlap between loci with *trans* genetic and *trans* environment. Of the 561 loci that could be evaluated for both effects, 9% showed evidence for both, and 42% showed evidence for either.

Candidate loci for thermotolerance. There are genes with temperature sensitive expression (in head tissue), that are also differentially expressed between the high and low parental lines (head tissue) in a QTL for thermotolerance (Williams-Simon et al. 2024). If one of these genes showed evidence for genetic regulatory variation (*cis* or *trans*) and an absence of *trans* effects between the environments, we consider this additional evidence of genetic regulatory variation worth considering in potential follow-up studies for understanding thermotolerance. There are 11 genes that meet these criteria (**Table S3**) including FBgn0035985, Cpr67B. This gene makes a cuticular protein, and is expressed in both the head and the crop, hindgut and rectal pad according to FlyAtlas (Krause et al. 2022). The gene contains an R&R motif (Rebers and Riddiford 1988), which can bind chitin in some circumstances (Togawa et al. 2004). How the other genes identified may be involved in thermotolerance is less obvious, some are linked to developmental transitions and therefore variation in these loci may affect the cuticle, and four of these are not named, and would not likely be selected for follow-up based on prior biological knowledge (FBgn0035983, FBgn0038337, FBgn0039656, FBgn0039719).

The RILxTester design is an unbiased screen for regulatory variation with respect to the QTL. We do not expect there to be an excess of regulatory variation in the QTL, and consistent with that expectation, the proportion of tests for H_{01} *cis* in the QTL region between the lines with the 'high' thermotolerance QTL and lines with 'low' thermotolerance QTL, is similar to the proportion of tests significant in the genome.

Conclusion

Combining independent data in an unbiased way has the potential to reveal novel relationships between genetic and phenotypic variation. Studies of allelic imbalance allow a direct inference of the presence of regulatory variation in gene expression between alleles at a particular locus. Further focusing on loci with regulatory effects that are not influenced by the environment increases the chances of identifying stable regulatory effects. While this is an efficient screen for winnowing down large numbers of candidate genes, it does not guarantee success, as regulation of gene expression is only one of many possible sources of phenotypic variation for any complex trait. Hunting for individual candidate genes and testing them for phenotypic effects assumes that there are large effect loci underneath the QTL worth identifying. There are many cases for which this has been a successful strategy. However, the importance of large numbers of small effect polygenes to quantitative variation remains a potent alternative hypothesis.

Genetic variation that impacts the regulation of gene expression is a potential underlying mechanism by which small effect polymorphisms could act on phenotype. The hypothesis here is that *cis* regulatory variation in allelic expression provides a background standing genetic variation that can be immediately be amplified by *trans* effects in the environment. This hypothesis has been discussed in many contexts since its initial proposal (Jacob 1977). One factor contributing to the ongoing difficulty in evaluating these hypotheses is the lack of efficient designs that enable a survey of a large number of alleles and provide a statistically powerful approach to interrogating *cis* and *trans* effects. The RILxTester design we propose here enables the evaluation of **n** alleles in **n** crosses and by comparing these crosses pairwise can be used to compare the **n** alleles for both *cis* and *trans* effects. This is markedly fewer assays than an eQTL study.

We compared our results to other studies of genetic variation in gene expression (**Table S5**). The genes associated with *cis* effects in this study, were enriched for genes showing evidence for heritability of gene expression in a diallel (Wayne et al. 2007); *cis* and *trans* effects in comparisons among RI lines of OregonR and 2b3 (Genissel et al. 2008), *cis* effects as estimated by allelic imbalance within an individual in a testcross with W1118 and DGRP lines (León-Novelo et al. 2018) and *cis* effects identified in F1's of interspecific hybrid crosses (Graze et al. 2012). This is a remarkable consistency of evidence for genetic variation in regulation of gene expression, across platforms (microarray, short read RNA-seq and long read RNA-seq), tissue types (head, whole body, abdomen), and experimental designs. This could be because different variants at the same locus contribute in multiple studies due to high levels of standing variation due to balancing selection, relaxed selection or variation maintained by mutation-selection balance under stabilizing selection. Another possibility is that allelic variants that impact expression produce a signal that in combination with the statistical approach is unusually robust to tissue, environment or other microenvironmental variation. In either case, this is support for the hypothesis that genetic regulation of gene expression may be one mechanism by which pleiotropic effects are regulated.

Methods

For a QTL analysis of thermotolerance, the most significant Quantitative Trait Loci (QTL) region identified by Williams-Simon et al. (Williams-Simon et al. 2024) was used to randomly select two RIL haplotypes with the shared allele for 'high' thermotolerance in QTL 1 (A5: 12272, 12279) and two RIL haplotypes with the 'low' thermotolerance allele (A4 : 11037 , 11255).

Fly Husbandry Populations were housed in narrow vials (Genesee Scientific – 32-116SB) at 25°C and 60% relative humidity, on a modified Bloomington Cornmeal Recipe (Beltz et al. 2025). The four RILs selected from the *Drosophila* Synthetic Population Resource (DSRP) (King et al. 2012) were crossed with a common female genotype W1118v(*drosophila stock center stock ID*). Each cross consisted of 10 female W1118 and 5 RIL male pupae. 72 hours after mating 60-70 F1 eggs were transferred to a new vial with fresh food. At 10 days post oviposition, 0-1 hour virgin females were collected by observing eclosure in 1 hour increments. Newly eclosed females were either dissected for abdominal tissue and placed in RNAlater (Sigma – MKCJ9161) + 100uL 0.1% PBST and at -20°C, or transferred to a new vial with fresh food. After 3-8 days, ovaries were removed (gonadectomized) and flies placed in RNAlater and stored at -20°C . Three replicates of ~50 individuals were collected for each cross and time point.

Library Preparation and Sequencing For each sample, mRNA was isolated (DynaBeads mRNA direct kit) from a pool of ~20 abdomens. ONT libraries were constructed using the ONT PCR-cDNA Barcoding Kit (SQK-PCB109) starting with polyA mRNA according to the manufacturer's protocol. Libraries were pooled to a total of 100fmol and run on a MinION Mk1c with real-time basecalling and demultiplexing (Guppy v6.1.5, MinKNOW v22.05.8). Read length and quality was evaluated with all samples passing the PycoQC metrics v2.5.2 (Leger and Leonardi 2019). Based on the MinION read counts, libraries were repooled prior to obtaining additional sequencing data on the ONT PromethION (Guppy v5.1.13, MinKNOW v23.04.5) at the University of Florida Interdisciplinary Center for Biotechnology Research (ICBR). Technical replicates (TRs) are defined as the same library run on different ONT flow cells (MinION or PromethION). TRs 1-3 were run on the MinION and TRs 4-6 were run on the PromethION.

Dorado basecalled reads were generated from the fast5 files by converting to pod5 formats (pod5 v 0.3.6) prior to basecalling by Dorado (v 0.5.2) (https://github.com/nanoporetech/dorado) using options --recursive --device "cuda:0,1" --kit-name SQK-PCB109 --trim none. Reads were demultiplexed using the demux mode of Dorado (v 0.5.2) with options --no-classify –emit-fastq. Fastq files are available at the SRA under BioProject PRJNA1134728. The fastq files generated by were processed using pychopper (v 2.7.1), the oriented fastq files were aligned to *D. melanogaster* 6.50 and the resulting sam files were converted to gtf using samtools (v 1.10) (Li et al. 2009) and bedtools (v 2.29.2) (Quinlan and Hall 2010). The resulting gtf files (67 technical replicates from 24 samples), the *D. melanogaster* 6.50 fasta reference file

(https://ftp.flybase.net/releases/FB2023_01/dmel_r6.50/ (Öztürk-Çolak et al. 2024)), and a design file were used as input to SQANTI-reads (Keil et al. 2025).

Construction of haplotype specific references SNP data for w1118 and the RILS were retrieved from the *Drosophila* Synthetic Population Resource data (King et al. 2012; Fear et al. 2016); genomic coordinates were converted from release 5 (Hoskins et al. 2007) to release 6 (Hoskins et al. 2015) using the coordinate converter tool in Flybase (Gramates et al. 2022). As part of quality control, 11, 5, 6, and 6 SNP positions were discarded in RIL11037, RIL11255, RIL12272, and RIL12279 respectively, because the reference base in the VCF file did not match the reference base in the genome. There were fewer SNPs in the w1118 VCF file compared to any of the individual RIL genotypes.

The dmel6 reference genome (Hoskins et al. 2015) was updated with w1118 and RIL specific SNPs respectively (Graze et al. 2012; Munger et al. 2014; Gobet et al. 2022). Following approaches published for other species, we used the alignments to the RIL parent, to evaluate and correct for missing SNP calls (Glinos et al. 2022). Long-reads were aligned and positions expected to have a SNP in the RIL but not the w1118, that is expected to be heterozygous were evaluated as follows. If there were a min of 5 reads and no evidence of polymorphism among the reads, the SNP was presumed to be a missed call in w1118 and the w1118 VCF file was updated to include the SNP. Haplotype specific gene regions (+/- 100 base pairs from the annotated transcription start and end site) were identified. Reads were mapped separately to the parental haplotype specific gene regions. Uniquely mapping reads were compared and identified as allele specific if they mapped with fewer mismatches to one parent (Graze et al. 2012) using the python script sam_compare_w_feature.py (Miller et al. 2021) available at https://github.com/McIntyre-Lab/BayesASE) and no haplotype bias was observed. The resulting allele counts are used as input into the Bayesian Model, and results are provided in **Supplementary File 1**.

Statistical Tests

Differences in allelic imbalance was deemed to be statistically significant using a Bayesian credible interval from the model bayesASE (León-Novelo et al. 2018; Miller et al. 2021). The statistical test used here has been explained in detail (León-Novelo et al. 2018). Briefly, under condition 1, the proportion of reads produced by the reference allele is θ_1 and allelic imbalance in the genotype/condition 1 occurs when $\theta_1 \neq 1/2$. θ_2 is defined analogously to θ_1 but under condition 2. The model bayesASE tests for allelic imbalance under condition 1 (H_{01}: $\theta_1 = 1/2$ vs

$H_{01}: \theta_1 \neq 1/2$), allelic imbalance under condition 2 ($H_{02}: \theta_2 = 1/2$ vs $H_{02}: \theta_2 \neq 1/2$) and if the levels of allelic imbalance differ across the two conditions differ ($H_{03}: \theta_3 = 0$ vs $H_{03}: \theta_3 \neq 0$ with $\theta_3 = \theta_1 - \theta_2$). The test of θ_1 is the test within individuals and the comparison between individuals is θ_3. The model explicitly accounts for differences in overall expression and uses a prior to account for potential map bias, including potential differences in mapping rates between the two crosses. We estimated the prior probability from the RNA long-read data using the Prior Calculation module from BayesASE (Miller et al. 2021). All comparisons were carried out with 100,000 iterations and a burn in of 10,000.

The statistical test θ_3 is used to evaluate *cis* and *trans* effects of the RIL alleles. As the three comparisons (*cis* genetic between individuals, *trans* genetic between individuals, and *trans* due to the environment) use the same statistical approach, the difference in detection will be due to effect size differences and not the properties of the test statistic.

Data Availability

Fastq files are deposited to the SRA BioProject PRJNA1134728. Results from the Bayesian model are available in individual files labeled by the test performed in Supplementary File 1.

Funding

This work was funded by GM GM137430 (LM, PS), R01GM128193 (LM) R35 GM149238 (EK), R35 GM133376 (RR), the National science Foundation Career award (RG), the LongTrec European Marie Curie Action Network (NN).

Acknowledgements

The Department of Molecular Genetics and Microbiology, The University of Florida Cancer Center, The University of Florida Genetics Institute, University of Florida Research Computing, HiPerGator.

Figure 1: *Cis*- and *trans*- sources of regulatory variation in tests of allelic imbalance. **A:** Comparing expression of the two alleles of one individual (RIL 1 vs tester, θ_1, RIL 2 vs tester , θ_2) tests for regulatory variation due to polymorphisms in *cis* **B:** Comparing allelic imbalance

between different conditions for the same genotypes (θ_3) tests for differences in *cis*-effects results from environmental changes in *trans* regulation. **C:** Comparing allelic imbalance between different genotypes. When the RIL allele is the same between two crosses, θ_3 is due to differences in the *trans* regulatory background (θ_3 *trans* source). D**:** Comparing allelic imbalance between different genotypes. If the RIL genotype is not shared, then θ_3 is a comparison between these RIL alleles; this is a test of a *cis*- effect, and will include some interactions due to *trans*-factors interacting with *cis* differences between the RILs (θ_3 *cis* + *trans* sources).

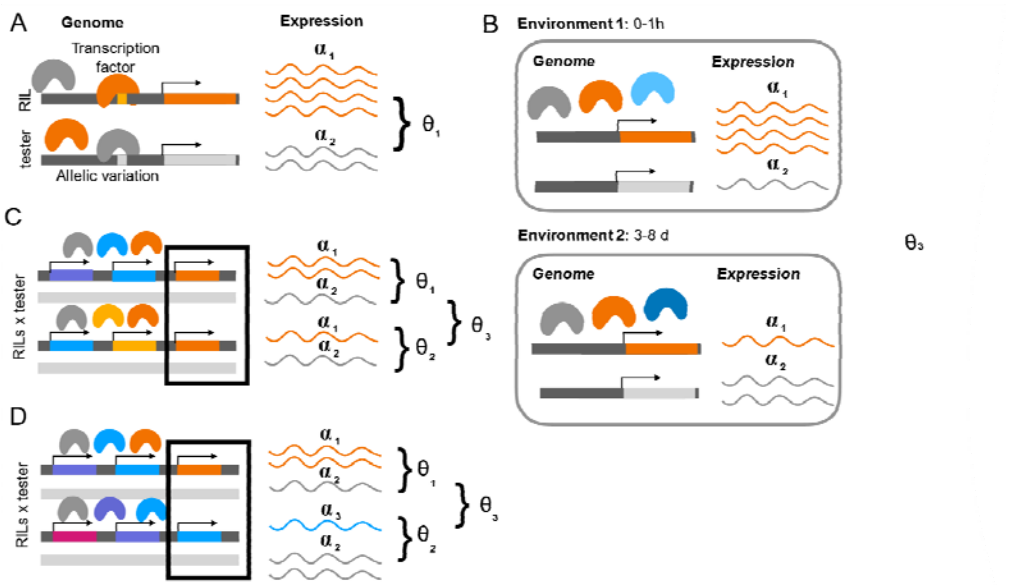


Figure 2: Number of genes for which H_{03} was rejected when comparing different genotypes sharing the RIL allele being tested (*Trans geno*), the same genotype in different times (*Trans env*) and different genotypes not sharing the RIL allele being tested (*Cis between*). The number of genes shared between the two tests for *trans* effect is shown in dark red. The number of genes shared between the test for *cis* between individuals and the test for *cis* within individuals is shown in dark blue.

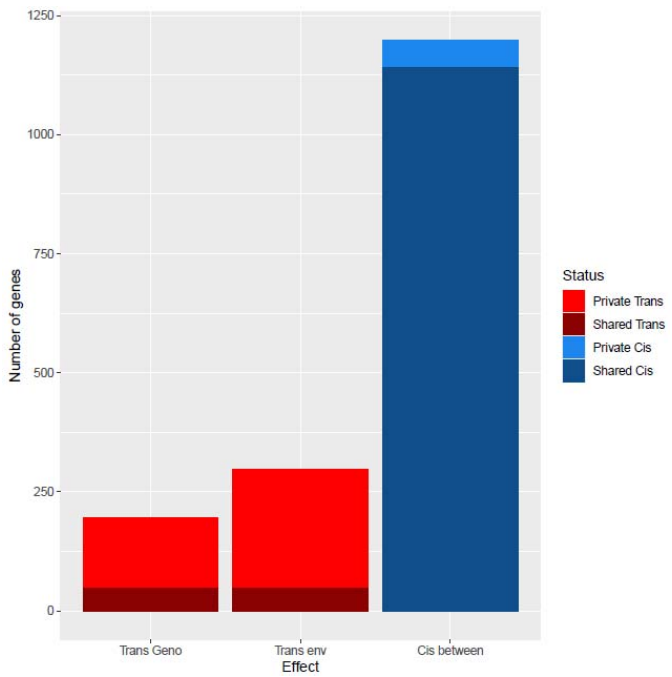
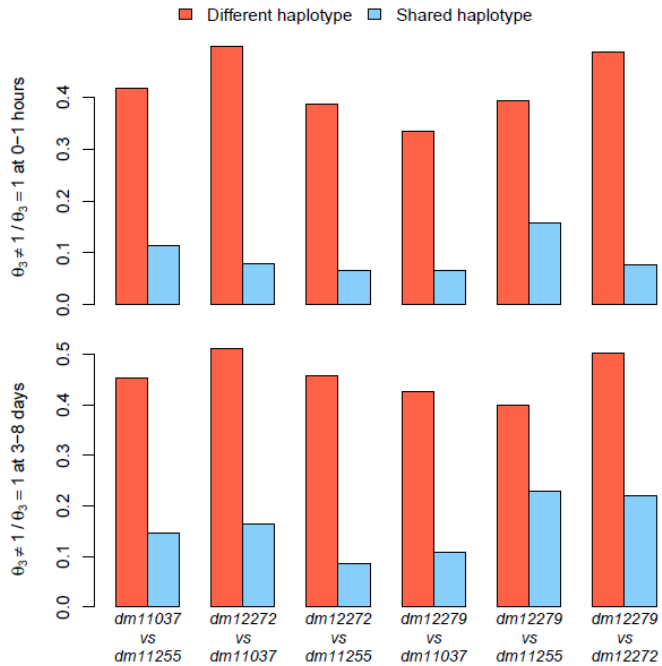


Figure 3: Proportion of genes showing different levels of allelic imbalance between lines (Rejection of H_{03} θ_3 is not zero), limited to regions in which the lines have different (red) or same (blue) haplotype.



Literature cited

Arnone MI, Davidson EH. 1997. The hardwiring of development: organization and function of genomic regulatory systems. Development. 124(10):1851–1864. doi:10.1242/dev.124.10.1851.

Ballinger MA, Mack KL, Durkin SM, Riddell EA, Nachman MW. 2023. Environmentally robust cis-regulatory changes underlie rapid climatic adaptation. Proceedings of the National Academy of Sciences. 120(39):e2214614120. doi:10.1073/pnas.2214614120.

Beltz JK, Bitter MC, Goldfischer A, Petrov DA, Schmidt P. 2025. Variation in Resource Environment Drives Adaptive Divergence in Drosophila melanogaster. :2024.12.10.626603. doi:10.1101/2024.12.10.626603. [accessed 2025 Apr 13]. https://www.biorxiv.org/content/10.1101/2024.12.10.626603v2.

Chen N-C, Solomon B, Mun T, Iyer S, Langmead B. 2021. Reference flow: reducing reference bias using multiple population genomes. Genome Biology. 22(1):8. doi:10.1186/s13059-020-02229-3.

Cubillos FA, Parts L, Salinas F, Bergström A, Scovacricchi E, Zia A, Illingworth CJR, Mustonen V, Ibstedt S, Warringer J, et al. 2013. High-Resolution Mapping of Complex Traits with a Four-Parent Advanced Intercross Yeast Population. Genetics. 195(3):1141–1155. doi:10.1534/genetics.113.155515.

Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, Pritchard JK. 2009. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. Bioinformatics. 25(24):3207–3212. doi:10.1093/bioinformatics/btp579.

Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, Baute J, et al. 2015. Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biology. 16(1):167. doi:10.1186/s13059-015-0716-z.

Durkin SM, Ballinger MA, Nachman MW. 2024. Tissue-specific and cis-regulatory changes underlie parallel, adaptive gene expression evolution in house mice. PLOS Genetics. 20(2):e1010892. doi:10.1371/journal.pgen.1010892.

Fear JM, León-Novelo LG, Morse AM, Gerken AR, Van Lehmann K, Tower J, Nuzhdin SV, McIntyre LM. 2016. Buffering of Genetic Regulatory Networks in Drosophila melanogaster. Genetics. 203(3):1177–1190. doi:10.1534/genetics.116.188797.

Genissel A, McIntyre LM, Wayne ML, Nuzhdin SV. 2008. Cis and Trans Regulatory Effects Contribute to Natural Variation in Transcriptome of Drosophila melanogaster. Molecular Biology and Evolution. 25(1):101–110. doi:10.1093/molbev/msm247.

Glinos DA, Garborcauskas G, Hoffman P, Ehsan N, Jiang L, Gokden A, Dai X, Aguet F, Brown KL, Garimella K, et al. 2022. Transcriptome variation in human tissues revealed by long-read sequencing. Nature. 608(7922):353–359. doi:10.1038/s41586-022-05035-y.

Gobet N, Jan M, Franken P, Xenarios I. 2022. Towards mouse genetic-specific RNA-sequencing read mapping. PLoS Comput Biol. 18(9):e1010552. doi:10.1371/journal.pcbi.1010552.

Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby MA, dos Santos G, Goodman JL, Goutte-Gattat D, Jenkins VK, Kaufman T, et al. 2022. FlyBase: a guided tour of highlighted features. Genetics. 220(4):iyac035. doi:10.1093/genetics/iyac035.

Graze RM, Novelo LL, Amin V, Fear JM, Casella G, Nuzhdin SV, McIntyre LM. 2012. Allelic imbalance in Drosophila hybrid heads: exons, isoforms, and evolution. Mol Biol Evol. 29(6):1521–1532. doi:10.1093/molbev/msr318.

Hansen TJ, Fong SL, Day JK, Capra JA, Hodges E. 2024. Human gene regulatory evolution is driven by the divergence of regulatory element function in both cis and trans. Cell Genomics. 4(4). doi:10.1016/j.xgen.2024.100536. [accessed 2025 Mar 16]. https://www.cell.com/cell-genomics/abstract/S2666-979X(24)00092-2.

Hashemi SM, Perry G, Rajcan I, Eskandari M. 2022. SoyMAGIC: An Unprecedented Platform for Genetic Studies and Breeding Activities in Soybean. Front Plant Sci. 13. doi:10.3389/fpls.2022.945471. [accessed 2025 Apr 27]. https://www.frontiersin.orghttps://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.945471/full.

Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E, Wan KH, Park S, Mendez-Lago M, Rossi F, et al. 2007. Sequence Finishing and Mapping of Drosophila melanogaster Heterochromatin. Science. 316(5831):1625–1628. doi:10.1126/science.1139816.

Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R, et al. 2015. The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Res. 25(3):445–458. doi:10.1101/gr.185579.114.

Jacob F. 1977. Evolution and Tinkering. Science. 196(4295):1161–1166. doi:10.1126/science.860134.

Jansen RC, Tesson BM, Fu J, Yang Y, McIntyre LM. 2009. Defining gene and QTL networks. Current Opinion in Plant Biology. 12(2):241–246. doi:10.1016/j.pbi.2009.01.003.

Keil N, Monzó C, McIntyre L, Conesa A. 2025. Quality assessment of long read data in multisample lrRNA-seq experiments using SQANTI-reads. Genome Res. doi:10.1101/gr.280021.124.

King EG, Macdonald SJ, Long AD. 2012. Properties and Power of the Drosophila Synthetic Population Resource for the Routine Dissection of Complex Traits. Genetics. 191(3):935–949. doi:10.1534/genetics.112.138537.

Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Mott R. 2009. A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in Arabidopsis thaliana. PLOS Genetics. 5(7):e1000551. doi:10.1371/journal.pgen.1000551.

Krause SA, Overend G, Dow JAT, Leader DP. 2022. FlyAtlas 2 in 2022: enhancements to the Drosophila melanogaster expression atlas. Nucleic Acids Research. 50(D1):D1010–D1015. doi:10.1093/nar/gkab971.

Krumlauf R. 1994. Hox genes in vertebrate development. Cell. 78(2):191–201. doi:10.1016/0092-8674(94)90290-9.

Lander ES, Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 121(1):185–199. doi:10.1093/genetics/121.1.185.

Leger A, Leonardi T. 2019. pycoQC, interactive quality control for Oxford Nanopore Sequencing. Journal of Open Source Software. 4(34):1236. doi:10.21105/joss.01236.

León-Novelo L, Gerken AR, Graze RM, McIntyre LM, Marroni F. 2018. Direct Testing for Allele-Specific Expression Differences Between Conditions. G3: Genes, Genomes, Genetics. 8(2):447–460. doi:10.1534/g3.117.300139.

León-Novelo LG, McIntyre LM, Fear JM, Graze RM. 2014. A flexible Bayesian method for detecting allelic imbalance in RNA-seq data. BMC Genomics. 15:920. doi:10.1186/1471-2164-15-920.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

Mackay IJ, Bansept-Basler P, Barber T, Bentley AR, Cockram J, Gosman N, Greenland AJ, Horsnell R, Howells R, O'Sullivan DM, et al. 2014. An Eight-Parent Multiparent Advanced Generation Inter-Cross Population for Winter-Sown Wheat: Creation, Properties, and Validation. G3 Genes|Genomes|Genetics. 4(9):1603–1610. doi:10.1534/g3.114.012963.

Mangandi J, Verma S, Osorio L, Peres NA, van de Weg E, Whitaker VM. 2017. Pedigree-Based Analysis in a Multiparental Population of Octoploid Strawberry Reveals QTL Alleles Conferring Resistance to Phytophthora cactorum. G3 Genes|Genomes|Genetics. 7(6):1707–1719. doi:10.1534/g3.117.042119.

Mattioli K, Oliveros W, Gerhardinger C, Andergassen D, Maass PG, Rinn JL, Melé M. 2020. Cis and trans effects differentially contribute to the evolution of promoters and enhancers. Genome Biology. 21(1):210. doi:10.1186/s13059-020-02110-3.

McGinnis W, Garber RL, Wirz J, Kuroiwa A, Gehring WJ. 1984. A homologous protein-coding sequence in drosophila homeotic genes and its conservation in other metazoans. Cell. 37(2):403–408. doi:10.1016/0092-8674(84)90370-2.

Miller BR, Morse AM, Borgert JE, Liu Z, Sinclair K, Gamble G, Zou F, Newman JRB, León-Novelo LG, Marroni F, et al. 2021. Testcrosses are an efficient strategy for identifying cis-regulatory variation: Bayesian analysis of allele-specific expression (BayesASE). G3 Genes|Genomes|Genetics. 11(5):jkab096. doi:10.1093/g3journal/jkab096.

Munger SC, Raghupathy N, Choi K, Simons AK, Gatti DM, Hinerfeld DA, Svenson KL, Keller MP, Attie AD, Hibbs MA, et al. 2014. RNA-Seq alignment to individualized genomes improves transcript abundance estimates in multiparent populations. Genetics. 198(1):59–73. doi:10.1534/genetics.114.165886.

Noble LM, Chelo I, Guzella T, Afonso B, Riccardi DD, Ammerman P, Dayarian A, Carvalho S, Crist A, Pino-Querido A, et al. 2017. Polygenicity and Epistasis Underlie Fitness-Proximal Traits in the Caenorhabditis elegans Multiparental Experimental Evolution (CeMEE) Panel. Genetics. 207(4):1663–1685. doi:10.1534/genetics.117.300406.

Ongom PO, Ejeta G. 2018. Mating Design and Genetic Structure of a Multi-Parent Advanced Generation Intercross (MAGIC) Population of Sorghum (Sorghum bicolor (L.) Moench). G3 Genes|Genomes|Genetics. 8(1):331–341. doi:10.1534/g3.117.300248.

Osada N, Miyagi R, Takahashi A. 2017. Cis- and Trans-regulatory Effects on Gene Expression in a Natural Population of Drosophila melanogaster. Genetics. 206(4):2139–2148. doi:10.1534/genetics.117.201459.

Öztürk-Çolak A, Marygold SJ, Antonazzo G, Attrill H, Goutte-Gattat D, Jenkins VK, Matthews BB, Millburn G, dos Santos G, Tabone CJ, et al. 2024. FlyBase: updates to the Drosophila genes and genomes database. Genetics. 227(1):iyad211. doi:10.1093/genetics/iyad211.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841–842. doi:10.1093/bioinformatics/btq033.

Raghavan C, Mauleon R, Lacorte V, Jubay M, Zaw H, Bonifacio J, Singh RK, Huang BE, Leung H. 2017. Approaches in Characterizing Genetic Structure and Mapping in a Rice Multiparental Population. G3 Genes|Genomes|Genetics. 7(6):1721–1730. doi:10.1534/g3.117.042101.

Rebers JE, Riddiford LM. 1988. Structure and expression of a Manduca sexta larval cuticle gene homologous to Drosophila cuticle genes. J Mol Biol. 203(2):411–423. doi:10.1016/0022-2836(88)90009-5.

Rockman MV. 2012. The Qtn Program and the Alleles That Matter for Evolution: All That's Gold Does Not Glitter. Evolution. 66(1):1–17. doi:10.1111/j.1558-5646.2011.01486.x.

Rockman MV, Kruglyak L. 2006. Genetics of global gene expression. Nat Rev Genet. 7(11):862–872. doi:10.1038/nrg1964.

Sandeep N, Kumar BMD, Chavan AL. 2025. Microsatellite Marker–Driven Genetic Diversity and Breeding Potential Assessment in Okra (Abelmoschus esculentus (L.) Moench): A Multi-parent Approach. Plant Mol Biol Rep. doi:10.1007/s11105-025-01561-x. [accessed 2025 Apr 27]. https://doi.org/10.1007/s11105-025-01561-x.

Sannemann W, Huang BE, Mathew B, Léon J. 2015. Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. Mol Breeding. 35(3):86. doi:10.1007/s11032-015-0284-7.

Scott MP, Weiner AJ. 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proceedings of the National Academy of Sciences. 81(13):4115–4119. doi:10.1073/pnas.81.13.4115.

Sherbina K, León-Novelo LG, Nuzhdin SV, McIntyre LM, Marroni F. 2021 Jul 12. Power Calculator for Detecting Allelic Imbalance Using Hierarchical Bayesian Model. biorXiv.:2021.07.10.451873.

Srivastava A, Morgan AP, Najarian ML, Sarsani VK, Sigmon JS, Shorter JR, Kashfeen A, McMullan RC, Williams LH, Giusti-Rodríguez P, et al. 2017. Genomes of the Mouse Collaborative Cross. Genetics. 206(2):537–556. doi:10.1534/genetics.116.198838.

Stern DL, Orgogozo V. 2008. THE LOCI OF EVOLUTION: HOW PREDICTABLE IS GENETIC EVOLUTION? Evolution. 62(9):2155–2177. doi:10.1111/j.1558-5646.2008.00450.x.

Togawa T, Nakato H, Izumi S. 2004. Analysis of the chitin recognition mechanism of cuticle proteins from the soft cuticle of the silkworm, Bombyx mori. Insect Biochem Mol Biol. 34(10):1059–1067. doi:10.1016/j.ibmb.2004.06.008.

Wayne ML, McIntyre LM. 2002. Combining mapping and arraying: An approach to candidate gene identification. Proceedings of the National Academy of Sciences. 99(23):14903–14906. doi:10.1073/pnas.222549199.

Wayne ML, Telonis-Scott M, Bono LM, Harshman L, Kopp A, Nuzhdin SV, McIntyre LM. 2007. Simpler mode of inheritance of transcriptional variation in male Drosophila melanogaster. Proceedings of the National Academy of Sciences. 104(47):18577–18582. doi:10.1073/pnas.0705441104.

Williams-Simon PA, Oster C, Moaton JA, Ghidey R, Ng'oma E, Middleton KM, King EG. 2024. Naturally segregating genetic variants contribute to thermal tolerance in a Drosophila melanogaster model system. Genetics. 227(1):iyae040. doi:10.1093/genetics/iyae040.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. Nature. 430(6995):85–88. doi:10.1038/nature02698.

Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, et al. 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 48(5):481–487. doi:10.1038/ng.3538.