RESEARCH ARTICLE

Cosmopolitan inversions have a major impact on trait variation and the power of different GWAS approaches to identify associations

Benedict Adam Lenhart◯*, **Alan O. Bergland**◯*

Department of Biology, University of Virginia, Charlottesville, Virginia, United States of America

* bal7cg@virginia.edu (BAL); aob2x@virginia.edu (AOB)

Abstract

The ability of genomic inversions to reduce recombination and generate linkage can have a major impact on genetically based phenotypic variation in populations. However, the increase in linkage associated with inversions can create hurdles for identifying associations between loci linked to inversions and the traits they impact. Therefore, the role of inversions in mediating genetic variation of complex traits remains to be fully understood. This study uses the fruit fly *Drosophila melanogaster* to investigate the impact of inversions on trait variation. We tested the effects of common inversions among a diverse assemblage of traits including aspects of behavior, morphology, and physiology, and identified that the cosmopolitan inversions In(2L)t and In(3R)Mo are associated with many traits. We compared the ability of different approaches of accounting for relatedness and inversion presence during genome-wide association to identify signals of association with SNPs. We report that commonly used association methods are underpowered within inverted regions, while alternative approaches such as leave-one-chromosome-out improve the ability to identify associations. In all, our research enhances our understanding of inversions as components of trait variation and provides insight into approaches for identifying genomic regions driving these associations.

Author summary

Genomic inversions are large mutations that flip the orientation of sections of DNA, and the presence of inversions has the potential to impact many traits at once. Inversions exist in many organisms, including humans and the fruit fly *Drosophila melanogaster*. We take existing knowledge on *Drosophila* trait variation and identify several inversions which impact many kinds of traits. Approaches such as GWAS can be used to identify the DNA mutations most associated with variation in a trait of interest, but many GWAS methods do not perform well when inversions contribute to variation in phenotype. We show that a common GWAS

approach in *Drosophila* is not only unable to find association within inversions, but is overall underpowered. In contrast, we show a different approach is better able to identify mutations within inversions that are potentially associated with fruit fly traits. These findings help scientists studying a wide range of organisms to better understand the phenotypic impact of inversions, and support the broad research of identifying gene associations within these regions of inversion.

Introduction

Genomic inversions facilitate adaptation by suppressing recombination and generating linkage between many genes and mutations, therefore affecting the genetic basis and evolution of complex traits [1–5]. The adaptive importance of inversions for the evolution of novelty, local adaptation, and speciation is clear from a wide variety of organisms across the tree of life (reviewed in [6–10]). Despite the prevalent role of inversions in evolution, the role of inversions underlying phenotypic variation is often overlooked within association studies [11].

Our understanding of the general importance of inversions in affecting trait variation largely comes from ecological genetics, wherein distinct morphs have been identified in natural populations and subsequently linked to inversions [12]. For instance, conspicuous behavioral, morphological, phenological and life-history variation has been linked to complex inversion polymorphisms in wild populations of birds [13], seaweed flies [14], monkey-flowers [7], and snails [15]. In these cases, and many others [16–18], distinct morphs and their patterns of segregation were first characterized [19–22], prior to identification of inversion genotypes. Therefore, it is less clear if inversions have a major impact on less conspicuous quantitative genetic variation that can be identified through forward mapping approaches, where the goal is to identify the genetic basis of phenotypic variation.

There are two main reasons that forward mapping approaches have potentially missed the impact of inversions on quantitative trait variation. The first reflects the design features of mapping approaches that utilize recombinant populations [23–25]. Because inversions reduce recombination, these mapping panels have intentionally used strains with colinear genomes to facilitate recombination and enable efficient QTL mapping. The second reflects statistical techniques of genome-wide association (GWA) studies of outbred wild or laboratory populations. Modern GWA approaches that factor out population structure may have missed important links between inversions and trait variation because inversions can have a major impact on estimates of population structure and relatedness [26–28]. Thus, the use of population structure or relatedness estimates as a co-factor in GWA analysis may have led to reduced power to detect association with SNPs linked to inversions. Even when inversion-trait associations can be drawn following GWA [15,29–31], it is challenging to identify the specific genetic architecture within inversion driving association due to the high linkage between loci within inversions [32,33]. Therefore, the role of inversions in less conspicuous quantitative genetic variation may have been overlooked in many species.

The fruit fly *Drosophila melanogaster* is an excellent model to assess the importance of inversions on quantitative trait variation. *D. melanogaster* possesses large inversions present at intermediate frequencies worldwide, and flies show evidence of local and rapid adaptation driven by inversions (reviewed in [34]). *D. melanogaster* inversions are known to impact a variety of traits [35–41]. For instance, In(3R)P presence is associated with body size, lifespan, and starvation resistance [37,41], and In(2L)t is associated with behavioral, stress-tolerance, and morphological traits [42–45].

The Drosophila Genetic Reference Panel (DGRP) provides an excellent resource for identifying the effects of cosmopolitan inversions on quantitative variation, and for exploring the role of various GWA methods in discovering associations. The DGRP is a collection of 205 inbred and fully genotyped *D. melanogaster* lines, initially collected from a farmer's market in North Carolina [46]. The lineages have been inbred, their genomes have been sequenced, and the presence of common inversions has been characterized for each line using a combination of polytene chromosome preparations [26], principal component analysis [47], and PCR [48]. Due to the availability of these resources, DGRP lines have become a common model for phenotyping studies across many traits [49]. To facilitate association studies with the DGRP, several websites have been developed with a standardized mapping approach that factors out inversions and other drivers of relatedness [26,49]. Indeed, GWA approaches that correct for population structure, cryptic relatedness, or inversions account for approximately 60% of DGRP studies from a representative sample of 36 papers (curated dataset;(49)), yet few (35%) report testing for associations with inversions or inversion linked markers (S1 Table).

In our study, we test the ability of different GWA mapping approaches to identify signatures of association with inversions and linked variants. We utilize published studies that measure phenotypic variation in the DGRP. First, we show that several cosmopolitan inversions have large effects on dozens of traits, in that they explain more trait variation than expected by SNPs of comparable frequencies. Next, we explore four genome-wide association strategies that differ in their genetic-relatedness matrices (GRMs) and the treatment of inversions as co-factors, and contrast the real GWA signal for each phenotype to 100 permutations. We generated three types of GRMs (i) using the full genome, (ii) using an LD-thinned genome, (iii) and using a leave-one-chromosome-out (LOCO) approach. In addition, we performed association analysis and permutations using the full-genome based GRM and factored out the effect of inversions following methods outlined in (26,49). We show that the result of the GWA greatly depends on the mapping strategy, and that only the LOCO approach resolves association signals that exceed permutations. We highlight one case study to show that LOCO can identify signals of SNP association with alcohol tolerance within In(2L)t that are missed with the common GWAS method. Finally, we use the output of the LOCO-GWA to test whether SNPs identified as top candidates under the different mapping strategies show different levels of enrichment for signatures of local adaptation, and whether signals of pleiotropy are resolvable at specific loci inside two inversions.

Materials and methods

Selection of trait data

We re-analyzed trait data collected on the DGRP [46]. We made use of the DGRPool resource, which has consolidated the phenotypic line averages of inbred DGRP lines from many publications [49]. We used the "curated" data, and removed traits from this dataset that describe genomic features such as genome size or transposon presence, or used less than 75 unique DGRP lines, ending up with 409 unique traits derived from 36 publications (S1 Table). Some traits measure the same or similar phenotype, and thus there is potential for our results to be biased toward more frequently measured phenotypes. Of these 36 studies, 19 of them were also represented in the independent meta-analysis reported in Nunez et al., 2024 [43]. We annotated these traits by classifying each trait into 5 general groups: "Behavior", "Life-History", "Morphology", "Physiology", and "Stress-resistance" (S1 Table).

The phenotypic impact of cosmopolitan inversions

We characterized the effect of cosmopolitan inversions In(2L)t, In(2R)NS, In(3L)P, In(3R)K, In(3R)P, and In(3R)Mo on these traits, as these are the inversions considered by the DGRP analysis webtools. For this analysis, we used the inversion classifications provided by (26). The 205 DGRP lines classified here (S2 Table) are the same lineages present in the DGRP genotyping data. For each trait, we used a simple linear model to test the effect of any single inversion using strains that were homozygous for either the inverted or standard allele. We did not include heterozygotes as there is no way to determine the frequency of the inverted allele within DGRP lines marked as heterozygous, nor to know the genotype of any individual that was phenotyped. In addition, we do not consider the impact of multiple inversions simultaneously, nor do we test interactions between inversions because of the relatively low frequency of inversions in the dataset.

For each phenotype, and for each inversion, we compared the results of the linear model with a null model using an ANOVA test, and counted the number of traits with significant association with any of the inversions with $p < 0.05$ (S1 Table). Next, we evaluated whether the extent of association between traits and inversion status is greater than expected relative to other random polymorphisms in the genome. The motivation for this analysis is to test if the inversions have a greater impact than expected by chance. For each inversion, we replicated the above linear modeling using 100 SNPs and small indels randomly selected from those identified at the same frequency as each inversion (±1%), and on the same chromosome arm, but at least 2MB from the inversion breakpoints to avoid areas of highest linkage disequilibrium. By comparing the number of traits significantly associated ($p < 0.05$) with the inversions and random polymorphisms, we can use the 100 random polymorphisms to approximate the effect of a given mutation on trait data expected by chance. Last, we calculated R^2 (coefficient of determination) for the observed and matched-polymorphism models to ask whether the inversions explain more variation than expected by other comparable SNPs and small indels in the genome. If a trait was significantly associated with an inversion using the linear model results, and if the R^2 of that model surpassed the 95% quantile of permutations, we assigned that trait to a group of "inversion-associated traits" used in downstream analysis.

One last extension of this analysis was to examine the impact of the ancestry of the DGRP lines on trait line averages. This is important because North American *D. melanogaster* populations, including the DGRP, result from secondary contact and admixture between European and African populations about 150 years ago [50–53]. Using estimates of the proportion of European and African ancestry of each DGRP line from Pool 2015 (53), we created linear models using the same statistical tools as described above. For each phenotype, and each inversion, we evaluated an Ancestry-only model (ancestry as a fixed effect), Inversion-only model (inversion genotype as a fixed effect), and a Full model (both ancestry and inversion genotype as fixed effects) against a null model using ANOVA. Observed models were compared against permutations in which phenotype line averages were shuffled. Last, we repeated this statistical framework, this time comparing the Full model against the Ancestry and Inversion models.

Principal component analysis of trait data

We used principal component analysis (PCA) to identify broader trends of inversion impact on phenotype. From the original five inversions, we only investigated inversions in which at least 5% of the DGRP samples used were homozygous for the inversion. In this case we focus on two sets of inversion-associated traits: traits associated with In(2L)t or In(3R)Mo (S3 Table). Missing data for any trait were imputed using the *imputePCA* function from missMDA v1.19 [54] using the "regularization" method. This approach uses the average phenotype value for initial imputation, and performs a secondary regularization step. We used the imputed data to conduct PCA using FactoMineR v2.8 [55]. We separated the principal component projections by inversion presence using the inversion genotype of the DGRP lines and then compared the PCs of "inverted" and "standard" groups using Student's t-test.

Principal component analysis of genomic data

To understand how inversions impact general patterns of multi-locus genetic variation, we performed a series of PCA on the DGRP SNP and small indel polymorphism data. We used three polymorphism selection strategies for this principal component analysis that mirrors polymorphism selection strategies used for the construction of the GRM (see below). The first version of polymorphism selection ("Full") used all SNPs and small indels across the autosomes and X chromosome with minor allele frequency (MAF) greater than 5% and sites with missing genotype data in less than 20% of DGRP lines. The second version ("LD") used SNPs and small indels with MAF > 5% and missing rate < 15%, and a low pairwise linkage disequilibrium ($R^2 < 0.2$). To ensure that SNPs were at least 5000 base pairs apart we used the *snpgdsLDpruning* function of the R-package SNPrelate v3.17 [56], with the *slide.max.bp* parameter set to 5000. The third version used a leave-one-chromosome out ("LOCO") approach [57] that used the same filtering and thinning strategy as the LD-pruning approach but in four parts, each missing one of the main chromosomal arms in order to reduce the effect of an inversion on relatedness within its own chromosomal arm. We used the *snpgdsPCA* function from *SNPrelate* v3.17 [56] for PCA. To quantify the effect of inversions on principal component space, we constructed linear models in the same manner described above, recording the R^2 of both the linear models and a set of permutations with the lines' inversion genotype shuffled. We designated a model outcome significant if its R^2 surpassed 95% of permutations.

Construction of GRMs

We developed three genomic relatedness matrixes (GRM) to address population structure in different ways. For the "Full" method, we use the GRM matrix that is supplied by the DGRP website and is commonly used in DGRP GWAS studies (http://dgrp2.gnets.ncsu.edu/, last accessed 04/20/2025). This approach uses the VanRaden method [58] to construct a GRM from all SNPs and small indels with a MAF > 0.05 and a missing rate < 20% [26]. For the "LD" method, we used LD pruning using the same parameters that we used for the LD-pruned PCA, described above, and constructed a GRM from the whole genome using the *snpgdsGRM* function in SNPRelate based on the Genome-wide Complex Trait Analysis (GCTA) method [59]. For the "LOCO" method, we generated sub-GRMs, each one drawing from the DGRP genome but ignored one chromosome arm ("2L", "2R", "3L", "3R", and "X") and using the same steps as described for the LD-thinned approach.

GWA analysis

We performed association mapping using mixed-effect models implemented in the R package *GMMAT* v1.3.2 [60]. This approach used either the "Full", "LD", and "LOCO" GRMs as a random effect to control for population structure and cryptic relatedness. In addition, we performed a fourth association mapping approach based on the GWA approach developed by Huang *et al.* 2014 [26], which we refer to as the "Factored-out" approach. The Factored-out approach first standardizes each trait by the effects of the inversions by regressing line mean data against inversion status using the model:

$$\text{trait} \sim \text{In(2L)t} + \text{In(2R)NS} + \text{In(3R)P} + \text{In(3R)K} + \text{In(3R)Mo}$$

Next, the residuals of this model are used as the trait or association analysis. We used the "Full" GRM with the Factored-out approach to replicate the association model implemented in by Huang et al. 2014 (26) and available on the DGRP online GWA tool (http://dgrp2.gnets.ncsu.edu/).

For each of the four GWA approaches, we compared a "full model" to a "reduced model." The reduced model is described by the formula:

$$y \sim \text{Wolbachia} + \text{GRM},$$

where y represents the line means for a particular trait (or residuals in the case of the Factored-out model). Wolbachia is an infectious symbiote known to affect aspects of Drosophila fitness [61–63], and is included as a cofactor in standard DGRP GWA approaches. Here we encode Wolbachia infection status as a fixed effect listed as present or absent, based on the tables published in Huang et al., 2014 [26]. GRM is a random effect genetic relatedness matrix. The full model is:

$$y \sim variant_i + Wolbachia + GRM,$$

where $variant_i$ is the fixed effect and an additive representation of the dosage of the ith SNP or small-indel reported for the DGRP. We contrasted the full and reduced models using the *glmm.score* function in the GMMAT package (v1.4.2), which filters out all variants with minor allele frequency < 5% and missing data > 15%. In the LOCO approach (57), we split the scoring of the genome into five sections for each of the major chromosomal arms, with each GWA using the sub-GRM constructed without that corresponding region. Our GWA approach scores inverted and non-inverted regions without distinction.

For each trait and GWA method, we conducted 100 permutations by randomly shuffling the trait data prior to fitting the reduced and full models.

GWA summary statistics

We compared the overall genomic signal from the GWA of each trait using statistics for the observed and permuted GWA models. We partitioned each chromosome into bins based on whether SNPs are inside or outside inversions as defined using coordinates in Corbett-Detig et al. 2012 [48]. For each bin, we calculated the proportion of SNPs with a p-value less than 10^{-5} as "hits", a common p-value threshold in DGRP studies [64–70]. We also calculated the genome inflation factor (GIF) as the ratio of the median observed variant-wise X^2 values within the bin divided by the expected median X^2 with 1 degree of freedom. We compared the summary statistics for the observed data to the statistics from the permutations for each trait and GRM method, reporting the proportion of traits where the statistic exceeds the 95th percentile of the trait's permutation-based distribution.

Enrichment tests

We tested if GWA hits identified via different GRM approaches prioritize SNPs that are potentially subject to temporally or spatially variable selection as identified by BayPass v2.1 [71]. We used data from the DEST dataset to obtain allele frequencies from *D. melanogaster* population*s* sampled across the North American East Coast [72], and Charlottesville, Virginia across multiple years [43]. We identified polymorphisms that are more differentiated than expected given population structure (XtX* outliers) and tested association between variants and environmental variables after correcting for population structure. Using this framework, we identified differences in allele frequencies between populations (XtX*) and Bayes Factor (BF) of association with environmental variables. We used latitude for the association of East Coast variants and maximum temperature two weeks prior to collection for Charlottesville variants [43]. We ran the software five times and report the mean statistic per SNP [73]. We generated a null distribution of XtX* and BF using the POD (Pseudo-observed Data) framework for 10-times the number of SNPs as the observed data, ran BayPass with five replicate iterations, and calculated empirical p-values for XtX* and BF using these POD simulations.

We identified the level of enrichment between top GWA variants and top BayPass variants. We identified the top hits within each GWA study by identifying the 500 hits with the lowest P-value, and top XtX* and BF variants as those that surpass 95% of the corresponding distribution from the simulated POD data. We computed Fisher's Exact test by contrasting the odds that top association hits for any trait are enriched for top XtX* and BF hits. We compared these Fisher's Exact test odds-ratios to odds-ratios constructed in the same way using the permuted GWA.

COLOC analysis

We tested if inverted regions of the genome are likely to have pleiotropic effects on phenotypic variation using the *coloc.abf* function from coloc v5.2.3 [74]. By treating the top principal component projections as dimensionality-reduced traits, we sought to identify regions in the genome with a shared association with multiple inversion-linked traits. We used the Factored-out and LOCO GWA frameworks to score the impact of SNPs genome-wide on the PC1 and PC2 loadings for the In(2L)t and In(3R)Mo associated traits. We identified areas of colocalized signal on PC1 and PC2 using a sliding window analysis across the genome with window size 10Kb and step size 5Kb, and compared the SNP GWA data using *coloc.abf*(). This analysis identified regions of SNPs likely associated with the traits differentiated along only PC1, only PC2, or regions of SNPs with a colocalized association across PC1 and PC2.

Results

Cosmopolitan inversions impact phenotypic variation

To study the role of inversions on genetically based trait variation in *D. melanogaster*, we re-analyzed data from publications that measured trait variation in the DGRP and were curated in the DGRPool database [49]. We analyzed 409 traits, categorizing them into five groups: morphological, life-history, stress resistance, physiological, and behavioral (S1 Table). We found that In(2L)t, In(3R)Mo, and In(3R)K are associated with more traits than expected given SNPs of the same frequency (Fig 1A). In(2L)t is especially associated with many behavioral traits including startle response, sleep, and movement, while In(3R)K and In(3R)Mo are associated with morphological traits such as femur and abdomen size (S1 Table). We found that the inversions explain ~10% of the variation in these traits, and that dozens of traits are explained better by inversion status than expected from random polymorphisms in the genome (Fig 1B). Signals of association with the inversions are not explainable by variation in African ancestry among DGRP lines (S1 Fig).
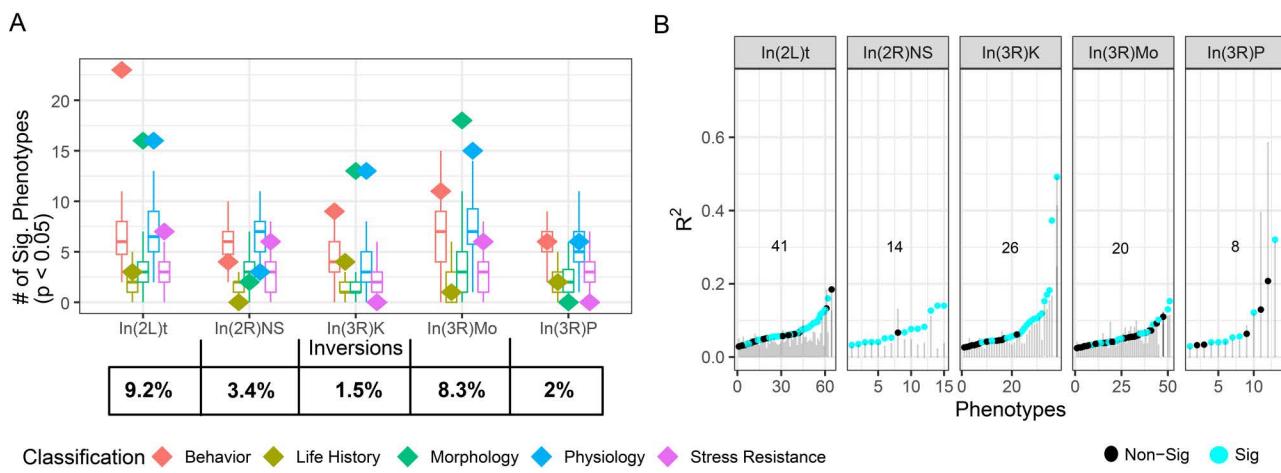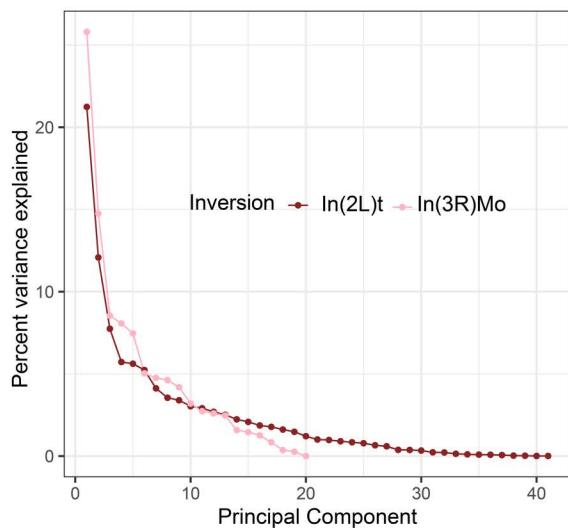

Fig 1. Cosmopolitan inversions have a major impact on trait variation. A) Diamonds, colored by trait category, indicate the number of traits significantly affected by inversion presence at $p < 0.05$, overlaying the same-frequency models shown with box plots. The percentage of lines homozygous for each inversion within the lineages tested is given at the bottom. **B)** The proportion of variation explained by each inversion (R^2) for traits significantly associated at $p < 0.05$ with each inversion, compared against the distribution of corresponding same-frequency models in grey. Statistically significant inversion model values that surpass the null distribution are colored cyan. The number in each panel of B is the count of phenotypes where R^2 exceeds permutations and traits that are significantly associated with the inversion with $p < .05$.

https://doi.org/10.1371/journal.pgen.1012012.g001
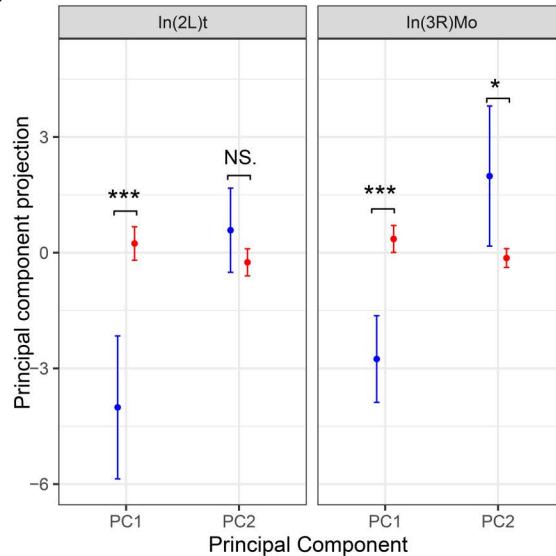
Principal component analysis of traits associated with In[2L]t and In[3R]Mo

To better understand the impact of inversions on phenotypic variation, we performed PCA on the traits that are associated with In(2L)t and In(3R)Mo and also explain more variation than expected by chance (In(2L)t: n = 41, In(3R)Mo: n = 20; **Fig 1**). The top two principal components (PC1 and PC2) explain over one third of the trait variation for both In(2L)t and In(3R)Mo (**Fig 2A**). Therefore, we restrict further analysis to these two principal components. In(2L)t significantly loads onto PC1 (t-test, t = -4.38, df = 16.68, p = 4.29e-4) of its associated trait set, while In(3R)Mo significantly loads
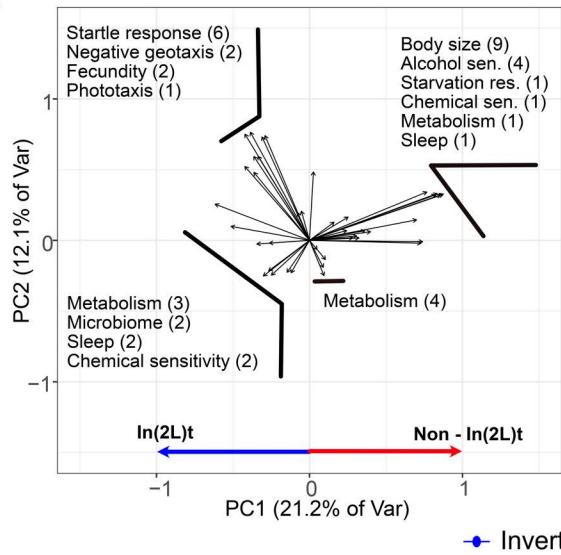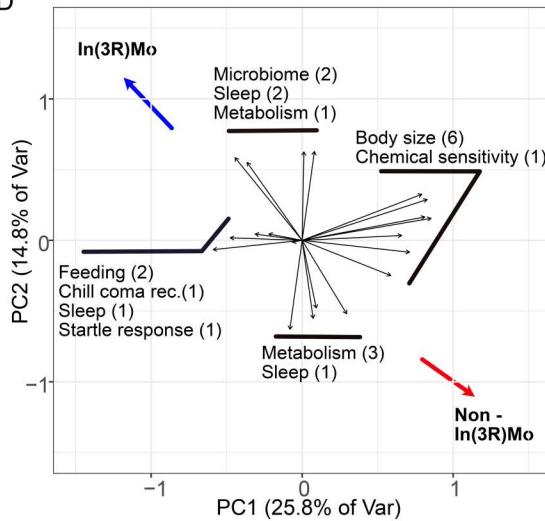

Fig 2. Principal component analysis of In(2L)t and In(3R)Mo. A) A scree plot showing the variance explained by principal components, colored by their associated inversion. **B)** The effect of In(2L)t and In(3R)Mo genotype on PC1 and PC2 projection values. Points represent the mean and confidence intervals represent two standard errors. **C)** PC loading values for the traits significantly impacted by In(2L)t. Labels are aggregated to show similar traits together, e.g., "Sleep (2)" corresponds to two sleep traits. Variance explained by each PC is given on the axis title. **D)** PC loading values for the traits significantly impacted by In(3R)Mo. In B, C, and D colors represent the homozygous genotype.

https://doi.org/10.1371/journal.pgen.1012012.g002

onto both PC1 (t-test, $t = -5.18$, df = 18.03, $p = 6.21e-5$) and PC2 (t-test, $t = 2.27$, df = 15.54, $p = 0.038$; **Fig 2B**) of its associated trait set. For the In(2L)t PCA, traits like body size and ethanol sensitivity have positive loadings on PC1 and traits like startle response and negative geotaxis have negative loading. Lines homozygous for In(2L)t have lower values of PC1, thus have higher startle response and higher activity levels (**Fig 2C**), among other differences (S3 Table). For the In(3R)Mo PCA, traits like body size have positive loadings on PC1 and traits like feeding and chill coma recovery have negative loading. Lines homozygous for In(3R)Mo have lower values of PC1, thus have lower body size (**Fig 2D**), among other differences (S3 Table).

Controlling the effect of inversions on PC and GRM space

The cosmopolitan inversions of *D. melanogaster* have been previously shown to have an impact on genome-wide patterns of genetic variation as summarized by principal components and genetic relatedness matrices [27,28,75]. Therefore, we tested if different polymorphism selection strategies can mitigate this impact. As previously reported [26], PCA of the "Full" genome shows that inversions strongly impact PC space. In(2L)t primarily impacts $PC1_{Full}$ ($F_{1,179} = 870$, $p = 2.89e-70$) and In(3R)Mo primarily impacts $PC2_{Full}$ ($F_{1,195} = 331$, $p = 5.54e-44$, **Fig 3A**). The "LD" polymorphism-set slightly reduces the impact on $PC1_{LD}$ for In(2L)t ($F_{1,179} = 175$, $p = 2.89e-28$) and the impact on $PC2_{LD}$ for In(3R)Mo ($F_{1,195} = 90.71$, $p = 6.93e-18$, **Fig 3A**). In contrast, PCA of the LOCO genome shows a sharply reduced impact of In(2L)t on $PC1_{LOCO}$ ($F_{1,179} = 5.84$, $p = 0.017$) and In(3R)Mo on $PC2_{LOCO}$ ($F_{1,195} = 0.02$, $p = 0.88$, **Fig 3A**).

We calculated the proportion of variation in the genetic PC1 and PC2 that is explained by inversion status, and contrasted that to a null distribution made via 100 permutations. We found for the Full and LD methods, In(2L)t and In(3R)Mo explained more variation for PC1 and PC2 than expected by chance (**Fig 3B**), with In(2L)t explaining the most variance within PC1 using the Full method and less using the LD method, while In(3R)Mo explained the most variance for PC2 within the Full method and less within the LD method. Meanwhile, within the LOCO method each inversion explains a near zero amount of variance for PC1 or PC2, and PC2 is no longer significantly impacted by either inversion (**Fig 3B**). For PC3 and PC4, the R^2 of each of the principal component ~ inversion genotype models were near zero, indicating there is little correlation between inversion genotype and these principal components (S2 Fig).

To identify how the presence of inversions impacts patterns of relatedness, we compared the relatedness of the DGRP lines using different polymorphism selection strategies. Across all approaches, relatedness is low within standard genotype lines with no cosmopolitan inversions (**Fig 3C**). This replicates observations in Huang et al. 2014 (26). However, when using the Full and LD-thinned approaches, relatedness is noticeable between lines that are both homozygous for any given inversion. In contrast, the LOCO approach on a given chromosome can drive relatedness for homozygous inverted lines near zero while still accounting for inversions on the other chromosomal arms (**Fig 3C**).

The LOCO approach can better capture signals for inversion-associated traits

After characterizing the impact of different GRM methods and presence of inversion in the DGRP data, we tested the strengths and weaknesses of the four GWA strategies (Factored-out, Full, LD, LOCO) on the magnitude of association signal. We compared the summary statistics from the observed trait GWA against permutations to see how many traits identified more signal than expected by chance. Using the Factored-out approach, we found that about 6% of traits have "hit-counts" that exceeds the largest 95% of the null distribution generated by permutation (**Fig 4**). In other words, the Factored-out approach largely fails to identify more associations than would be expected by random chance, as 6% is about the number of traits that would surpass permutations as false positives. We found that the Full and LD-thinned approaches also fail to identify many more significant associations than expected given the permutations. The LOCO method surpasses permutation significantly more often than the Factored-out method within inverted regions (Fisher's Exact Test - FET, 2L: $p = 9.47e-9$, 2R: $p = 5.17e-4$, 3L: $p = 6.66e-10$, 3R: $p = 1.61e-7$), as well as outside the inverted region (FET, 2L: $p = 1.2e-9$, 2R: $p = 7.71e-4$, 3L: $p = 1.61e-5$, 3R: $p = 9.19e-13$) (**Fig 4A**). Similarly, the GIF of LOCO surpasses
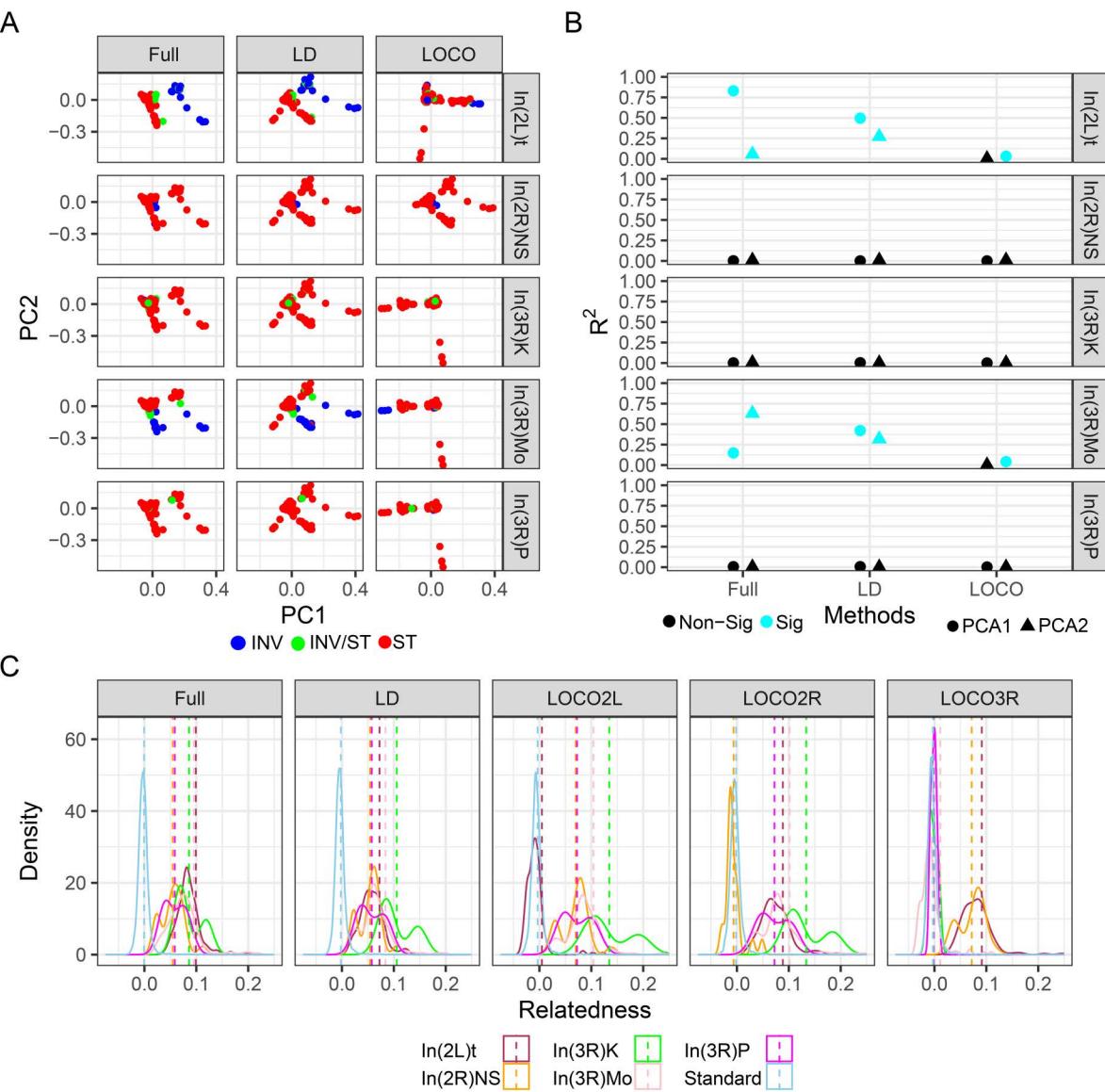
Fig 3. Inversions influence population genetic structure and relatedness. A) The first and second genomic PCs for each sample colored by the genotype of that sample. **B)** The R^2 values for models comparing PC1 and PC2 to inversion, colored by which values exceed a distribution of permutations. **C)** The distribution of pairwise relatedness values between each of the samples, indicated by a solid line colored by genotype of the samples and split across different methods, with a dotted line indicating means.

permutation significantly more often than the Factored-out method within inverted regions (Fisher's Exact Test - FET, 2L: p = 7.05e-45, 2R: p = 3.14e-12, 3L: p = 7.42e-6, 3R: p = 2.21e-29), as well as outside the inverted region (FET, 2L: p = 1.59e-19, 2R: p = 1.38e-13, 3L: p = 2.35e-5, 3R: p = 1.15e-21: (**Fig 4A**). However, this increase in GIF via LOCO is not uniform across the chromosome. The GIF of traits scored with LOCO is significantly higher when scored in inverted regions than non-inverted regions on both 2L (FET, p = 4.34e-8) and on 3R (FET, p = 0.014).

As a case study to highlight how statistical methods impact signals of association in GWA studies, we focused on male alcohol tolerance as measured by Morozova et al., 2015 [76]. This phenotype is correlated with In(2L)t (**Fig 4B** and S1 Table) and is among the phenotypes that our analyses focused on. Performing GWA on male ethanol tolerance using
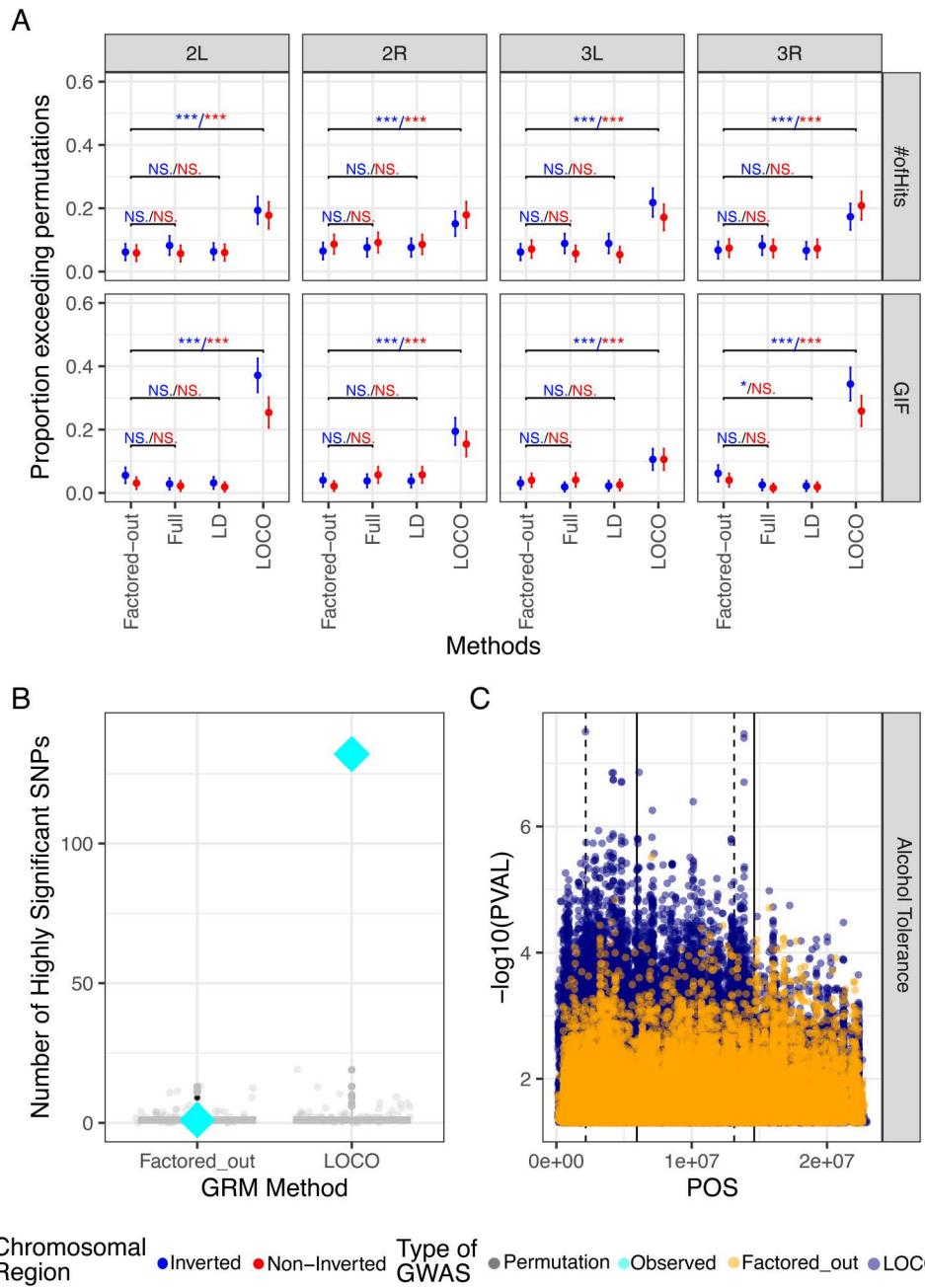
Fig 4. The LOCO approach can better capture signal genome-wide. A) The proportion of significant hits and GIF for each GWAS output are compared across each method and colored by location relative to inversions. The proportion of traits that exceed their corresponding permutations is given on the y axis, along with binomial confidence intervals. The color of the significance annotation refers to the chromosomal region under comparison. (* = p < 0.05, ** = p < 0.01, *** = p < 0.001, NS. = p >= 0.05) **B)** The number of significant hits (p < 1e-5) on 2L for male ethanol tolerance for LOCO and Factored-out GWA is compared against permutations. Color refers to observed or permuted GWAS results. **C)** The strength of association of variants with male ethanol tolerance is shown across 2L. The left-most solid line indicated the *Gdph* gene, and the right-most indicates the *Adh* gene. The dotted lines indicated the breakpoints of In(2L)t. Color refers to the GWA approach used.

the factored-out and LOCO approaches yields starkly different results. The Manhattan plot of the GWA results for the LOCO method shows a strong signal of association with SNPs near In(2L)t, a signal that is absent using the Factored-out approach. Indeed, when comparing the GWA signal from the real data to permutations, we observe that the LOCO method yields far stronger signals of association than permutations, while the factored-out method does not exceed permutations. Intriguingly, the strongest signal of association for male ethanol tolerance is found at a SNP near the distal breakpoint of In(2L)t, whereas signals of association at the closely linked alcohol dehydrogenase (*Adh*) gene is not apparent (**Fig 4C**).

Enrichment tests

To compare the utility of the four approaches to identify biologically meaningful loci, we characterized the ability of LOCO and Factored-out methods to identify loci thought to be important for local adaptation. Using estimates of allele frequencies of *D. melanogaster* collected across seasons and across latitudes [72, 77], we used the Baypass [71] software to identified SNPs that are more strongly differentiated across the North American east coast, or within Charlottesville, VA through time (XtX* outliers). In addition, we identified the strength of association between SNPs and latitude for the East Coast samples and between SNPs and temperature in the two weeks prior to sampling for the Charlottesville samples. To understand which methods could successfully identify enrichment within inverted regions, we compared enrichment signals between inversion associated and not-associated traits. There was a significant jump in enrichment between GWA hits and candidate SNPs that have large BF association with maximum temperature for the LOCO method on 2L (FET, p = 0.019), but not for hits derived from the Factored-out method (**Fig 5**). Correspondingly, there was an increase
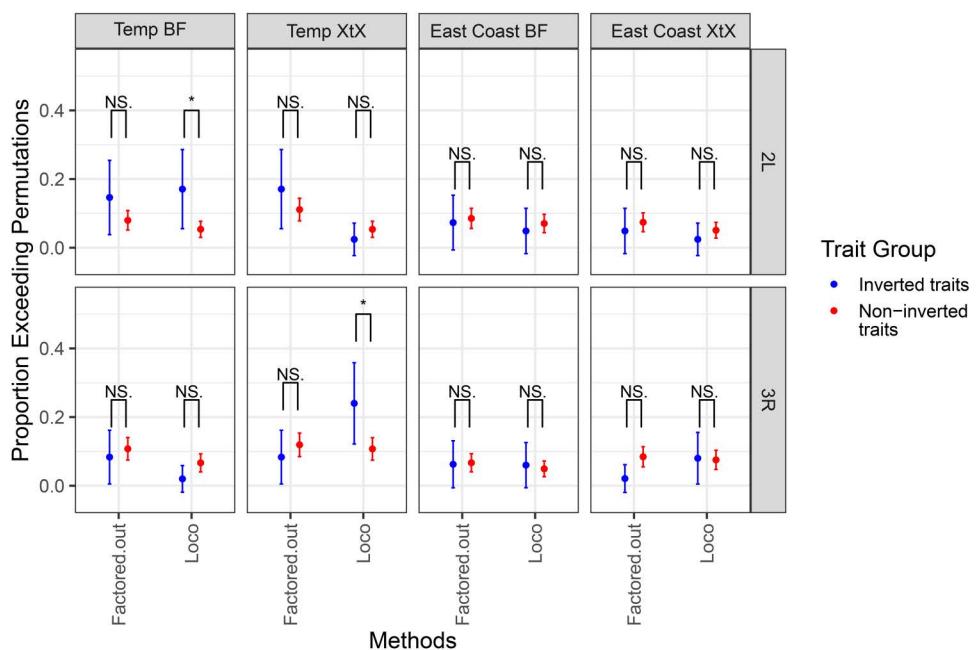
Fig 5. Enrichment of GWAS hits with environmentally associated SNPs. We used two population sets: a temporal set collected in Charlottesville, and a latitudinal set collected across the North American east coast. BF (Bayes Factor) outliers are top loci associated with temperature (Charlottesville temporal population set) and latitude (East Coast population set). XtX outliers are the most strongly differentiated SNPs through time (Charlottesville) or space (East Coast). We report the proportion of traits for whom the enrichment between BayPass SNPs and real GWA SNPs exceeds the 95% largest enrichment with the GWA permutations. Error bars represent 95% binomial confidence intervals. Color indicates whether the traits are associated with inversion. (* = p < 0.05, ** = p < 0.01, *** = p < 0.001, NS. = p >= 0.05).

https://doi.org/10.1371/journal.pgen.1012012.g005

in enrichment between GWAS hits and differentiation across max temperatures for the LOCO methods on 3R (FET, p = 0.011) but not for the Factored-out method (**Fig 5**). There was no difference reported within the East Coast enrichments with GWA hits between inverted and non-inverted associated traits.

COLOC enrichment within the genome

To compare the ability of GWA approaches to identify potentially pleiotropic associations of inversions with orthogonal multivariate-traits, we calculated the probability that regions of the genome share polymorphisms that affect multi-dimensional traits (co-localization). Using the top principal component projections from **Fig 2** as dimensionality reduced traits, we scored the effect of SNPs and small indels genome-wide on PC1 and PC2 using the LOCO and Factored-out GWA approaches. We followed with a sliding window analysis to identify loci within the genome are that likely associated with traits differentiated along only PC1, the traits differentiated along only PC2, or for both PC1 and PC2. With the LOCO method, we identified variants within the In(2L)t inverted regions and near the breakpoints that have high association likelihood with PC1 of the In(2L)t-linked traits (**Fig 6A**) while there was little association on other chromosome arms (**S3 Fig**). In contrast, for In(3R)Mo areas of likely association were identified across 3R for both PC1 and PC2, with the peaks aligning with other inversion breakpoints on 3R (**Fig 6B**) and several peaks observed on other chromosomes (**S4 Fig**). Notably, the peaks of the highest likelihood of association differed between PC1 and PC2, suggesting that distinct loci within the
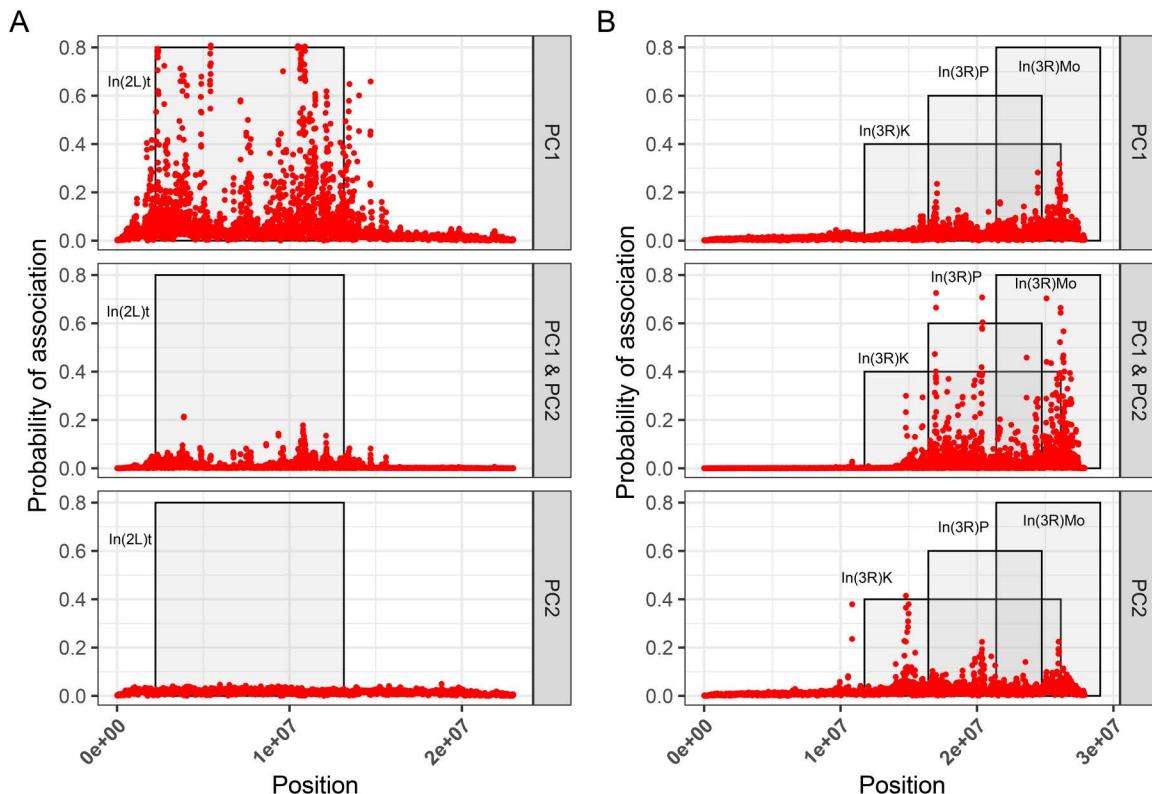
Fig 6. The LOCO method reveals areas of likely colocalized association for inversion-related traits. A) Results of a sliding window analysis examining enrichment between SNPs scored using LOCO for PC1 and PC2 of In(2L)t. The y-axis shows the likelihood of association, and the x-axis shows position on the genome. The grey shaded regions show the zone of cosmopolitan inversions on the chromosome arm. **B)** Same analysis as in A, but for traits associated with In(3R)Mo. D likelihood.

https://doi.org/10.1371/journal.pgen.1012012.g006

inverted regions influenced different sets of traits. In contrast, SNPs scored using the Factored-out method failed to capture any signal of likely association with either PC1, PC2, or both (S5 Fig).

Discussion

Genomic inversions can simultaneously influence multiple traits and provide a mechanism for adaptation. Associations between inversions and phenotypic variation have been identified across the tree of life, along with evidence that natural selection acts upon these genomic features [6,7,8]. Here, we find that inversions within *D. melanogaster* impact a suite of diverse traits (Figs 1A, 2C, and 2D), and specific statistical methods are better equipped to map associations between inversion-linked loci and these traits. Inversions should be considered areas of interest, rather than areas to be skipped over within association studies, as their presence here is shown to explain large parts phenotypic variation (Figs 1B and 2B), and strongly impacts estimates of relatedness derived from polymorphism data (Fig 3). We illustrate that different GWA approaches have different power in the number and strength of associations they can identify (Fig 4). Compared to several commonly used methods, the LOCO approach is better able to identify variants linked to inversions that are associated with key traits. We show that these variants are enriched for loci that could underlie local adaptation and that these variants are likely pleiotropic (Figs 5, 6A, and 6B).

Previous work has linked *D. melanogaster* inversions to phenotypic variation, implicating these mutations in changes to body size, wing size, longevity, and more [35–41]. Despite these known impacts, only 13 out of the 36 publications aggregated here report any test for association between inversions and their trait(s) of study (S1 Table). Here we reexamined the impact of inversions on a large body of diverse trait data, and show that inversions like In(3R)Mo and In(2L)t significantly affect more traits than would be expected by chance, including inversion-trait associations not previously identified (S1 Table and Fig 1). In(3R)Mo varies across latitudinal clines in multiple continents [3], and the overlapping inversion In(3R)P is thought to facilitate local adaptation [34,36,78,79]. In our analysis, In(3R)Mo is likely more enriched than In(3R)P due to its higher frequency. Here we confirm that inversions on chromosome 3R impact body size (Fig 2D). Similarly, highlighting the association between In(2L)t and activity (Figs 1A and 2C) provides new avenues for investigation for the ongoing link to this inversion and seasonal adaptation [43,80]. Taken together, we show that inversion presence explains considerable trait variance for specific traits (Figs 1B and 2B), indicating these inversions should be a major factor in consideration for association studies.

Inversions provide a challenge for association studies, as the increased linkage disequilibrium and relatedness within inverted samples can elevate the false discovery rate [28,75]. Many modern GWA techniques thus seek to mitigate the impact of relatedness by using top principal components as cofactors [81] or factoring our relatedness using GRMs as a random-effect [82]. The Factored-out approach described here employs such methods, using genome-wide GRMs and additionally factors out the effect of inversions prior to genome-wide association mapping. Of the studies we analyzed, 21/36 used this method or an equivalent for GWA with the DGRP (S1 Table). However, we report that only about 6% of GWA using this method find more hits than from random permutations (Fig 4), indicating a lack of power and a potentially high false-positive rate amongst many published DGRP studies. Other GWA methods, such as thinning the relatedness matrix for linkage disequilibrium, fare little better (Fig 4). In contrast, LOCO is designed to identify association when there is high LD within the genome, by avoiding proximal contamination between highly linked SNPs while still partially accounting for population structure [57]. Recent association studies have used LOCO methods of establishing relatedness while investigating association studies within inversions and other areas of high LD [15,83]. For example, Calboli *et al.,* 2022 established an association between an agriculturally relevant trout disease and an inversion used a LOCO method, but not with their accompanying "Full" genome method [84]. Here we provide new evidence that LOCO has the potential to outperform other methods at identifying association signal, especially when inversions are present in mapping populations (Fig 4).

To highlight the differences in GWA signal between the Factored-out and LOCO approach we focused on alcohol tolerance. Alcohol tolerance is an ecologically relevant trait for *D. melanogaster*, given their preferred habitat of rotting and

fermenting fruit [85, 86]. While the yeast that colonizes rotting fruit provides an important food source for larval [87, 88] and adult flies [89], the ethanol produced as a byproduct of fermentation can act as a strong selective pressure [90, 91]. A principal *Drosophila* enzyme involved in metabolizing ethanol is *Adh*, and genetic variation in the *Adh* gene and protein has been a focus of empirical population genetics since the inception of the field [92–94]. Notably, the *Adh* gene harbors two allozyme variants in *D. melanogaster* (Fast and Slow; [95]), generated by a single non-synonymous polymorphism (K192T; [92]), and these variants display clinal patterns of allele frequency change consistent with the action of spatially [96, 97] and temporally varying selection [98]. As a consequence, the role of the *Adh* polymorphism in explaining heritable genetic variation in ethanol tolerance has been a long-standing area of research [99]. Recent work using genetically engineered variants clearly shows that enzymatic activity of these alleles contributes to genetic differences in adult ethanol tolerance [100]. However, a DGRP GWA study of adult ethanol tolerance using the factored-out approach [76] did not recover any signal of association at *Adh,* or other genes with natural enzyme polymorphism linked to ethanol tolerance such as *Gdph* and *Mdh1* [101–103]. Morozova *et al.* attribute the lack of signal at these and other classical ethanol tolerance genes to limited statistical power or physiological buffering.

The *Adh* gene resides on chromosome 2L and is several megabases outside of the boundaries of In(2L)t. Yet, polymorphism at *Adh* is tightly linked to In(2L)t [104, 105]. This linkage could result from recombination interference that is generated by inversion heterozygotes [106]. In addition, it is possible that epistatic selection is maintaining linkage disequilibrium between *Adh* and In(2L)t [107–109]. Epistasis is plausible given that *Gdph* resides within the boundary of In(2L)t and is tightly linked to the inversion and *Adh* [44, 110]. There is some empirical support for epistasis from selection experiments [101]and wild population surveys [110], but this hypothesis lacks support from transgenic approaches [103]. Nonetheless, linkage disequilibrium between *Adh* and In(2L)t, coupled with possible epistatic interactions between these loci, could impact signals of association for ethanol tolerance.

In our analysis, analysis choices for GWA has a strong impact on signals of association with male ethanol tolerance (**Fig 4B** and **4C**). Using the factored-out approach, we do not recover any signal of association at *Adh* (**Fig 4C**) similar to the original study [76]. In addition, using the factored-out approach, there is no signal of association greater than we expect by chance (**Fig 4B**). However, the LOCO approach yields much more interesting signals of association. Notably, SNPs tightly linked to In(2L)t are strongly associated with male ethanol tolerance, consistent with our ANOVA results (**Fig 1**). The strongest signal of association in the genome is a SNP that resides immediately adjacent to the inversion breakpoint, and another noticeable peak of association is immediately adjacent to the *Gdph* locus (**Fig 4C**). In the Factored-out approach, there is not a strong signal generated by *Adh*, *Gdph*, or In(2L)t. While the causal roles of In(2L)t, *Gdph*, and *Adh* on male ethanol tolerance in the DGRP remain unclear, a major candidate locus is missed when performing GWA using the Factored-out approach. GWA approaches that avoid proximal contamination may therefore be important in GWA, especially when inversions or other large structural variants are present.

Incorporating data on clinal allele variation indicates that the LOCO method may provide advantages in identifying loci involved in inversion-mediated adaptation. Several studies have indicated that In(2L)t could mediate seasonal adaption [80,111], and one study suggested that behavioral traits could contribute to rapid seasonal adaptation [79]. We investigated the enrichment between the top GWA hits and an independent set of environmentally relevant alleles. We discovered that top GWA hits within In(2L)t are enriched for seasonally associated loci, and that GWA hits within In(3R)Mo are enriched with loci that differentiate over latitudinal clines (**Fig 5**). The overlap of environmentally varying alleles and top GWA loci can be identified from LOCO-based GWA, but not from the Factored-out approach (**Fig 5**).

We identify differences in the ability of GWAS approaches to identify regions within the genome that are associated with multiple traits (pleiotropy). The ability of inversions to pleiotropically effect multiple traits has already been noted in salmonids [84, 112], and mice [6,113–115]. In principal component analysis, different aspects of body size can load onto the top principal component, reflecting some unifying aspect of body size development [116]. In contrast, phenotypic variation across principal components can reflect a different degree of pleiotropy across orthogonal traits. Thus, we characterized

areas of high association between PC1 and PC2 of the inversion linked trait sets to illustrate such pleiotropy using a colocalization test. Only the LOCO approach identifies that the areas of highest association with multiple inversion-linked traits are near their corresponding inversions, as one might expect (**Fig 6A** and **6B**). However, within the inverted regions there are peaks of higher likelihood of association, similar to the finding in Nunez et al., 2024 of peaks of SNP-phenotype enrichment within In(2L)t [43]. Peaks of association with PC1 of In(3R)Mo may indicate loci relevant to traits such as body size, while a peak for PC2 may indicate different loci relative to traits such as metabolic storage and sleep, and the peaks for both PCs indicate areas of likely pleiotropic effect (**Fig 2D**). While LOCO can aid in identifying these areas of likely association, this signal cannot be recapitulated using the Factored-out approach (S5 Fig).

Inversions have the potential to be a fruitful area of investigation within association studies. Despite the evidence across taxa that inversions can influence many classes of traits (6,7,78), inversions are sometimes presented as a statistical hindrance [28,75]. Efforts such as the creation of popular mapping populations from largely co-linear genotypes [23–25], and the use of multiple methods to factor out inversion presence within the DGRP [26] represent steps to account for these mutations. To be clear, the phenotyping and association studies from the DGRP and other mapping populations have produced many important and foundational insights. However, models that remove or ignore inversions miss a valuable opportunity. Methods like LOCO offer tools toward building association studies to identify relevant loci linked to inverted regions (84,85). Inversions can play a significant role in the traits of humans and across many forms of life [117–120]. Improving our ability to connect inversion to traits will motivate future work to better understand how these complex mutations contribute to trait regulation and formation.

Supporting information

S1 Fig. The addition of ancestry proportion does not remove the broad impact of inversion genotype on phenotype. A) The number of phenotypes with significant associations is shown as diamonds for the Ancestry and Inversion model, as well as for a Full model that uses both ancestry and inversion genotype as fixed effect. A set of paired 100 permutations of each model is shown as a box and whisker plot. Results are split across five cosmopolitan inversions, and colored by trait classification. **B)** The same plot as in (A), now showing a comparison between the Full and Ancestry models, as well as the Full and Inversion models.
(DOCX)

S2 Fig. Genomic principal components PC3 and PC4 have little correlation with inversion genotype. A) The third and fourth genomic PCs for each sample colored by the genotype of that sample. **B)** The R^2 values for models comparing PC3 and PC4 to inversion, colored by which values exceed a distribution of permutations.
(DOCX)

S3 Fig. Signal of loci association with In(2L)t is mostly adjacent to the inversion. The same results of the association study using the LOCO method from Fig 6 are shown across the genome, showing the likelihood of a SNP's association with PC1, PC2, or both from the In(2L)t PCA analysis.
(DOCX)

S4 Fig. Signals of loci association with In(3R)Mo is elevated on 3R. The same results of the association study using the LOCO method from Fig 6 are shown across the genome, showing the likelihood of a SNP's association with PC1, PC2, or both from the In(3R)Mo PCA analysis.
(DOCX)

S5 Fig. The Factored-out method fails to identify areas of likely association. A) Results of a sliding window analysis examining enrichment between SNPs on 2L scored using Factored-out for PC1 and PC2 of In(2L)t, the y- axis shows the strength of enrichment and the x-axis shows position on the genome. Grey shaded region show the zone of

cosmopolitan inversions on the chromosome arm. **B)** Same analysis as in A, but considering chromosome arm 3R and inversion In(3R)Mo.
(DOCX)

S1 Table. Phenotype metadata. This table includes the background information for the DGRP studies and individual phenotypes that passed quality control and are used throughout this publication. For each phenotype the name and doi of the originating study is listed. Additionally, the table includes metadata relating to the number of DGRP lines with data for each trait, and the relationship of each trait to cosmopolitan inversions.
(XLSX)

S2 Table. The inversion genotype of DGRP lines. This table indicates the genotype for the five cosmopolitan inversions examined in this study, for each of the 205 lineages genotyped within the DGRP. This data is directly taken from the DGRP website (http://dgrp2.gnets.ncsu.edu/, last accessed 04/20/2025).
(XLSX)

S3 Table. The principal components of inversion-related traits. This table further describes the sets of traits found to be significantly associated with In(2L)t or In(3R)Mo, showing the top 5 principal components for the PCA run on each of the two sets of traits.
(XLSX)

Acknowledgments

Author contributions

Conceptualization: Benedict Adam Lenhart, Alan O. Bergland.

Data curation: Benedict Adam Lenhart.

Formal analysis: Benedict Adam Lenhart, Alan O. Bergland.

Funding acquisition: Alan O. Bergland.

Investigation: Benedict Adam Lenhart.

Methodology: Benedict Adam Lenhart, Alan O. Bergland.

Project administration: Alan O. Bergland.

Resources: Alan O. Bergland.

Software: Benedict Adam Lenhart, Alan O. Bergland.

Supervision: Alan O. Bergland.

Validation: Benedict Adam Lenhart.

Visualization: Benedict Adam Lenhart.

Writing – original draft: Benedict Adam Lenhart, Alan O. Bergland.

Writing – review & editing: Benedict Adam Lenhart, Alan O. Bergland.

References

1. Charlesworth B, Charlesworth D. Selection of new inversions in multi-locus genetic systems. Genet Res. 1973;21(2):167–83. https://doi.org/10.1017/s0016672300013343

2. Dobzhansky T. Genetic nature of species differences. The American Naturalist. 1937;71(735):404–20.

3. Kapun M, Fabian DK, Goudet J, Flatt T. Genomic Evidence for Adaptive Inversion Clines in Drosophila melanogaster. Mol Biol Evol. 2016;33(5):1317–36. https://doi.org/10.1093/molbev/msw016 PMID: 26796550

4. Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics. 2006;173(1):419–34. https://doi.org/10.1534/genetics.105.047985 PMID: 16204214

5. Villoutreix R, de Carvalho CF, Soria-Carrasco V, Lindtke D, De-la-Mora M, Muschick M. Large-scale mutation in the evolution of a gene complex for cryptic coloration. Science. 2020;369(6502):460–6.

6. Harringmeyer OS, Hoekstra HE. Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. Nat Ecol Evol. 2022;6(12):1965–79. https://doi.org/10.1038/s41559-022-01890-0 PMID: 36253543

7. Lowry DB, Willis JH. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. 2010;8(9):e1000500. https://doi.org/10.1371/journal.pbio.1000500 PMID: 20927411

8. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, et al. A common inversion under selection in Europeans. Nat Genet. 2005;37(2):129–37. https://doi.org/10.1038/ng1508 PMID: 15654335

9. Westram AM, Faria R, Johannesson K, Butlin R, Barton N. Inversions and parallel evolution. Philosophical Transactions of the Royal Society B: Biological Sciences. 2022;377(1856):20210203. https://doi.org/10.1098/rstb.2021.0203

10. Gorkovskiy A, Verstrepen KJ. The Role of Structural Variation in Adaptation and Evolution of Yeast and Other Fungi. Genes (Basel). 2021;12(5):699. https://doi.org/10.3390/genes12050699 PMID: 34066718

11. Berdan EL, Barton NH, Butlin R, Charlesworth B, Faria R, Fragata I, et al. How chromosomal inversions reorient the evolutionary process. J Evol Biol. 2023;36(12):1761–82. https://doi.org/10.1111/jeb.14242 PMID: 37942504

12. Wellenreuther M, Bernatchez L. Eco-Evolutionary Genomics of Chromosomal Inversions. Trends Ecol Evol. 2018;33(6):427–40. https://doi.org/10.1016/j.tree.2018.04.002 PMID: 29731154

13. Tuttle EM, Bergland AO, Korody ML, Brewer MS, Newhouse DJ, Minx P, et al. Divergence and Functional Degradation of a Sex Chromosome-like Supergene. Curr Biol. 2016;26(3):344–50. https://doi.org/10.1016/j.cub.2015.11.069 PMID: 26804558

14. Mérot C, Berdan EL, Cayuela H, Djambazian H, Ferchaud A-L, Laporte M, et al. Locally Adaptive Inversions Modulate Genetic Variation at Different Geographic Scales in a Seaweed Fly. Mol Biol Evol. 2021;38(9):3953–71. https://doi.org/10.1093/molbev/msab143 PMID: 33963409

15. Koch EL, Morales HE, Larsson J, Westram AM, Faria R, Lemmon AR, et al. Genetic variation for adaptive traits is associated with polymorphic inversions in Littorina saxatilis. Evol Lett. 2021;5(3):196–213. https://doi.org/10.1002/evl3.227 PMID: 34136269

16. Brown KS join(' ', Benson WW. Adaptive Polymorphism Associated with Multiple Mullerian Mimicry in Heliconius numata (Lepid. Nymph.). Biotropica. 1974;6(4):205. https://doi.org/10.2307/2989666

17. Küpper C, Stocks M, Risse JE, Dos Remedios N, Farrell LL, McRae SB, et al. A supergene determines highly divergent male reproductive morphs in the ruff. Nat Genet. 2016;48(1):79–83. https://doi.org/10.1038/ng.3443 PMID: 26569125

18. White BJ, Collins FH, Besansky NJ. Evolution of Anopheles gambiae in relation to humans and malaria. Annual Review of Ecology, Evolution, and Systematics. 2011;42:111–32.

19. Lowry DB, Rockwood RC, Willis JH. Ecological reproductive isolation of coast and inland races of Mimulus guttatus. Evolution. 2008;62(9):2196–214.

20. Butlin RK, Read IL, Day TH. The effects of a chromosomal inversion on adult size and male mating success in the seaweed fly, Coelopa frigida. Heredity. 1982;49(1):51–62. https://doi.org/10.1038/hdy.1982.64

21. Johannesson B. Shell morphology of Littorina saxatilis Olivi: The relative importance of physical factors and predation. Journal of Experimental Marine Biology and Ecology. 1986;102(2–3):183–95. https://doi.org/10.1016/0022-0981(86)90175-9

22. Lowther JK. Polymorphism in the white-throated sparrow, Zonotrichia albicollis (Gmelin). Can J Zool. 1961;39(3):281–92.

23. Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet. 2004;36(11):1133–7. https://doi.org/10.1038/ng1104-1133 PMID: 15514660

24. Crombie TA, McKeown R, Moya ND, Evans KS, Widmayer SJ, LaGrassa V, et al. CaeNDR, the Caenorhabditis Natural Diversity Resource. Nucleic Acids Res. 2024;52(D1):D850–8. https://doi.org/10.1093/nar/gkad887 PMID: 37855690

25. Lister C, Dean C. Recombinant inbred lines for mapping RFLP and phenotypic markers in Arabidopsis thaliana. The Plant Journal. 1993;4(4):745–50. https://doi.org/10.1046/j.1365-313x.1993.04040745.x

26. Huang W, Massouras A, Inoue Y, Peiffer J, Ràmia M, Tarone AM, et al. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. 2014;24(7):1193–208. https://doi.org/10.1101/gr.171546.113 PMID: 24714809

27. Li H, Ralph P. Local PCA Shows How the Effect of Population Structure Differs Along the Genome. Genetics. 2019;211(1):289–304. https://doi.org/10.1534/genetics.118.301747 PMID: 30459280

28. Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV, et al. Long-range LD can confound genome scans in admixed populations. Am J Hum Genet. 2008;83(1):132–5; author reply 135-9. https://doi.org/10.1016/j.ajhg.2008.06.005 PMID: 18606306

29. Ayala D, Zhang S, Chateau M, Fouet C, Morlais I, Costantini C, et al. Association mapping desiccation resistance within chromosomal inversions in the African malaria vector Anopheles gambiae. Mol Ecol. 2019;28(6):1333–42. https://doi.org/10.1111/mec.14880 PMID: 30252170

30. González JR, Ruiz-Arenas C, Cáceres A, Morán I, López-Sánchez M, Alonso L, et al. Polymorphic Inversions Underlie the Shared Genetic Susceptibility of Obesity-Related Diseases. Am J Hum Genet. 2020;106(6):846–58. https://doi.org/10.1016/j.ajhg.2020.04.017 PMID: 32470372

31. Harbison ST, McCoy LJ, Mackay TFC. Genome-wide association study of sleep in Drosophila melanogaster. BMC Genomics. 2013;14:281. https://doi.org/10.1186/1471-2164-14-281 PMID: 23617951

32. Cáceres A, González JR. Following the footprints of polymorphic inversions on SNP data: from detection to association tests. Nucleic Acids Res. 2015;43(8):e53. https://doi.org/10.1093/nar/gkv073 PMID: 25672393

33. Feuk L. Inversion variants in the human genome: role in disease and genome architecture. Genome Med. 2010;2(2):11. https://doi.org/10.1186/gm132 PMID: 20156332

34. Kapun M, Flatt T. The adaptive significance of chromosomal inversion polymorphisms in Drosophila melanogaster. Mol Ecol. 2019;28(6):1263–82. https://doi.org/10.1111/mec.14871 PMID: 30230076

35. Aulard S, David JR, Lemeunier F. Chromosomal inversion polymorphism in Afrotropical populations of Drosophila melanogaster. Genet Res. 2002;79(1):49–63. https://doi.org/10.1017/s0016672301005407 PMID: 11974603

36. de Jong G, Bochdanovits Z. Latitudinal clines in Drosophila melanogaster: Body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. J Genet. 2003;82(3):207–23.

37. Durmaz E, Benson C, Kapun M, Schmidt P, Flatt T. An inversion supergene in Drosophila underpins latitudinal clines in survival traits. J Evol Biol. 2018;31(9):1354–64. https://doi.org/10.1111/jeb.13310 PMID: 29904977

38. García-Vázquez E, Sánchez-Refusta F. Chromosomal polymorphism and extra bristles of Drosophila melanogaster: joint variation under selection in isofemale lines. Genetica. 1988;78(2):91–6.

39. Hoffmann AA, Sgrò CM, Weeks AR. Chromosomal inversion polymorphisms and adaptation. Trends Ecol Evol. 2004;19(9):482–8. https://doi.org/10.1016/j.tree.2004.06.013 PMID: 16701311

40. Hoffmann AA, Rieseberg LH. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annual Review of Ecology, Evolution, and Systematics. 2008;39(1):21–42.

41. Kapun M, Schmidt C, Durmaz E, Schmidt PS, Flatt T. Parallel effects of the inversion In(3R)Payne on body size across the North American and Australian clines in Drosophila melanogaster. J Evol Biol. 2016;29(5):1059–72. https://doi.org/10.1111/jeb.12847 PMID: 26881839

42. Kamping A, Van Delden W. The role of fertility restoration in the maintenance of the inversion In(2L)t polymorphism in drosophila melanogaster. Heredity (Edinb). 1999;83 (Pt 4):460–8. https://doi.org/10.1038/sj.hdy.6885980 PMID: 10583548

43. Nunez JCB, Lenhart BA, Bangerter A, Murray CS, Mazzeo GR, Yu Y, et al. A cosmopolitan inversion facilitates seasonal adaptation in overwintering Drosophila. Genetics. 2024;226(2):iyad207. https://doi.org/10.1093/genetics/iyad207 PMID: 38051996

44. van Delden W, Kamping A. The Association Between the Polymorphisms at the Adh and Agpdh Loci and the In(2l)t Inversion in Drosophila Melanogaster in Relation to Temperature. Evolution. 1989;43(4):775–93. https://doi.org/10.1111/j.1558-5646.1989.tb05176.x PMID: 28564191

45. Lenhart BA, Bergland AO. The inversion In(2L)t impacts complex, environmentally sensitive behaviors in Drosophila melanogaster. bioRxiv. 2025.

46. Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D. The Drosophila melanogaster Genetic Reference Panel. Nature. 2012;482(7384):173–8.

47. Nowling RJ, Manke KR, Emrich SJ. Detecting inversions with PCA in the presence of population structure. PLoS One. 2020;15(10):e0240429. https://doi.org/10.1371/journal.pone.0240429 PMID: 33119626

48. Corbett-Detig RB, Cardeno C, Langley CH. Sequence-Based Detection and Breakpoint Assembly of Polymorphic Inversions. Genetics. 2012;192(1):131–7.

49. Gardeux V, Bevers RPJ, David FPA, Rosschaert E, Rochepeau R, Deplancke B. DGRPool: A web tool leveraging harmonized Drosophila Genetic Reference Panel phenotyping data for the study of complex traits. bioRxiv. 2023.

50. Bergland AO, Tobler R, González J, Schmidt P, Petrov D. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in Drosophila melanogaster. Mol Ecol. 2016;25(5):1157–74. https://doi.org/10.1111/mec.13455 PMID: 26547394

51. Corbett-Detig R, Nielsen R. A Hidden Markov Model Approach for Simultaneously Estimating Local Ancestry and Admixture Time Using Next Generation Sequence Data in Samples of Arbitrary Ploidy. PLoS Genet. 2017;13(1):e1006529. https://doi.org/10.1371/journal.pgen.1006529 PMID: 28045893

52. Kao JY, Lymer S, Hwang SH, Sung A, Nuzhdin SV. Postmating reproductive barriers contribute to the incipient sexual isolation of the United States and Caribbean Drosophila melanogaster. Ecol Evol. 2015;5(15):3171–82. https://doi.org/10.1002/ece3.1596 PMID: 26357543

53. Pool JE. The Mosaic Ancestry of the Drosophila Genetic Reference Panel and the D. melanogaster Reference Genome Reveals a Network of Epistatic Fitness Interactions. Mol Biol Evol. 2015;32(12):3236–51. https://doi.org/10.1093/molbev/msv194 PMID: 26354524

54. Josse J, Husson F. missMDA: A package for handling missing values in multivariate data analysis. Journal of Statistical Software. 2016;70:1–31.

55. Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software. 2008;25:1–18.

56. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28(24):3326–8.

57. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet. 2014;46(2):100–6. https://doi.org/10.1038/ng.2876 PMID: 24473328

58. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008;91(11):4414–23. https://doi.org/10.3168/jds.2007-0980 PMID: 18946147

59. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet. 2011;19(7):807–12. https://doi.org/10.1038/ejhg.2011.39 PMID: 21407268

60. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. Am J Hum Genet. 2016;98(4):653–66.

61. Burdina EV, Gruntenko NE. Physiological aspects of Wolbachia pipientis–Drosophila melanogaster relationship. J Evol Biochem Phys. 2022;58(2):303–17.

62. Maistrenko OM, Serga SV, Vaiserman AM, Kozeretska IA. Longevity-modulating effects of symbiosis: insights from Drosophila-Wolbachia interaction. Biogerontology. 2016;17(5–6):785–803. https://doi.org/10.1007/s10522-016-9653-9 PMID: 27230747

63. Zug R, Hammerstein P. Bad guys turned nice? A critical assessment of Wolbachia mutualisms in arthropod hosts. Biol Rev Camb Philos Soc. 2015;90(1):89–111. https://doi.org/10.1111/brv.12098 PMID: 24618033

64. Durham MF, Magwire MM, Stone EA, Leips J. Genome-wide analysis in Drosophila reveals age-specific effects of SNPs on fitness traits. Nat Commun. 2014;5:4338. https://doi.org/10.1038/ncomms5338 PMID: 25000897

65. Krefl D, Bergmann S. Cross-GWAS coherence test at the gene and pathway level. PLoS Comput Biol. 2022;18(9):e1010517. https://doi.org/10.1371/journal.pcbi.1010517 PMID: 36156592

66. Marriage TN, King EG, Long AD, Macdonald SJ. Fine-mapping nicotine resistance loci in Drosophila using a multiparent advanced generation inter-cross population. Genetics. 2014;198(1):45–57. https://doi.org/10.1534/genetics.114.162107 PMID: 25236448

67. Mitchell CL, Latuszek CE, Vogel KR, Greenlund IM, Hobmeier RE, Ingram OK, et al. α-amanitin resistance in Drosophila melanogaster: A genome-wide association approach. PLoS One. 2017;12(2):e0173162. https://doi.org/10.1371/journal.pone.0173162 PMID: 28241077

68. Vaisnav M, Xing C, Ku H-C, Hwang D, Stojadinovic S, Pertsemlidis A, et al. Genome-wide association analysis of radiation resistance in Drosophila melanogaster. PLoS One. 2014;9(8):e104858. https://doi.org/10.1371/journal.pone.0104858 PMID: 25121966

69. Vonesch SC, Lamparter D, Mackay TFC, Bergmann S, Hafen E. Genome-Wide Analysis Reveals Novel Regulators of Growth in Drosophila melanogaster. PLoS Genet. 2016;12(1):e1005616. https://doi.org/10.1371/journal.pgen.1005616 PMID: 26751788

70. Watanabe LP, Riddle NC. GWAS reveal a role for the central nervous system in regulating weight and weight change in response to exercise. Sci Rep. 2021;11(1):5144. https://doi.org/10.1038/s41598-021-84534-w PMID: 33664357

71. Olazcuaga L, Loiseau A, Parrinello H, Paris M, Fraimout A, Guedot C. A whole-genome scan for association with invasion success in the fruit fly Drosophila suzukii using contrasts of allele frequencies corrected for population structure. Molecular Biology and Evolution. 2020;37(8):2369–85.

72. Kapun M, Nunez JCB, Bogaerts-Márquez M, Murga-Moreno J, Paris M, Outten J, et al. Drosophila Evolution over Space and Time (DEST): A New Population Genomics Resource. Mol Biol Evol. 2021;38(12):5782–805. https://doi.org/10.1093/molbev/msab259 PMID: 34469576

73. Blair LM, Granka JM, Feldman MW. On the stability of the Bayenv method in assessing human SNP-environment associations. Hum Genomics. 2014;8(1):1. https://doi.org/10.1186/1479-7364-8-1 PMID: 24405978

74. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383. https://doi.org/10.1371/journal.pgen.1004383 PMID: 24830394

75. Seich Al Basatena N-K, Hoggart CJ, Coin LJ, O'Reilly PF. The effect of genomic inversions on estimation of population genetic parameters from SNP data. Genetics. 2013;193(1):243–53. https://doi.org/10.1534/genetics.112.145599 PMID: 23150602

76. Morozova TV, Huang W, Pray VA, Whitham T, Anholt RRH, Mackay TFC. Polymorphisms in early neurodevelopmental genes affect natural variation in alcohol sensitivity in adult drosophila. BMC Genomics. 2015;16:865. https://doi.org/10.1186/s12864-015-2064-5 PMID: 26503115

77. Nunez JCB, Coronado-Zamora M, Gautier M, Kapun M, Steindl S, Ometto L, et al. Footprints of Worldwide Adaptation in Structured Populations of Drosophila melanogaster Through the Expanded DEST 2.0 Genomic Resource. Mol Biol Evol. 2025;42(8):msaf132. https://doi.org/10.1093/molbev/msaf132 PMID: 40824865

78. Kapun M, Fabian DK, Goudet J, Flatt T. Genomic Evidence for Adaptive Inversion Clines in Drosophila melanogaster. Mol Biol Evol. 2016;33(5):1317–36. https://doi.org/10.1093/molbev/msw016 PMID: 26796550

79. Rane RV, Rako L, Kapun M, Lee SF, Hoffmann AA. Genomic evidence for role of inversion 3RP of Drosophila melanogaster in facilitating climate change adaptation. Mol Ecol. 2015;24(10):2423–32. https://doi.org/10.1111/mec.13161 PMID: 25789416

80. Machado HE, Bergland AO, Taylor R, Tilk S, Behrman E, Dyer K. Broad geographic sampling reveals the shared basis and environmental correlates of seasonal adaptation in Drosophila. eLife. 2021;10:e67577.

81. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–9. https://doi.org/10.1038/ng1847 PMID: 16862161

82. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38(2):203–8. https://doi.org/10.1038/ng1702 PMID: 16380716

83. Baran Y, Quintela I, Carracedo A, Pasaniuc B, Halperin E. Enhanced localization of genetic samples through linkage-disequilibrium correction. Am J Hum Genet. 2013;92(6):882–94. https://doi.org/10.1016/j.ajhg.2013.04.023 PMID: 23726367

84. Calboli FCF, Koskinen H, Nousianen A, Fraslin C, Houston RD, Kause A. Conserved QTL and chromosomal inversion affect resistance to columnaris disease in 2 rainbow trout (Oncorhyncus mykiss) populations. G3 (Bethesda). 2022;12(8):jkac137. https://doi.org/10.1093/g3journal/jkac137 PMID: 35666190

85. Keller A. Drosophila melanogaster's history as a human commensal. Curr Biol. 2007;17(3):R77-81. https://doi.org/10.1016/j.cub.2006.12.031 PMID: 17276902

86. McKenzie JA, Parsons PA. Alcohol tolerance: An ecological parameter in the relative success of Drosophila melanogaster and Drosophila simulans. Oecologia. 1972;10(4):373–88. https://doi.org/10.1007/BF00345738 PMID: 28307067

87. Begon M. Yeasts and Drosophila. The Genetics and Biology of Drosophila. Academic Press. 1982. p. 345–84.

88. Anagnostou C, Dorsch M, Rohlfs M. Influence of dietary yeasts onDrosophila melanogasterlife-history traits. Entomologia Exp Applicata. 2010;136(1):1–11. https://doi.org/10.1111/j.1570-7458.2010.00997.x

89. Spieth HT. Courtship behavior in Drosophila. Annu Rev Entomol. 1974;19:385–405. https://doi.org/10.1146/annurev.en.19.010174.002125 PMID: 4205689

90. Fry JD. Mechanisms of naturally evolved ethanol resistance in Drosophila melanogaster. J Exp Biol. 2014;217(Pt 22):3996–4003. https://doi.org/10.1242/jeb.110510 PMID: 25392459

91. McKenzie JA, Parsons PA. Microdifferentiation in a natural population of Drosophila melanogaster to alcohol in the environment. Genetics. 1974;77(2):385–94. https://doi.org/10.1093/genetics/77.2.385 PMID: 4211152

92. Kreitman M. Nucleotide polymorphism at the alcohol dehydrogenase locus of Drosophila melanogaster. Nature. 1983;304(5925):412–7. https://doi.org/10.1038/304412a0 PMID: 6410283

93. Aquadro CF, Desse SF, Bland MM, Langley CH, Laurie-Ahlberg CC. Molecular population genetics of the alcohol dehydrogenase gene region of Drosophila melanogaster. Genetics. 1986;114(4):1165–90. https://doi.org/10.1093/genetics/114.4.1165 PMID: 3026893

94. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991;351(6328):652–4. https://doi.org/10.1038/351652a0 PMID: 1904993

95. Johnson FM, Denniston C. Genetic Variation of Alcohol Dehydrogenase in Drosophilia Melanogaster. Nature. 1964;204:906–7. https://doi.org/10.1038/204906a0 PMID: 14235730

96. Oakeshott JG, Gibson JB, Anderson PR, Knibb WR, Anderson DG, Chambers GK. Alcohol Dehydrogenase and Glycerol-3-Phosphate Dehydrogenase Clines in Drosophila melanogaster on Different Continents. Evolution. 1982;36(1):86.

97. Berry A, Kreitman M. Molecular analysis of an allozyme cline: alcohol dehydrogenase in Drosophila melanogaster on the east coast of North America. Genetics. 1993;134(3):869–93. https://doi.org/10.1093/genetics/134.3.869 PMID: 8102342

98. Kamping DWV. A long-term study on interactions between the Adh and αGpdh allozyme polymorphisms and the chromosomal inversion In(2L)t in a seminatural population of D. melanogaster. J Evol Biol. 1999;12(4):809–21.

99. van Delden W. The Alcohol Dehydrogenase Polymorphism in Drosophila melanogaster. Evolutionary Biology. Springer US. 1982:187–222. https://doi.org/10.1007/978-1-4615-6968-8_4

100. Siddiq MA, Thornton JW. Fitness effects but no temperature-mediated balancing selection at the polymorphic Adh gene of Drosophila melanogaster. Proceedings of the National Academy of Sciences. 2019;116(43):21634–40.

101. Cavener DR, Clegg MT. Dynamics of correlated genetic systems. IV. Multilocus effects of ethanol stress environments. Genetics. 1978;90(3):629–44.

102. Geer BW, Heinstra PW, McKechnie SW. The biological basis of ethanol tolerance in Drosophila. Comp Biochem Physiol B. 1993;105(2):203–29. https://doi.org/10.1016/0305-0491(93)90221-p PMID: 8359013

103. Eanes WF, Merritt TJS, Flowers JM, Kumagai S, Zhu C-T. Direct evidence that genetic variation in glycerol-3-phosphate and malate dehydrogenase genes (Gpdh and Mdh1) affects adult ethanol tolerance in Drosophila melanogaster. Genetics. 2009;181(2):607–14. https://doi.org/10.1534/genetics.108.089383 PMID: 19033156

104. Watanabe TK, Watanabe T. Enzyme and chromosome polymorphisms in Japanese natural populations of Drosophila melanogaster. Genetics. 1977;85(2):319–29.

105. Inoue Y, Tobari YN, Tsuno K, Watanabe TK. Association of Chromosome and Enzyme Polymorphisms in Natural and Cage Populations of Drosophila Melanogaster. Genetics. 1984;106(2):267–77. https://doi.org/10.1093/genetics/106.2.267 PMID: 17246191

106. Koury SA. Predicting recombination suppression outside chromosomal inversions in Drosophila melanogaster using crossover interference theory. Heredity (Edinb). 2023;130(4):196–208. https://doi.org/10.1038/s41437-023-00593-x PMID: 36721031

107. van Delden W, Boerema AC, Kamping A. The alcohol dehydrogenase polymorphism in populations of Drosophila melanogaster. I. Selection in different environments. Genetics. 1978;90(1):161–91. https://doi.org/10.1093/genetics/90.1.161 PMID: 100371

108. Malpica J, Vassallo J. Epistatic interaction between the inversion In(2L)t and the Adh locus in Drosophila melanogaster. Heredity. 1987;58(2):227–31.

109. McKechnie SW, Geer BW. The epistasis of Adh and Gpdh allozymes and variation in the ethanol tolerance of Drosophila melanogaster larvae. Genet Res. 1988;52(3):179–84. https://doi.org/10.1017/s0016672300027634 PMID: 3149599

110. Alonso-Moraga A, Muñoz-Serrano A. Allozyme polymorphism and linkage disequilibrium of Adh and α-Gpdh loci in wine cellar and field populations of Drosophila melanogaster. Experientia. 1986;42(9):1048–50.

111. Sanchez-Refusta F, Santiago E, Rubio J. Seasonal fluctuations of cosmopolitan inversion frequencies in a natural population of Drosophila melanogaster. Genet Sel Evol. 1990;22(1):47–56.

112. Pearse DE, Barson NJ, Nome T, Gao G, Campbell MA, Abadía-Cardoso A, et al. Sex-dependent dominance maintains migration supergene in rainbow trout. Nat Ecol Evol. 2019;3(12):1731–42. https://doi.org/10.1038/s41559-019-1044-6 PMID: 31768021

113. Hager ER, Harringmeyer OS, Wooldridge TB, Theingi S, Gable JT, McFadden S, et al. A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. Science. 2022;377(6604):399–405.

114. Furusawa C, Kaneko K. Formation of dominant mode by evolution in biological systems. Phys Rev E. 2018;97(4):042410.

115. Mei H, Cuccaro ML, Martin ER. Multifactor dimensionality reduction-phenomics: a novel method to capture genetic heterogeneity with use of phenotypic variables. Am J Hum Genet. 2007;81(6):1251–61. https://doi.org/10.1086/522307 PMID: 17999363

116. Berner D. Size correction in biology: how reliable are approaches based on (common) principal component analysis?. Oecologia. 2011;166(4):961–71. https://doi.org/10.1007/s00442-011-1934-z PMID: 21340614

117. García-Ríos E, Nuévalos M, Barrio E, Puig S, Guillamón JM. A new chromosomal rearrangement improves the adaptation of wine yeasts to sulfite. Environ Microbiol. 2019;21(5):1771–81. https://doi.org/10.1111/1462-2920.14586 PMID: 30859719

118. Giner-Delgado C, Villatoro S, Lerga-Jaso J, Gayà-Vidal M, Oliva M, Castellano D, et al. Evolutionary and functional impact of common polymorphic inversions in the human genome. Nat Commun. 2019;10(1):4222. https://doi.org/10.1038/s41467-019-12173-x PMID: 31530810

119. Huang K, Rieseberg LH. Frequency, origins, and evolutionary role of chromosomal inversions in plants. Front Plant Sci. 2020;11.

120. Merrikh CN, Merrikh H. Gene inversion potentiates bacterial evolvability and virulence. Nat Commun. 2018;9(1):4662. https://doi.org/10.1038/s41467-018-07110-3 PMID: 30405125