**RESULTS [summary of claims]**

**Fly populations are temporally structured:** We used principal component analysis (**PCA**) to characterize patterns of genetic variation across our pooled samples from the DEST dataset. Consistent with previous analyses [**REF**], the primary signal of the PCA reveals spatial structure. PC1 separates samples from Europe and North America (**Fig 1A**), whereas PC2 separate the eastern and Western phylogeographic clusters in Europe. Aside from continent-level structure, the PCs also contains latent signatures of temporal structure (**Table S1**). The overall pattern of temporal structure can be visualized through the vector formed by the mean projections between the first and last samples collected in every population (arrows in **Figs 1A and 1B; Fig S1**).

**Temporal structure is driven by seasonal boom-bust demography:** The pattern of temporal structure observed in our global data is consistent with a demographic scenario of boom-bust seasonal demography. To test this hypothesis, we conducted genetic differentiation analysis (*F*ST) comparing genetic patternswithin a year’s growing season (~ 13 generations), relative to populations separated by overwintering across years (i.e., last winter/first spring flies, ~2 generations). Based on the dynamics of boom-and-bust demography [**REF**], we expect populations to become more differentiated across the two generation of overwinter, due to strong winter drift, than across the 13 generations of the growing season. Consistent with the boom-bust model, our populations show the expected patterns and significantly different *F*ST distributions (**Fig 1C; Table S2**). Notably, this pattern of increased differentiation is observed across multiple years of overwintering (Δy**; Fig 1D; Table S3**). We further explored the overwintering signal using forward genetic simulations [**REF**] designed to emulate the boom-bust cycle. Comparing our simulations results to the patterns observed in Virginia (the population with the highest sampling resolution) provide support for overwintering scenarios and suggest that populations experience a winter collapse of approximately X% of their maximum summer size [Supp evidence connor].

**Temporal drift drives genetic structure across the genome:** PCA analyses of Virginian samples across distinct chromosomal arms reveals two salient patterns. First, we observe that chromosome arms 2R, 3L and 3R recapitulate the patters of temporal structure seen genome wide (**Fig 2A**). Second, chromosome arm 2L shows a unique pattern, distinct from the rest of the genome (**Fig 2A;** top left). To assess the potential drivers of these genetic patterns we conducted a correlation analysis comparing PC projections, calculated using a locus resampling scheme, to several variables of interest (e.g., year of collection, frequency of cosmopolitan inversions, as well as the effective coverage, Nc, as a nuisance variable). Our resampling analysis shows that PCs 1 and 2 are significantly correlated with year of collection, though the correlation of PC1 is the weakest in 2L (**Fig 2B; Fig S3**) [Supp evidence 🡪 correlation values; plot]. Notably, the frequency of the cosmopolitan inversion In(2L)t is the strongest predictor for PC1 projections in 2L. – **other pops [currently doing this]**

We formally tested the hypothesis that most of the fly genome experiences genetic drift in response to short term temporal evolution. Accordingly, we regressed levels of allele frequency of each locus (AF*i*) onto the year of collection (as a factor) using a generalized linear model framework with a quasibinomial error term (GLM*qb*). Formally, we fitted both a null and an effect model. The null model where AF are regressed against their mean (μAF) and the effect model using collection year. We compared the goodness of fit of these models using a likelihood ratio test (LRT). We refer to this LRT comparison as the AF~Y comparison. We summarized the AF~Y analysis using two methods. First, we explored the distribution of ranked normalized LRT *P*-values (rn*P*v) and the wZa statistic, both across genomic windows (100 Kb width and 50 Kb steps). *P*-value rank normalization is a robust method to assess whether regions of the genome are enriched for the temporal signal. Particularly, we assessed if genomic windows harbor more outlier loci than expected by chance (using a significance threshold of α = 0.05). We conducted this analysis for each chromosome arm, conditioned on whether windows were located inside or outside of a cosmopolitan inversion. We also compared results for both real and permuted configurations of the data. The rn*P*v analysis indicates that the temporal signal is enriched across the genome regardless of cosmopolitan inversions (**Fig 3A; Supp fig 🡪 raw output**). The wZa analysis is based on Stouffer’s *P*-value aggregation method and represents the strength of signal across windows. Consistent with the rn*P*v results, all chromosome arms show strong temporal signal for the AF~Y model (**Fig 3C; Supp fig 🡪 raw output**). Overall, these results are consistent with the idea that temporal drift drives genome-wide levels of standing genetic variation in wild fly populations.

**Loci within In(2L)t responds to temperature selection:** Concordant with our observations, previous data had proposed that In(2L)t may be a locus under seasonal selection **[REF]**. Accordingly, we tested the association between AF*i* and the mean temperature 30 days prior to sampling (T30). In this context, we are utilizing aggregated temperatures as a proxy predictor for the large environmental process of seasonality [Is this justified in the lit?]. We fitted two different GLM*qb* models, one fitting AF to the year of collection alone (a null model), and a second one fitting AF to both the year of collection and T30 as additive effects. The LRT of these models, henceforth AF~Y+T30, allow us to discover outlier temperature loci, while explicitly controlling for temporal structure. We applied the same statistics as above, rn*P*v and wZa. Both statistics reveal that In(2L)t is strongly enriched for loci responding to temperature (**Fig 3B and 3D**) [Supp evidence 🡪 rnPvs for the whole genome]. Window level values for rn*P*v and wZa, within 2L, reveal five regions of interest (i.e., signal peaks) within the inversion which drive the overall signal (**Fig 3E**). These regions are putative targets of natural selection and are named based on their genomic location: window (*w*)*4.6*, *w5.1*, *w6.2*, *w6.8*, and *w9.5* (**Supp fig 🡪** table with Mb coordinates).

**Temperature haplotypes in In(2L)t show partial selective sweeps and long distance LD:** We conducted follow-up analyses on in(2L)t using our alternative dataset of wild-caught individual sequencies from Virginia. We classified the inversion status of the individual sequences by training a linear support vector machine classifier (SVM) based on 47 highly informative inversion markers present in the DGRP panel. Of our 203 phased wild-caught Virginian samples,152 are homozygous for the standard 2L configuration (i.e., No Inv.), nine are homozygotes for the inversion (Inv.), and 42 are heterozygous. We used this data to calculate Tajima’s *D*, a statistic used to detect departures from neutral models of evolution. Our results reveal that samples which are homozygous for In(2L)t inversion harbors regions with strong departures from neutrality (D ≠ 0), two of which match our AF~Y+T30 outlier windows (*w6.2* and *w9.5*; **Fig 3F**). The nature of the signal in these windows (D < -1.8) is consistent with a partial sweep targeting only inverted backgrounds [**REF**]. Furthermore, levels of linkage disequilibrium (LD) across the phased data show strong levels of linkage within and among our windows of interest (**Fig 3G**), with the strongest linkage observed within *w6.2* (**Fig 3H**), and between *w5.1* and *w6.2* (**Fig 3I**).

**Temperature haplotypes in In(2L)t are present across the North American east coast:** We combined our phased Virginia data (homozygous samples only) with inbreed samples from the DGRP (North Carolina), as well as from Pennsylvania and Maine (whose inversion status were determined with our SVM) to create a joint dataset. A sequence plot of this data reveals shared haplotypes across North America present at the AF~Y+T30 outlier windows (**Fig 4A**). Notably, *w6.2* and *w9.5*, show the putatively swept haplotypes corresponding to the Troughs of the Tajima’s D analysis. For *w9.5* these haplotypes are seen across all samples. For *w6.2* the swept haplotype is seen only in Virginia and Maine. Using the LD estimates from the individual data, we identified mutations with strong and consistent r2 in Virginia. We used the average frequency of these mutations as estimators of haplotype frequencies in the pooled data (i.e., *haplotag* mutations). Notably, we observe that the frequencies of the inversion, and haplotypes within, are inversely proportional to seasonal temperature (**Fig 4B**) [Supp evidence 🡪 individual SNP plots].

**??In(2L)t’s response to temperature is concordant with XYZ (world-wide analysis) …**

Pending

**??Behavior phenotypes are associated with genetic variation at In(2L)t outlier loci.**

Pending

**Fig 1: Signatures of overwintering in flies.**

1. PCA of selected DEST samples. Dimensions 1 and 2 are shown. The arrow path (i.e., time path) indicates the temporal identity of the samples, the front of the path is guided by the most recent samples, and the origin of the path is guided by the oldest samples.
2. Same as A but for dimensions 2 and 3.
3. Genetic differentiation (*F*ST) across all within growing season (red) and between year (green, i.e., across overwintering periods).
4. *F*ST values across multiple years of collection (Δy is the difference in years of collection).

A screenshot of a computer

Description automatically generated with medium confidence

**Fig 2: Temporal structure in Virginian flies.**

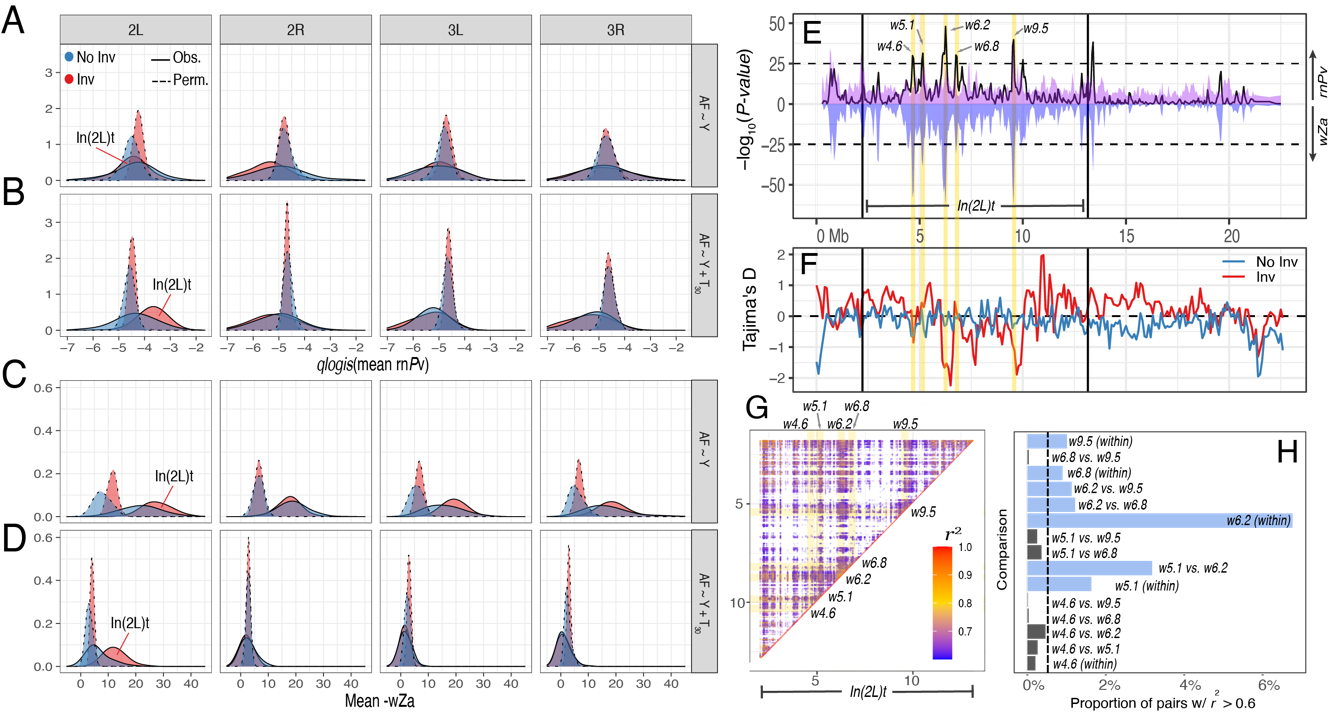
1. Chromosome level PCA. Color-coded as a function of collection year.
2. Correlation analyses of PCA dimensions build using a bootstrap-and-resample approach compared to two explanatory variables: year of collection and inversion frequency of In(2L)t. Square correlation values are shown. Chromosome arms are indicated in colors, with gray being permutations of the data.

Chart, scatter chart

Description automatically generated

**Fig 3: Temporal and temperature models in Virginia.**

1. Distribution of mean rn*P*vs (transformed using the logarithmic distribution) across the genome for the AF~Y model. Color indicates whether rn*P*vs come from a region inside or outside a cosmopolitan inversion. Solid lines are the observed data, dashed lines represent permutations.
2. Same as B, but for the AF~Y+T30 model.
3. Distribution of the wZA statistic across the genome for the AF~Y model (colors and line type as the same as A and B).
4. Same as C, but for the AF~Y+T30 model.
5. Reflected Manhattan plot showing the two-summary metrics. On the positive y-axis, the black line shows the significance of the rn*P*v enrichment test of AF~Y+T30 in 2L. The purple surface shows the 99th percentile value of permutations. On the negative y-axis the blue surface shows the significance of the wZa enrichment test. The boundaries of In(2L)t are indicated by vertical lines. The five windows of interest (*w4.6*, *w5.1*, *w6.2*, *w6.8*, and *w9.5*) are highlighted in yellow.
6. Values of Tajima’s D across 2L for two groups of homozygous samples with the inversion (red) and standard (blue) karyotypes.
7. Pairwise LD values (*r*2) for all AF~Y+T30 loci within In(2L)tmarkers (inversion markers are also included). Only values LD values greater than 0.6 are shown.
8. Proportion of pairs with *r*2 > 0.6 within and across windows of interest. The dashed line represents the mean proportion across In(2L)t.



**Fig 4: Haplotype structure of temperature loci.**

1. Sequence alignment plot showing the haplotype structures within windows of interest from the AF~Y+T30 model. The red and blue color represent ancestral and derived based calls, respectively, relative to *D. simulans*. The two flanking panels represent markers, and the horizontal black line denotes inverted vs. standard haplotypes. The dendrogram on the left side represent sample clustering based on haplotype identity. The color of the tips indicates the population of origin of the haplotype. The bottom panel shows the functional annotation of the individual loci as indicated by colors. The black tick-marks on the top indicate *haplotag* loci, highly linked mutations used to estimate haplotype frequencies.
2. Estimated frequencies of the 5 haplotype windows in the pooled data as a function of mean temperature (30 d prior to collection; left) and day of collection (right). Different collection years are shown by shape and color. In the right panel, the color shows mean temperature values.

A picture containing text

Description automatically generated