

## Supplementary Information for

Genome-wide variation and transcriptional changes in diverse developmental processes underly the rapid evolution of seasonal adaptation

Edwina J. Dowle, Thomas H.Q. Powell, Meredith M. Doellman, Peter J. Meyers, McCall B. Calvert, Kimberly K.O. Walden, Hugh M. Robertson, Stewart H. Berlocher, Jeffrey L. Feder, Daniel A. Hahn, and Gregory J. Ragland

Gregory Ragland, Department of Integrative Biology, University of Colorado Denver, Denver, CO, 80217, USA

Email: [gregory.ragland@ucdenver.edu](mailto:gregory.ragland@ucdenver.edu)

Edwina Dowle, Department of Anatomy, University of Otago, Dunedin, 9016, New Zealand

Email: [eddy.dowle@otago.ac.nz](mailto:eddy.dowle@otago.ac.nz)

This PDF file includes:

- Supplementary text
- Figures S1 to S7
- Tables S1 to S5
- Legends for Datasets S1 to S5
- SI References

Other supplementary materials for this manuscript include the following:

- Datasets S1 to S5

## Supplementary Methods

### Genome sequencing and assembly

DNA was extracted from both single and pooled adults of *Rhagoletis zephyria* (collected from East Lansing, MI) by grinding in liquid nitrogen before lysing overnight in an SDS solution with Proteinase K at 50C. The homogenate was treated with RNaseA at 37C, and proteins/debris were collected after high-salt precipitation overnight and centrifugation at 4C. After ethanol precipitation, the DNA was resuspended in 10 mM Tris and evaluated on an agarose gel and by Qubit (Invitrogen) fluorometric quantification.

We used several different library construction protocols for genomic DNA sequencing. Three independent 500bp-insert paired-end Illumina shotgun libraries were constructed from the DNA of a single adult female *R. zephyria*. And, a 1.5kb-insert paired-end shotgun library was constructed from DNA from two (pooled) female flies. The 500bp-insert and 1.5kb-insert libraries were prepared using Illumina TruSeq DNaseq Sample Prep kits and sequenced with the Illumina TruSeq SBS v3 sequencing kit for 100nt on an Illumina HiSeq 2000. Each 500bp-insert library was loaded onto a single full lane, while the 1.5kb-insert library was loaded at half-density onto a single lane. In addition to the sequencing above, a 5kb insert mate-pair library was constructed from a pool of 14 mixed sex adults, and 10 and 20kb insert libraries were constructed from a single pool of 50 mixed sex adults using an Illumina TruSeq DNaseq Sample Prep kit with a custom linker ligated between the read ends to enable mate-pair recovery or the Nextera Mate-Pairs Sample Prep kit from Illumina. These libraries were sequenced with the Illumina TruSeq SBS v.3 sequencing kit for 100 cycles on an Illumina HiSeq2000. The 10kb and 20kb insert libraries were made from DNA extracted from *Rhagoletis pomonella*; these reads were used primarily for scaffolding the *R. zephyria* genome. All bases were called with Casava 1.8.

To complement the traditional paired-end and insert library sequencing above, we also prepared and sequenced Illumina TruSeq Synthetic Long-Read (TSLR) libraries. Four independent TSLR libraries were constructed from a pool of mixed-sex adults using the TSLR Sample Prep kit (Illumina). Reads were sequenced 150nt from each end on 4 lanes of an Illumina HiSeq2500 and assembled using the Illumina TruSeq Long-Read Assembly Application in BaseSpace.

All libraries were prepared and sequenced at the William M. Keck Center at the University of Illinois at Urbana-Champaign.

Sequence reads were evaluated with FASTQC and then trimmed to remove 5' bias and 3' low quality bases using the FASTX Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) or Trimmomatic-0.32 [1]. Long-insert mate-pair libraries were filtered to remove redundancy and retain mates of appropriate distance and orientation using custom scripts. Trimmed reads were error-corrected per library with Quake (counting 19-mers) prior to assembly [2].

The *R. zephyria* genome was initially assembled in SOAPdenovo 2.04 [3] using the 500bp-insert library reads and the kmer parameter set to K=49. Scaffolding was initiated in SOAPdenovo using progressively longer insert *R. zephyria* libraries with additional scaffolding using the 10kb and 20kb-insert *R. pomonella* mate-pair libraries. Gaps were closed using GapCloser v1.12 [4] and the short-insert library reads. SOAPdenovo estimated average sequence coverage for the genome at 38X.

Assembly contiguity was improved using PBJelly2 [5] for further gap-filling using the *R. zephyria* TSLR reads. The SOAPdenovo assembly was first filtered for sequences less than 1kb and then potential haplotype copies at >90% identity using the dedupe algorithm in the bbmap package (<https://sourceforge.net/projects/bbmap/>) before running PBJelly2 with default parameters.

An initial screen for contaminant sequences using BLAST against the NCBI nr database identified 285 assembly scaffolds and contigs with high sequence identity to fungi and bacteria, which were manually removed. The assembly was additionally screened for vector sequences using a custom script. A final screen during the NCBI genome submission process identified 40 contigs identical to Wolbachia endosymbiont sequences, which were also removed from the assembly.

A Trinity assembly of the *R. pomonella* transcriptome [6] was leveraged for additional genome scaffolding. Non-redundant transcripts were selected at a 95% identity cut-off with the UCLUST algorithm in USEARCH [7] and only retained if they had a predicted CDS greater than 100 amino acids using TransDecoder from the Trinity package [8]. BLAT [9] was used with default settings to align the filtered transcripts to the genome prior to scaffolding with L\_RNA\_scaffolder [10].

Genome assembly completeness was assessed using the BUSCO 3.0.1 [11] pipeline and the Endopterygota\_odb9 dataset of 2,442 conserved ortholog sequences. Complete models were built for 96% of the orthologs (85% single-copy, 11% duplicated), with only 2% orthologous models fragmented and 2% orthologous models not detected.

Raw data accessions for the genome: BioProject = PRJNA321204, BioSample = SAMN04977950, SRA runs =SRR3670117, SRR3670118, SRR3670120, SRR3680351, SRR3670652, SRR3676006, SRR3676007, SRR3676008, SRR3676010

#### *Fly collection, rearing, subject selection, and tissue collection for RNAseq*

Infested apple and hawthorn fruits were collected from Urbana, Illinois (IL), USA in the late summer and fall of 2014 and transported to the University of Florida. Emerging larvae were collected and held at ~24°C 14:10 LD for 12 days to complete pupation and enter diapause; all pupae failing a screen for low metabolic rate (non-diapause pupae have high metabolic rates) were excluded from the study. Between days 9 and 11 after

larvae emerged from the fruit, the metabolic rate of each pupa was measured using stop-flow respirometry, following the methods of [12]. Only flies that showed the distinct low, baseline metabolic rate indicative of diapause development in this species [12] were placed into overwintering treatments for subsequent RNAseq analysis.

Diapausing pupae were transferred to simulated overwinter conditions at 3.5°C at 85% Relative Humidity (RH). While still in a cold room, diapausing pupae were randomly assigned to overwintering treatments (with assignments stratified across the range of larval emergence times) in groups of at least 40 individuals in a petri dish for a single pooled replicate. At monthly intervals from 2 to 6 months, we dissected out the Central Nervous System (CNS) of 20 individuals per replicate (the remaining pupae in each treatment were subsequently used for eclosion phenology phenotyping). Pupae were taken from the overwintering cold room in a cooler, and one pupae was removed from the cold at a time for dissection under Ringer's solution and immediately crushed with a pestle in cold Trizol solution. Once 20 CNS samples were crushed in the solution, the total pooled sample was homogenized using a bead-beater.

### Differential expression analysis

The full experiment included RNAseq libraries from 2 populations X 5 time points X 5 replicate pools per time point, yielding 50 total samples. Samples had an average of 30K SE reads (range 57K: 14.6K). Adapter contamination and low quality bases were removed using Trimmomatic [1] and reads were mapped to the *Rhagoletis zephyria* genome via the spliced aligner STAR [13]. Total mapped reads statistics averaged >90% across samples. Read counts per gene were generated using RSEM [14], and read counts per exon were generated using featureCounts [15]. All downstream analyses were performed using the R package edgeR [16]. We filtered out all transcripts that were not represented by at least one count in half of the samples. Count data were normalized using the weighted trimmed mean of M-values (TMM) method [17] and a Multi-Dimensional Scaling (MDS) plot was generated using the top 500 differentially expressed genes (determined strictly by fold-change difference among libraries using a log counts per million metric). The MDS plot showed two hawthorn fly 4-month samples

that were clear outliers (Fig. S5); further inspection showed that these samples had abnormally small library sizes, and they were removed from further analysis, leaving 3 apple fly 4-month replicate samples.

We fit a full generalized linear model (negative binomial error distribution) in edgeR including the fixed effects of time and host race, translating to a design matrix with 5 and 1 parameter(s) for time and host race, respectively. We did not explicitly model an interaction term, as statistical power to detect interactions would be low. Rather, we tested for the marginal effects of host race *within* months using contrasts (see below). Log2 fold changes and their statistical significance (with an FDR correction for multiple testing) were estimated from the full model using linear contrasts and a likelihood ratio test. Transcripts showing time effects common to both host races were identified as those transcripts exhibiting at least one significant ( $FDR < 0.05$ ) difference between a 3 – 6 month time point and the initial time point in both host races. Estimating expression relative to the earliest time point is a well-established approach for time series analysis of transcriptomic data [18]. We estimated host race differences by identifying all transcripts significantly differentially expressed between the host races within any time point. Trajectories of transcript expression across time were all estimated as log2 fold changes of months 3 – 6 relative to the 2-month time point.

### Time series clustering analysis

We used functions from the WGCNA package in R [19] to perform clustering analysis of differential expression over time. For clustering transcripts differentially expressed over time but not between host races, all differential expression was calculated relative to the 2 month time point for haw flies. For clustering transcripts differentially expressed over time and between host races, differential expression for each host race was calculated relative to the 2 month time point for that host race. We calculated the adjacency matrix using the Pearson correlation as the similarity metric, and calculated the distance matrix using one minus the topological overlap matrix calculated using the *TOMsimilarity* function. We determined initial clustering assignments using the dynamic tree cut

method of [19] implemented with the *cutreeDynamic* function with minimum module size of 30, *deepSplit* = 2, and *pamRespectsDendro* = FALSE. Next, we calculated the eigengene trajectories using *moduleEigengenes*, then merged all modules with a distance < 0.25. We tried various values for the *deepSplit* parameter and the merge threshold to produce more or fewer modules, but all analyses revealed the same pattern of progressive (monotonic) change over time.

### Network analysis

To estimate the connectivity of each transcript to the transcriptional network, we calculated pair-wise correlations (absolute value of the Pearson correlation coefficient) of expression trajectories over time between all transcripts significantly differentially expressed within at least one time point compared to the first (2 month) time point. We used the *findThreshold* function in the R package *coexnet* to identify a reasonable threshold to define an edge, estimated as 0.94 from our data. We constructed graphs with transcripts as nodes and retained correlations (23,755) as edges and estimated hub scores for each using the *igraph* package. The hub score is Kleinberg's hub centrality score, a metric of the centrality of each node. The associated R script is available here: [https://github.com/gjragland/Diverse-developmental-processes-influence-seasonal-phenology/blob/master/AnalyzeTrajectoriesNetwork\\_PNAS.r](https://github.com/gjragland/Diverse-developmental-processes-influence-seasonal-phenology/blob/master/AnalyzeTrajectoriesNetwork_PNAS.r)

### Brain size phenotyping

To test whether the brains of *R. pomonella* exhibit noticeable morphological development during diapause development, we took morphometric brain measurements from 6 – 12 pupae once a month, from 2 – 8 months post pupariation. Flies were sampled from Lansing, Michigan in Fall 2017; thus, they originated from a different population than that sampled for the RNAseq/Poolseq experiments. However, all populations of *R. pomonella* from the Midwest of the United States that have been studied in the context of diapause development follow the same developmental progression [20-22], and it is highly unlikely that brain morphogenesis is variable within

a species. Developing larvae and pupae were reared as described in the main text. The brains of higher flies undergo stereotypical, gross morphological changes after the molt into the pupal stage but prior to molt into the pharate adult, i.e., stage p1 to p4 in Bainbridge and Bownes [23], most noticeably the lateral widening of the left and right hemispheres and the development of the optic lobes [24]. We measured the anterior-posterior height and the lateral width of both hemispheres (Fig. S2) on all measured pupae, in addition to pupal wet weight. Linear regression of the average height and width of the hemispheres on body mass and month sampled (2 – 8 months) revealed no detectable effect of time (Table S1).

### Emergence time phenotyping

Feder et al. [25] showed that: 1) there is a curvilinear relationship between overwinter length and the timing of post-winter emergence (time to adult emergence, starting on the day of transfer from cold to warm conditions) in *R. pomonella*, and 2) that apple flies emerge earlier than hawthorn flies for most overwinter lengths. That study [25] was performed on flies collected from Michigan, USA and we wished to confirm that a similar pattern emerges using a similar experimental design with pupae collected from Urbana Illinois, USA, the field site used in the current study. A subset of diapausing pupae from the 2014 Urbana, IL field collections described in the methods section of the main text (same pre-winter and overwinter temperatures) were monitored for emergence at each of the RNAseq sampling time points (2 – 6 months) by transferring from the winter 4 °C treatment to 24 °C at 95% RH at the designated overwinter month (mean n = 190 per month) and scoring time to adult emergence. As expected, both host races demonstrated a curvilinear relationship between emergence time and winter length, with the longest emergence times associated with 3 and 4 months overwintering (Fig. S3). Apple flies emerged earlier than hawthorn flies at all overwintering treatments from 2 – 6 months.

The curvilinear relationship between emergence time and winter length seen for both host races in Figure S3 may seem at first to be at odds with diapause termination being



due to progressive changes in differential gene expression as pupae overwinter. If this were the case, then adult emergence time would be expected to decrease with increasing winter length as pupae become increasingly capacitated to terminate diapause through time. Indeed, in many insects with chilling requirements or for which low temperatures accelerate diapause development, there is a relatively linear, inverse relationship between emergence time and winter length [26]. And such a relationship is seen for both apple and haw *R. pomonella* flies following the 3-month treatment (Fig. S3). However, in both host races, pupae in the 2-month treatment eclosed earlier as adults than all other flies in the same race except those in the 6-month treatment (Fig. S3). Thus, in the absence of an extended chilling period, a portion of *R. pomonella* flies do not seem to have yet achieved a state of deep diapause and have the potential to terminate diapause if exposed to permissive temperature. This capacity appears to be lost in these flies after 3-months of chilling.

The curvilinear relationship between time spent in diapause and the propensity to terminate diapause is not unique to *R. pomonella*, as several insects have been shown to display the same pattern and be more prone to terminating diapause early and late in winter compared to in the middle of the diapause developmental trajectory [27, 28]. Two different phenotypic classes of early eclosion flies have been observed in *R. pomonella* related to this phenomenon [21]. First, in addition to the typical diapause phenotype that generally requires chilling for diapause termination, there are “non-diapause” phenotypes that can emerge without any exposure to cold. These non-diapause phenotypes never enter diapause and were removed from the current study using respirometric phenotyping (see ‘Fly collection, rearing, subject selection, and tissue collection for RNAseq’ methods above). A second class of early eclosion flies has been termed “shallow diapause” (SD) phenotypes that do enter diapause but are capable of terminating diapause without chilling. Moreover, when exposed to short winter durations at or below 2 months, SD phenotypes will emerge more rapidly than the chill dependent (CD) phenotypes. This pattern is described in [25], but at that time faster emergence at short winter lengths was attributed to non-diapause pupae, which were included in [25] but were excluded from the current study. Yet, in the current study we also observe

faster emergence at shorter winter lengths, consistent with the expectation of more rapid development for the SD phenotypes. [21, 29]. Thus, the relatively rapid emergence time for flies in the 2-month overwinter treatments reflect the influence of these SD phenotypes. This observation is not incompatible with the hypothesis that the rate of development influences the duration of diapause. Rather, it appears that the rate of diapause development is higher in SD phenotypes than CD phenotypes up to 2-months of overwintering, a difference that is genetically determined [29]. Indeed, the enrichment signal for transcripts differentially expressed between the host races at 2 months also strongly implicates developmental genes (see main text), and the incidence of the SD phenotype is higher in apple flies [21, 29]. After this time, diapause development for SD flies seems to converge on that for CD pupae, although studies remain to be conducted comparing emergence times between SD and CD phenotypes at longer overwinter lengths. However, additional genetic analyses suggest that genotypes associated with SD also affect emergence time at longer overwinter lengths [21, 29]. The ecological significance of the SD phenotype remains to be determined. However, the possibility exists that SD flies may constitute a potential second generation in regions of the southern USA where multiple species of hawthorn hosts exist with widely varying fruiting times ranging from May to early December [30]. At more northern sites such as Urbana, Illinois, the focus of the current study, phenotypic expression of early eclosion (second generation) may be repressed in SD flies by the onset of prolonged chilling during winter.

#### *Read cleaning, mapping, and variant detection for poolseq analyses*

Adapters were removed from Illumina reads using cutadapt [31] before reads were quality trimmed using Trimmomatic [1]. Cleaned reads were mapped to the *R. zephyria* reference genome via bwa v 0.7.16 [32]. PCR duplicates were removed using picardTools (<http://broadinstitute.github.io/picard/>) and reads were realigned around indels via GATK v 3.8 [33]. Reads counts per sample at each SNP variant were generated in ANGSD v 0.916 [34] using the following filters; minimum mapping score 30, minimum base score 30, MAF > 0.05 and individual depth per locus of 10 reads. All

MAF estimates were based on read counts. Read counts at each indel were generated using GATK via 'unified genotyper' using the same filters as above.

Given the fragmentation of the genome assembly, our analyses focused largely on individual variants rather than window-based analysis, an approach we have used in previous genome-wide analyses [35-37], though we did consider the distributions of variants across chromosomes and previously defined regions varying in linkage disequilibrium (LD; [37]. In total 28,133,855 SNPs and 1,667,189 indels passed our filters. On average, ~98% of reads mapped per sample, and sequencing depth estimates varied from 20-40X depending on sample and method of estimation (Dataset S5). To achieve maximum time separation between the eclosion bulks, sex ratios were not equal among all pools (Table S5). We performed several comparisons of loci on the sex-associated chromosome compared to loci on other chromosomes, and found no evidence for spurious genetic associations driven by sex-ratios (see next section).

*Tests for the influence of sample sex bias, coverage, and window vs. single SNP analysis in Poolseq data*

To test whether sex ratio differences between sequenced pools (Table S5) yielded spurious genetic associations, we made several comparisons of the behaviour of loci mapping to Chromosome 5 (the chromosome harbouring sex-determining elements). Several lines of evidence suggested that observed genetic associations with emergence time were not confounded by associations with sex. Distributions of the absolute value of allele frequency differences between emergence time bulks had a lower rather than higher median value on Chromosome 5 compared to loci on all other chromosomes (Fig. S6). And, the correlation for the strength of the genetic association with emergence timing in apple vs. haw flies remained when loci on chromosome 5 were removed (With chromosome 5,  $r = 0.96$ , 21,105 SNPs; without chromosome 5,  $r = 0.96$  (20,072 SNPs). Sex ratio, as measured as the proportion of males in a pool, differed by 0.33 and 0.17 between early and late emergence bulks in apple and hawthorn flies, respectively. We also re-calculated the correlation between emergence bulk differences in hawthorn vs.

apple flies using only loci significantly different at  $FDR < 0.05$  AND with the absolute value of the frequency difference greater than 0.40, exceeding the predicted frequency difference if it were driven by sex ratios (note: the correlation calculated in the main manuscript does not use the  $> 0.40$  cutoff). This yielded a correlation coefficient of 0.97 (15,013 SNPs), close to and slightly greater than the calculated correlation coefficient including all loci significant at  $FDR < 0.05$  ( $r = 0.96$ , 21,105 SNPs).

To test for the influence of sequencing depth on our genetic associations, we calculated a modest, negative Pearson correlation between the absolute value of the allele frequency difference between emergence pools (i.e., the strength of the genetic association with emergence timing) and sequencing depth ( $r = -0.13$  for haw flies,  $r = 0.12$  for apple flies). We then tested whether this relationship might influence estimates of allele frequency difference correlations. With a minimum read depth of 60 for both apple and haw flies and an FDR cutoff  $\leq 0.05$  (from the per-locus Fisher test of association with emergence time) we calculated the Pearson correlation between allele frequency difference of emergence pools for apple flies and allele frequency difference of emergence pools for haw flies as  $r = 0.96$  with 26,829 loci, and at min depth  $= 100\times$   $r = 0.96$  with 4,048 loci. This is the same correlation coefficient estimated including all 53,335 SNPs significantly associated ( $FDR \leq 0.05$ ) with emergence time in both apple and haw flies, as reported in the main text. These results suggest that the results are robust to variation in sequencing depth.

Window-based analyses are often used to identify genomic ‘outlier’ regions associated with a trait. This can be an effective approach for loci of major effect, but is less effective for polygenic variation. For example, selection on many loci will lead to comparably weak divergence, spread across many loci [38]. However, we conducted a 500bp, window-based analysis to corroborate our general estimates of allele frequency differences, and to corroborate the overlap between associations with emergence timing in apple flies, associations with emergence timing in haw flies, and host race differences. We used absolute allele frequency differences because we did not have individual genotype variation to infer phase. We estimated the mean absolute allele

frequency difference (between emergence pools and between host races) for each 500bp, non-overlapping window in the genome that contained at least one SNP. These estimates yielded distributions of allele frequency differences with shorter tails compared to the distributions of (absolute value) allele frequency differences for individual SNPs, as predicted, and with slightly higher medians. This is simply a consequence of the right-skewness of the distribution. Overall, the distributions were similar in shape and position, suggesting that individual SNPs provide reasonable estimates (Fig S7). The tails of all distributions exceeded 0.2, in-line with previous observations for individually-sequenced RAD loci [36, 37]. Without any filtering of windows based on frequency difference value, there was a significant correlation between the allele frequency difference between emergence pools in apple flies compared to haw flies ( $r = 0.42$ ,  $p \ll 0.001$ ,  $df = 1524108$ ). This correlation dropped to  $r = 0.21$  ( $p \ll 0.001$ ,  $df = 8976$ ) when considering only loci that surpassed the 0.975 quantile for allele frequency difference between emergence pools in both host races. However, this is to be expected because these data only contain one, extreme end of the distribution because of the outlier cut-off and because there is no directionality included (only absolute values). This would be analogous to calculating a correlation for only the data points in the far, upper right corner of a point cloud on a bivariate plot with a correlation of  $r \sim 1$ . However, Fisher's exact tests revealed that there was substantially and significantly more overlap than expect by chance alone between a) windows in the upper 0.975 quantile for differences between emergence pools for haw flies, and windows in the upper 0.975 quantile for differences between emergence pools for apple flies ( $OR = 15.4$ ,  $p \ll 0.001$ ), and b) windows in the upper 0.975 quantile for differences between emergence pools for either host race and windows in the upper 0.975 quantile for differences between host races (apple emergence vs. host race differences –  $OR = 6.5$ ,  $p \ll 0.001$ ; haw emergence vs. host race differences –  $OR = 6.5$ ,  $p \ll 0.001$ ). Thus, analysis of outlier windows reveals similar patterns to those observed using individual SNPs (see main text).

### Enrichment analyses and potential effects of LD

Because the majority of SNPs associated with emergence timing mapped to regions of chromosomes 1 and 3 with elevated LD, enrichment analyses could be influenced by linkage. We evaluated whether enrichment in the emergence associated loci might be driven by enrichment of genes in the same categories mapping to regions of high and medium LD on chromosomes 1 and 3 as defined in Ragland et al. [37]. Previous studies have mapped a total of 5,079 RAD loci to the 5 main *R. pomonella* chromosomes. A total of 4,908 of these 5,079 RAD loci have been mapped to regions within chromosomes that display high, intermediate, and low levels of LD [35, 37, 39]. We assigned SNP variants that mapped to within 200bp of mapped RAD loci to chromosomes and LD groups. From this set, we compiled a list of SNPs mapping to high and medium LD regions of chromosomes 1 and 3. These represent a roughly random set of loci mapping to those regions, as they are based on restriction site distributions. From that list we compiled a list of flybase gene annotations for the subset of those SNPs mapping to a gene model and submitted the list to the DAVID functional annotation tool (<https://david.ncifcrf.gov/>) for enrichment analysis. We found that the set of all SNPs mapping to high and medium LD regions of chromosomes 1 and 3 were only enriched for a single ‘cluster’ of genes involved in ion transport. Singleton categories of ‘axon’ and ‘neurogenesis’ were very moderately enriched (FDR = 0.08 and 0.3, respectively). In contrast, the set of SNPs and Indels associated with emergence timing (FDR < 0.01 in both apple and haw flies) was highly enriched for many functional category clusters of development and neurogenesis related genes: e.g., ‘axon extension’ and ‘cell growth’ in Annotation Cluster 2, ‘neuron development’ and ‘neuron differentiation’ in Annotation Cluster 4, ‘regulation of cell development’ and ‘regulation of cell morphogenesis’ in Annotation Cluster 14, ‘transcription’ and ‘transcription regulation’ in Annotation Cluster 16, and the unclustered terms ‘developmental protein’ and ‘DNA binding’ (Dataset S3). Thus, the set of SNPs associated with emergence timing are enriched for these functional categories even relative to the background of genes in the high and medium LD regions on chromosomes 1 and 3 (note that many also map to other chromosomes and LD regions).

#### Permutation tests for correlations of genetic associations

We performed several correlation analyses to test for correlations of genome-wide associations with emergence time in the two populations or between associations with emergence timing and host race divergence. Because an individual SNP locus was an experimental unit in these analyses, linkage disequilibrium leads to non-independence among data points. To account for this influence, we compared Pearson correlation coefficients to an empirical null distribution generated by randomly sampling SNPs from the genome that had roughly the same base pair position distribution compared to the focal SNPs. We used an R script (<https://github.com/gjragland/Diverse-developmental-processes-influence-seasonal-phenology/blob/master/permCor.r>) to execute the following steps:

- 1) Identify all focal SNPs co-localizing to the same contig or scaffold as additional focal SNPs, leaving the following groups; all singleton SNPs (the only focal SNP on a contig/scaffold), and groups of SNPs all mapping to the same contig/scaffold.
- 2) For each group of focal SNPs mapping to the same scaffold/contig:
  - a. Calculate the base pair range
  - b. Randomly sample (without replacement) scaffolds/contigs harbouring SNPs with the same or greater base pair range as those of the focal SNPs
  - c. Choose a random starting SNP position that allows for sampling of a requisite number of SNPs (number of focal SNPs in the group) across a base pair range greater than or equal to the range of the focal SNPs in the group.
  - d. Calculate the quantiles of base pair distribution of the focal SNP group.
  - e. Sample the SNPs from the sampled scaffold/contig closest to each quantile of each focal SNP in the group. This step samples SNPs that best approximate the base pair (physical distance) distribution of the focal SNPs in the group.
- 3) For each singleton SNP, randomly sample one SNP from one scaffold/contig (without replacement) from the pool of scaffolds/contigs not sampled in step 2.

- 4) Calculate the Pearson correlation coefficient using data from the randomly sampled SNPs. For example, if the comparison were the strength of the genetic association with emergence time between apple and haw flies, the two vectors would be the allele frequency differences between early and late pools at the sampled SNPs for apple and haw flies.
- 5) Repeat steps 1) through 4) for 5,000 iterations, generating the empirical null distribution.
- 6) Compare the point estimate of the Pearson correlation against the permuted distribution to determine a p-value.

#### Permutation tests for cis regulatory variation

We tested whether SNPs associated with either emergence timing or host race differences tended to localize to within 5kb of the Transcription Start Site (TSS) of a gene significantly differentially expressed (DE) either over time or between host races, depending on the comparison, more often than expected by chance. The point estimate,  $n$ , was simply the number of SNPs within 5kb of a TSS of a DE gene. We generated the null distribution by randomly sampling  $n$  coordinates from the genome, counting the number within 5kb of a TSS if a DE gene, then repeating 10,000 times. We then compared the point estimate against the permuted distribution to determine a p-value. The associated R script is available here: <https://github.com/giragland/Diverse-developmental-processes-influence-seasonal-phenology/blob/master/TSS.r>

#### Tajima's $D$ analyses for selection on host race differences

To test for evidence of selective sweeps we estimated Tajima's  $D$  in a sliding windows analysis (step 500 and window 1kb) across the whole genome for the randomly sampled pools of apple and haw flies using PoPoolation with a minimum base quality score of 30, minimum minor allele count of 2, and minimum coverage of 20 reads [40]. We used small windows because larger windows would have exceeded the size of

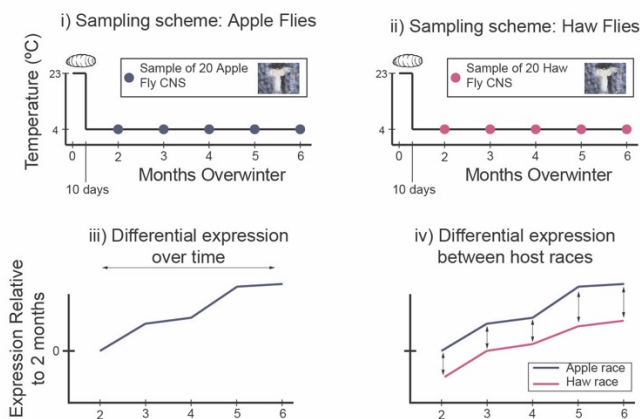


many scaffolds in the tail of the scaffold size distribution for the *R. zephyria* genome assembly.

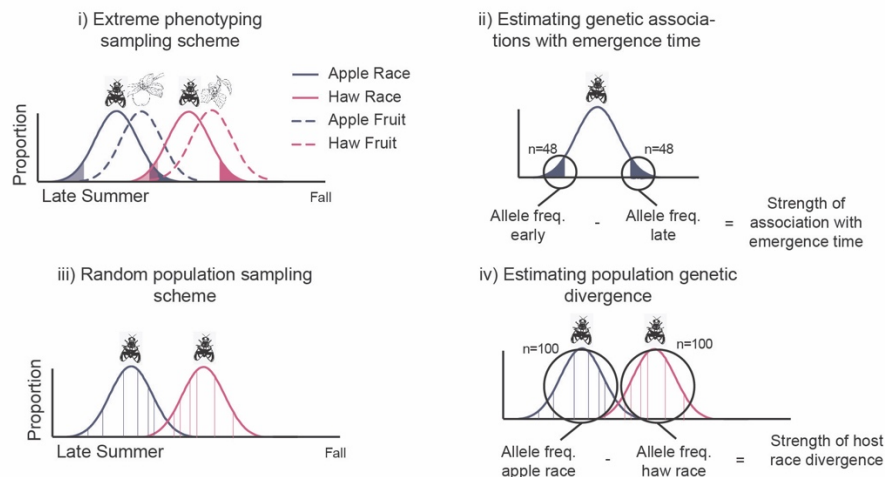
## Supplementary Figures

**Figure S1.** Conceptual diagram of the RNAseq and Poolseq experiments. **A)** For RNAseq, we reared 10-day post-pupariation pupae at 4°C, and sampled pools of 20 CNS at 2, 3, 4, 5, and 6 months for both (i) apple-infesting and (ii) haw-infesting populations. We estimated differential expression over time at months 3 – 6 relative to expression at 2 months (iii) in each host race, and estimated differential expression between host races at each sampled month (iv). **B)** For Poolseq, we took both an extreme phenotyping and random population sample approach. (i) Apple flies emerge on average ~3 weeks earlier than haw flies to track the earlier fruiting time of apple trees. We sampled the 48 earliest and 48 latest emerging flies for both apple and haw-infesting populations (see main paper for details), and (ii) estimated frequency differences between these pools as a metric of the strength of genetic association with emergence time. (iii) For population comparisons, we took random samples of 100 flies across the entire emergence distribution for both apple and haw-infesting flies, then (iv) estimated population divergence as the magnitude of the allele frequency difference between the apple fly and haw fly pools.

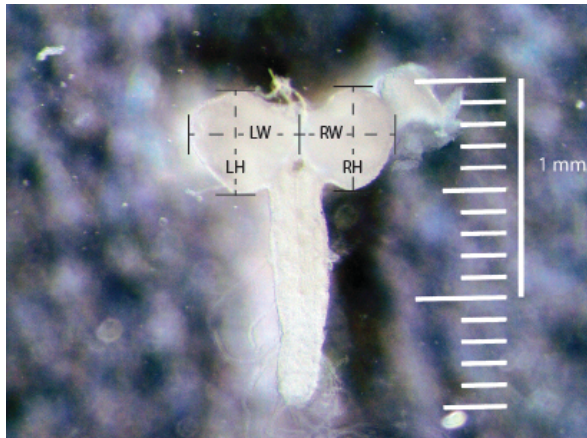
### A. RNAseq experimental design



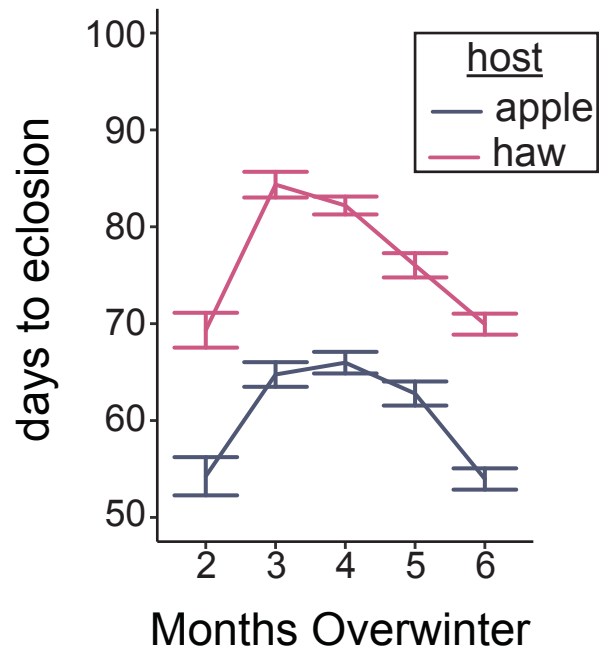
### B. Poolseq experimental design



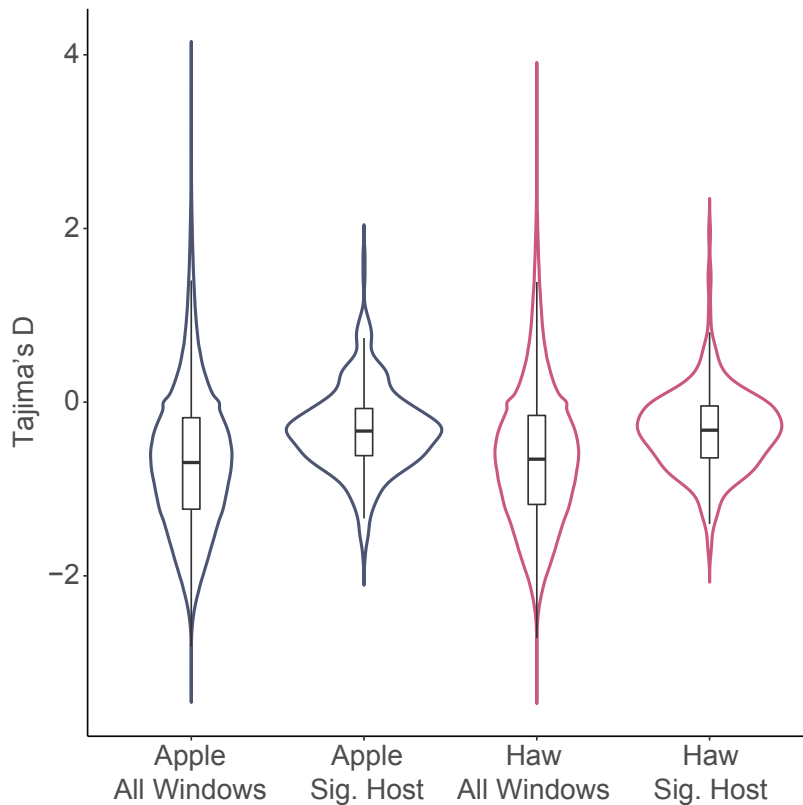
**Figure S2.** Representative brain from a 20 week old diapausing pupa, including dimensions for measurements of the anterior-posterior height of the left (LH) and right (RH) brain hemispheres, and the lateral width of each hemisphere (LW and RW). The elongated structure below the hemispheres is the fused, ventral nerve cord.



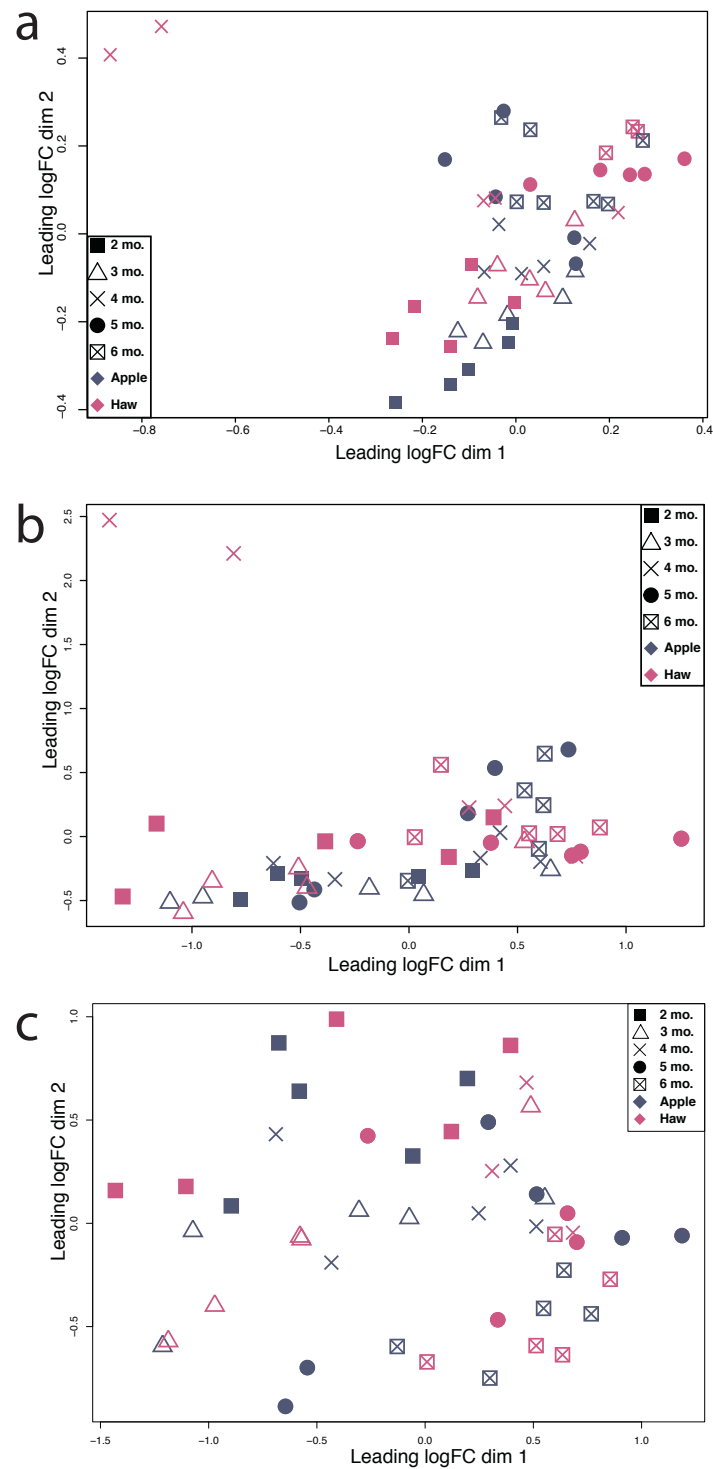
**Figure S3.** Emergence timing phenotypes vary between the two host races sampled from Urbana, IL for 2 – 6 months overwinter treatments. After mortality, median sample size per population per treatment = 78, range = 61 - 86.



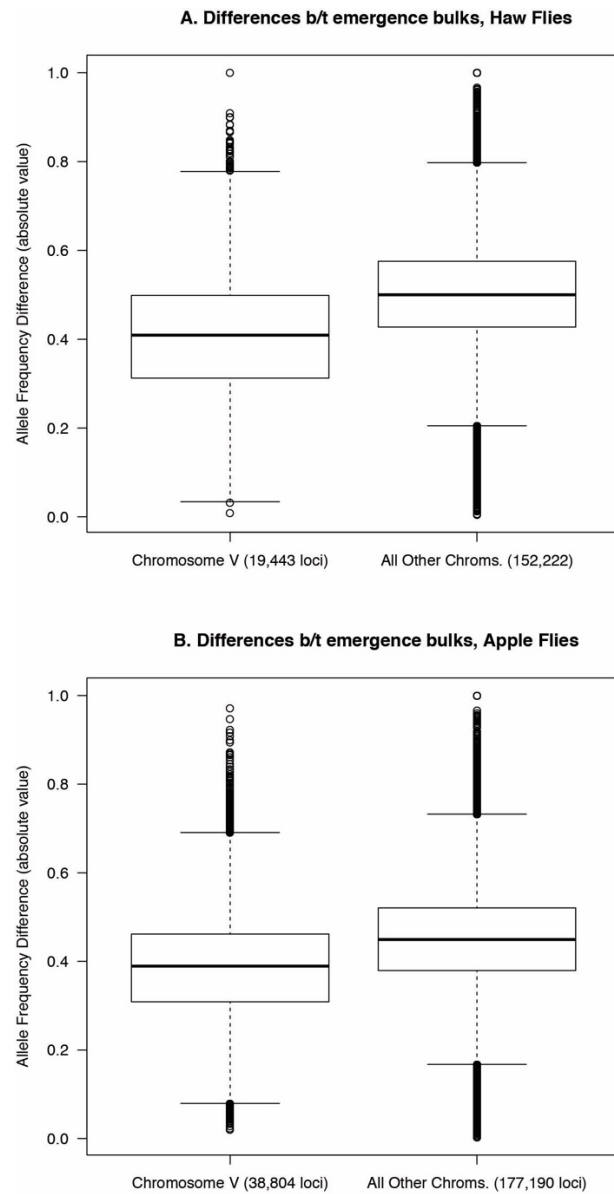
**Figure S4.** Distributions of Tajima's D values (1kb sliding windows, 500bp increment) across the genome ('All Windows';  $n = 2,144,514$  1kb windows) and for windows surrounding SNPs significantly associated with host race divergence ( $FDR < 0.05$ ;  $n = 456$  1kb windows) calculated from the pool of randomly sampled apple (blue violins) and haw (pink violins) flies. Low values of Tajima's D, indicating low nucleotide diversity, are associated with selective sweeps and other demographic processes (e.g., population growth). However, windows containing significant SNPs on average showed slightly elevated Tajima's D values relative to genomic background, providing no evidence for selective sweeps occurring in highly differentiated regions.



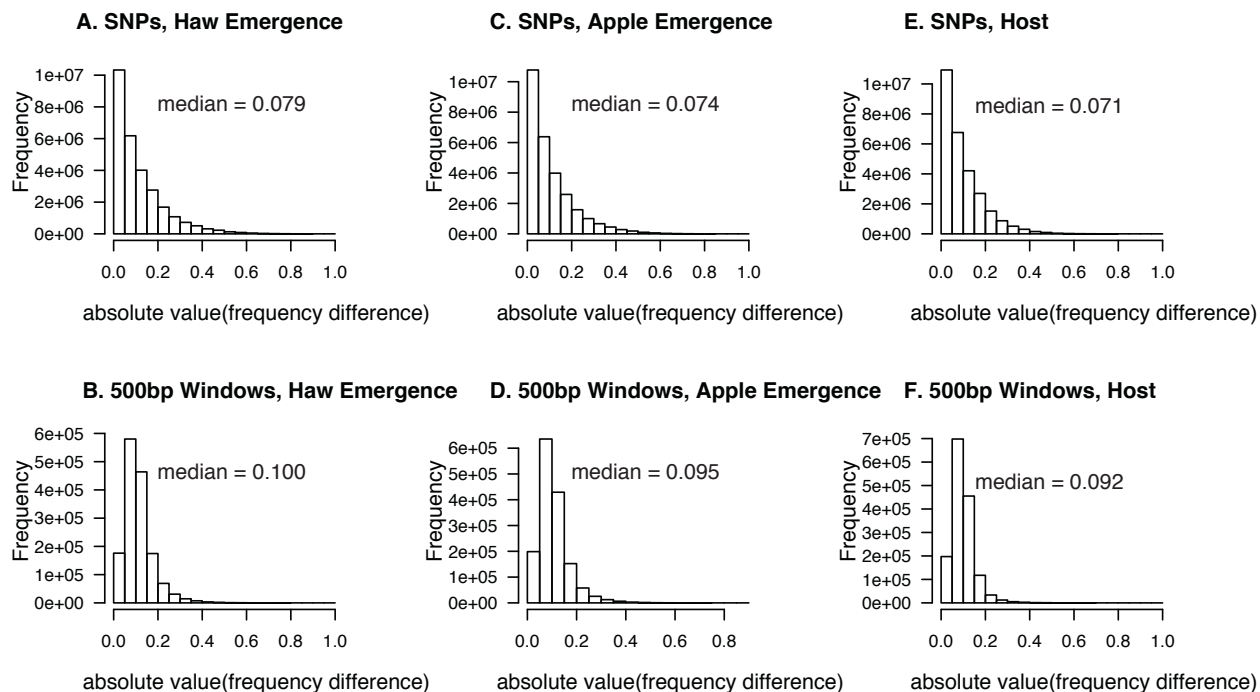
**Figure S5.** MDS plot of RNAseq data showing two clear outliers, haw fly 4-month samples (upper left), in the analysis of the full data set (a) and of only the top 500 DE transcripts, as recommended in the edgeR manual (b). Panel c) shows the MDS plot after outlier removal for the top 500 DE transcripts.



**Figure S6.** Comparison of the magnitude of association with emergence (absolute value of the allele frequency difference between early and late emergence bulks) for loci on Chromosome 5 vs. loci on all other chromosomes for A) hawthorn flies and B) apple flies. This comparison includes only loci that could be mapped to chromosomes and were significantly associated with emergence time ( $FDR < 0.05$ ).



**Figure S7.** Distributions of the absolute value of allele frequency differences between emergence time pools within host races and between host races for the mean of 500bp windows (top row) and for individual SNP estimates (bottom row).





## Supplementary Tables

**Table S1.** T-tests of parameter estimates from a linear regression of the average height of the left and right brain hemispheres on body mass and month sampled (fitted using 'lm' in R). A) Left and B) right hemispheres. Note that the 'Months' term is non-significant in both models, suggesting that brain size as measured does not change appreciably over time.

A)

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.35	0.053	6.5	9.3e-08
Months	0.00082	0.0040	0.20	0.84
mass	0.015	0.0048	3.15	0.0031

B)

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	0.49	0.088	5.6	1.7e-06 ***
Months	-0.0086	0.0066	-1.3	0.21
mass	0.0016	0.0079	0.20	0.84

**Table S2.** Counts, Odds ratios (OR) and p-values for tests for enrichment of SNPs significantly associated with emergence timing from a RADseq study of sympatric populations at Fennville, MI [37] in the set of SNPs significantly associated with emergence timing in this study (flies from Urbana, IL). The OR values are the odds of a significant SNP in the Urbana study if the SNP was also significant in the Fennville study over the odds of a significant SNP in the Urbana study if the SNP was not significant in the Fennville study. This test was performed for two data sets: SNPs associated with emergence timing in apple in both studies (A) and in haw in both studies (B), using the fisher.test function in R.

<b>A. Apple</b>	Significant Urbana	Non-Significant Urbana	OR	p-val
Significant Fennville	117	1086	5.3	< 2.2e-16
Non-significant Urbana	144	7141		
<b>B. Haw</b>	Significant Urbana	Non-Significant Urbana	OR	p-val
Significant Fennville	136	1507	8.1	< 2.2e-16
Non-significant Urbana	75	6766		

**Table S3.** Counts, Odds ratio (OR), and p-value for the test for enrichment of genes harboring SNPs associated with increased development time in *Drosophila melanogaster* from [41] (FDR < 0.01) in the set of genes from the *Rhagoletis zephyria* assembly harboring SNPs or Indels associated with emergence time in both host races of *R. pomonella* (FDR < 0.01 in both host races). The *D. melanogaster* data were processed as described for *R. pomonella* SNP calling and allele frequency estimation, except that *D. melanogaster* SNPs were filtered if read depth did not exceed 10x in both the selected and control line.

	Significant <i>R. pom.</i>	Non-Significant <i>R. pom.</i>	OR	p-val
Significant <i>D. mel.</i>	152	682	1.5	5.5e-05
Non-significant <i>D. mel.</i>	1183	7870		

**Table S4.** Counts, Odds ratios (OR) and p-values for tests for enrichment of SNPs significantly associated with emergence timing in the set of SNPs significantly associated with host race divergence. We conducted two separate tests for SNPs associated with emergence time in apple (A) and haw (B) host races. All tests were performed using the fisher.test function in R.

<b>A. Apple</b>	Significant Emergence	Non-Significant Emergence	OR	p-val
Significant Host Race	473	459,090	2.5	< 2.2e-16
Non-significant Host Race	11,354	27,662,938		
<b>B. Haw</b>	Significant Emergence	Non-Significant Emergence	OR	p-val
Significant Host Race	549	585,911	2.3	< 2.2e-16
Non-significant Host Race	11,278	27,536,117		

**Table S5.** Numbers of males/females and sex ratios for each of the six sequenced pools of flies. Males emerge later than females, and thus were disproportionately represented in the late emerging pools.

<b>Poolseq pool</b>	<b>Number males</b>	<b>Number females</b>	<b>Proportion male</b>
Apple late	38	10	0.79
Apple early	22	26	0.46
Apple random	54	44	0.56
Haw late	36	12	0.75
Haw early	28	20	0.58
Haw random	58	42	0.58

## Legends for Datasets S1 to S5

### Legend for **Dataset S1**

Functional enrichment analysis of transcripts differentially expressed across time, as performed using the DAVID online tool (see methods). Excel file containing all flybase IDs for flybase-annotated transcripts significantly (FDR 0.05) differentially expressed between at least two time points in both apple and haw flies, and functional enrichment statistics from DAVID. Submitted lists include the full list of all IDs, and the list of IDs falling into each of 6 co-regulated modules. Both clustered and un-clustered categories [42] are reported for each list.

### Legend for **Dataset S2**

Functional enrichment analysis of transcripts differentially expressed between host races, as performed using the DAVID online tool (see methods). Excel file containing all flybase IDs for flybase-annotated transcripts significantly (FDR 0.05) differentially expressed between at least two time points in both apple and haw flies, and functional enrichment statistics from DAVID. Submitted lists include the full list of all IDs, and the list of IDs falling into each of 7 co-regulated modules. Both clustered and un-clustered categories are reported for each list.

### Legend for **Dataset S3**

Functional enrichment analysis of variants (SNP and indels) significantly (FDR < 0.01) associated with emergence time in both apple and haw flies. Excel file containing all flybase IDs of flybase annotations for SNPs/Indels designated as upstream variants, downstream variants, or variants within a gene model by SnpEff [43], and functional enrichment statistics from DAVID. Both clustered and un-clustered categories are reported.

### Legend for **Dataset S4**

Functional enrichment analysis of variants (SNP and indels) either a) significantly (FDR < 0.05) associated with emergence time in apple flies and with host race genetic divergence, or b) significantly (FDR < 0.05) associated with emergence time in haw flies and with host race genetic divergence. Excel file containing all flybase IDs of flybase annotations for SNPs/Indels designated as upstream variants, downstream variants, or variants within a gene model by SnpEff, and functional enrichment statistics from DAVID. Both clustered and un-clustered categories are reported.

### Legend for **Dataset S5**

Mapping statistics for Poolseq data. Statistics vary slightly depending on method but coverage depth was always >20.

## References

1. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data*. Bioinformatics, 2014. **30**(15): p. 2114-2120.
2. Kelley, D.R., M.C. Schatz, and S.L. Salzberg, *Quake: quality-aware detection and correction of sequencing errors*. Genome Biology, 2010. **11**(11): p. R116.
3. Xie, Y., et al., *SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads*. Bioinformatics, 2014. **30**(12): p. 1660-1666.
4. Luo, R., et al., *SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler*. GigaScience, 2012. **1**(1): p. 18.
5. English, A.C., W.J. Salerno, and J.G. Reid, *PBHoney: identifying genomic variants via long-read discordance and interrupted mapping*. BMC bioinformatics, 2014. **15**(1): p. 180.
6. Meyers, P.J., et al., *Divergence of the diapause transcriptome in apple maggot flies: winter regulation and post-winter transcriptional repression*. Journal of Experimental Biology, 2016. **219**(17): p. 2613-2622.
7. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST*. Bioinformatics, 2010. **26**(19): p. 2460-2461.
8. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data without a reference genome*. Nature biotechnology, 2011. **29**(7): p. 644.
9. Kent, W.J., *BLAT—the BLAST-like alignment tool*. Genome research, 2002. **12**(4): p. 656-664.
10. Xue, W., et al., *L\_RNA\_scaffolder: scaffolding genomes with transcripts*. BMC genomics, 2013. **14**(1): p. 604.
11. Simão, F.A., et al., *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs*. Bioinformatics, 2015. **31**(19): p. 3210-3212.
12. Ragland, G.J., et al., *Biphasic metabolic rate trajectory of pupal diapause termination and post-diapause development in a tephritid fly*. Journal of Insect Physiology, 2009. **55**(4): p. 344-350.
13. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner*. Bioinformatics, 2013. **29**(1): p. 15-21.
14. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome*. BMC bioinformatics, 2011. **12**(1): p. 323.
15. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features*. Bioinformatics, 2013. **30**(7): p. 923-930.
16. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics, 2010. **26**(1): p. 139-140.
17. Robinson, M.D. and A. Oshlack, *A scaling normalization method for differential expression analysis of RNA-seq data*. Genome biology, 2010. **11**(3): p. R25.
18. Ernst, J. and Z. Bar-Joseph, *STEM: a tool for the analysis of short time series gene expression data*. BMC bioinformatics, 2006. **7**(1): p. 191.
19. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis*. BMC bioinformatics, 2008. **9**(1): p. 559.

20. Boller, E. and R. Prokopy, *Bionomics and management of Rhagoletis*. Annual review of entomology, 1976. **21**(1): p. 223-246.
21. Dambroski, H.R. and J.L. Feder, *Host plant and latitude-related diapause variation in Rhagoletis pomonella: a test for multifaceted life history adaptation on different stages of diapause development*. Journal of Evolutionary Biology, 2007. **20**: p. 2101-2112.
22. Ragland, G.J., et al., *Environmental interactions during host race formation: host fruit environment moderates a seasonal shift in phenology in host races of Rhagoletis pomonella*. Functional Ecology, 2012. **26**(4): p. 921-931.
23. Bainbridge, S.P. and M. Bownes, *Staging the metamorphosis of Drosophila melanogaster*. Development, 1981. **66**(1): p. 57-80.
24. Rabinovich, D., O. Mayseless, and O. Schuldiner, *Long term ex vivo culturing of Drosophila brain as a method to live image pupal brains: insights into the cellular mechanisms of neuronal remodeling*. Frontiers in cellular neuroscience, 2015. **9**: p. 327.
25. Feder, J.L., et al., *The effects of winter length on the genetics of apple and hawthorn races of Rhagoletis pomonella (Diptera: Tephritidae)*. Evolution, 1997. **51**(6): p. 1862-1876.
26. Hodek, I., *Controversial aspects of diapause development*. European Journal of Entomology, 2013. **99**(2): p. 163-173.
27. Denlinger, D., *Insect diapause: dynamics of hormone sensitivity and vulnerability to environmental stress*, in *Endocrinological Frontiers in Physiological Insect Ecology*, F. Sehnal, A. Zabaza, and D.L. Denlinger, Editors. 1988, Wroclaw Tech. Univ. Press: Wroclaw, Poland. p. 309-324.
28. Denlinger, D., et al., *Hormonal control of diapause*. Comprehensive Insect Molecular Science, 2004. **3**: p. 615-650.
29. Calvert, M.B., et al., *The genomics of trait combinations and their influence on adaptive divergence*. bioRxiv, 2020: p. 2020.06.19.161539.
30. Powell, T.H., et al., *On the scent of standing variation for speciation: behavioral evidence for native sympatric host races of Rhagoletis pomonella (Diptera: Tephritidae) in the southern United States*. Evolution: International Journal of Organic Evolution, 2012. **66**(9): p. 2739-2756.
31. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads*. EMBnet. journal, 2011. **17**(1): p. 10-12.
32. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform*. bioinformatics, 2009. **25**(14): p. 1754-1760.
33. McKenna, A., et al., *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data*. Genome research, 2010. **20**(9): p. 1297-1303.
34. Korneliussen, T.S., A. Albrechtsen, and R. Nielsen, *ANGSD: analysis of next generation sequencing data*. BMC bioinformatics, 2014. **15**(1): p. 356.
35. Doellman, M., et al., *Genomic differentiation during speciation-with-gene-flow: comparing geographic and host-related variation in divergent life history adaptation in Rhagoletis pomonella*. Genes, 2018. **9**(5): p. 262.
36. Egan, S.P., et al., *Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow*. Ecology letters, 2015. **18**(8): p. 817-825.



37. Ragland, G.J., et al., *A test of genomic modularity among life-history adaptations promoting speciation with gene flow*. Molecular ecology, 2017. **26**(15): p. 3926-3942.
38. Pritchard, J.K. and A. Di Rienzo, *Adaptation—not by sweeps alone*. Nature Reviews Genetics, 2010. **11**(10): p. 665-667.
39. Doellman, M.M., et al., *Standing geographic variation in eclosion time and the genomics of host race formation in Rhagoletis pomonella fruit flies*. Ecology and evolution, 2019. **9**(1): p. 393-409.
40. Kofler, R., et al., *PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals*. PloS one, 2011. **6**(1): p. e15925.
41. Burke, M.K., et al., *Genome-wide analysis of a long-term evolution experiment with Drosophila*. Nature, 2010. **467**(7315): p. 587.
42. Huang, D.W., et al., *DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists*. Nucleic acids research, 2007. **35**(suppl\_2): p. W169-W175.
43. Cingolani, P., et al., *A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3*. Fly, 2012. **6**(2): p. 80-92.