Report

Current Biology

Tissue-Specific *cis*-Regulatory Divergence Implicates *eloF* in Inhibiting Interspecies Mating in *Drosophila*

Highlights

- RNA-seq on *Drosophila* hybrids identifies tissue-specific *cis*-regulatory changes
- *eloF*, a fatty acid elongase, is a major determinant of hydrocarbon pheromones
- Ablation of *eloF* increases interspecies mating nearly to intraspecies levels
- This approach can be applied broadly to rapidly pinpoint key evolutionary drivers

Authors

Peter A. Combs, Joshua J. Krupp, Neil M. Khosla, Dennis Bua, Dmitri A. Petrov, Joel D. Levine, Hunter B. Fraser

Correspondence

pcombs@stanford.edu (P.A.C.), hbfraser@stanford.edu (H.B.F.)

In Brief

Combs et al. use RNA-seq in hybrid *Drosophila* to identify genes related to cuticular hydrocarbon (CHC) differences, including *eloF*. Ablation of *eloF* alters CHC profiles and increases interspecies mating nearly to intraspecies levels. This approach can be applied broadly to rapidly pinpoint key evolutionary drivers in a wide range of species.

Combs et al., 2018, Current Biology *28*, 1–7
December 17, 2018
https://doi.org/10.1016/j.cub.2018.10.036

CellPress

Please cite this article in press as: Combs et al., Tissue-Specific *cis*-Regulatory Divergence Implicates *eloF* in Inhibiting Interspecies Mating in *Drosophila*, Current Biology (2018), https://doi.org/10.1016/j.cub.2018.10.036

Current Biology
Report

Tissue-Specific *cis*-Regulatory Divergence Implicates *eloF* in Inhibiting Interspecies Mating in *Drosophila*

Peter A. Combs,[1,*] Joshua J. Krupp,[2] Neil M. Khosla,[1] Dennis Bua,[1] Dmitri A. Petrov,[1] Joel D. Levine,[2] and Hunter B. Fraser[1,3,*]
[1]Department of Biology, Stanford University, Stanford, CA 94305, USA
[2]Department of Biology, University of Toronto, Mississauga, Mississauga, ON L5L 1C6, Canada
[3]Lead Contact
*Correspondence: pcombs@stanford.edu (P.A.C.), hbfraser@stanford.edu (H.B.F.)
https://doi.org/10.1016/j.cub.2018.10.036

SUMMARY

Reproductive isolation is a key component of speciation. In many insects, a major driver of this isolation is cuticular hydrocarbon pheromones, which help to identify potential intraspecific mates [1–3]. When the distributions of related species overlap, there may be strong selection on mate choice for intraspecific partners [4–9] because interspecific hybridization carries significant fitness costs [10]. *Drosophila* has been a key model for the investigation of reproductive isolation; although both male and female mate choices have been extensively investigated [6, 11–16], the genes underlying species recognition remain largely unknown. To explore the molecular mechanisms underlying *Drosophila* speciation, we measured tissue-specific *cis*-regulatory divergence using RNA sequencing (RNA-seq) in *D. simulans* × *D. sechellia* hybrids. By focusing on *cis*-regulatory changes specific to female oenocytes, the tissue that produces cuticular hydrocarbons, we rapidly identified a small number of candidate genes. We found that one of these, the fatty acid elongase *eloF*, broadly affects the hydrocarbons present on *D. sechellia* and *D. melanogaster* females, as well as the propensity of *D. simulans* males to mate with them. Therefore, *cis*-regulatory changes in *eloF* may be a major driver in the sexual isolation of *D. simulans* from multiple other species. Our RNA-seq approach proved to be far more efficient than quantitative trait locus (QTL) mapping in identifying candidate genes; the same framework can be used to pinpoint candidate drivers of *cis*-regulatory divergence in traits differing between any interfertile species.

RESULTS AND DISCUSSION

D. simulans and *D. sechellia* are closely related sister species, which have accumulated ~0.018 substitutions per site in the several hundred thousand years since they last shared a common ancestor [17, 18]. They are believed to have diverged in allopatry [19], though currently, their ranges overlap and hybrids can be found in the wild [20]. In laboratory conditions, *D. sechellia* males will readily mate with *D. simulans* females, producing sterile male and fertile female hybrid offspring, although in the reciprocal cross, *D. simulans* males will not readily mate with *D. sechellia* females [21]. Male mate choice has been estimated to account for over 70% of the reproductive isolation between these species [16], but the gene(s) accounting for this isolation are unknown. This divergence in male mate choice likely involves species-specific differences in female cuticular hydrocarbons (CHCs), as well as in the male response to these CHCs; in this work, we focused on female CHCs.

Allele-Specific Expression Identifies Fatty Acid Elongases as a Major Differentiator between *D. simulans* and *D. sechellia* Female Oenocytes

Although quantitative trait loci (QTLs) affecting CHCs have been mapped, these contain many CHC-related genes [22–24] and fine-mapping has not been reported. As a complementary approach, we reasoned that genes responsible for major changes in female CHCs may share three key characteristics: (1) *cis*-regulatory divergence in female oenocytes; (2) female-specific expression; and (3) oenocyte-specific expression. Although these are not required, any genes meeting all three criteria would be promising candidates.

cis-regulatory divergence can be measured genome-wide via high-throughput sequencing of RNA (RNA-seq) in interspecific hybrids. Hybrids are required because comparisons between species reflect both *cis*- and *trans*-acting changes; measuring allele-specific expression (ASE) in F1 hybrids controls for *trans*-acting changes, because each allele experiences the same *trans*-regulatory environment within the hybrid nuclei [25]. Thus, differential expression of the two alleles in a hybrid can only be explained by *cis*-regulatory divergence.

In order to identify genes with *cis*-regulatory divergence specific to female oenocytes, we constructed RNA-seq libraries from hybrid *D. simulans* × *D. sechellia* tissues (Figure 1A; see STAR Methods). To measure female specificity, we included samples from both male and female oenocytes, and to measure oenocyte specificity, we included samples from male and female fat bodies (an adjacent non-CHC-producing tissue) [26]. We

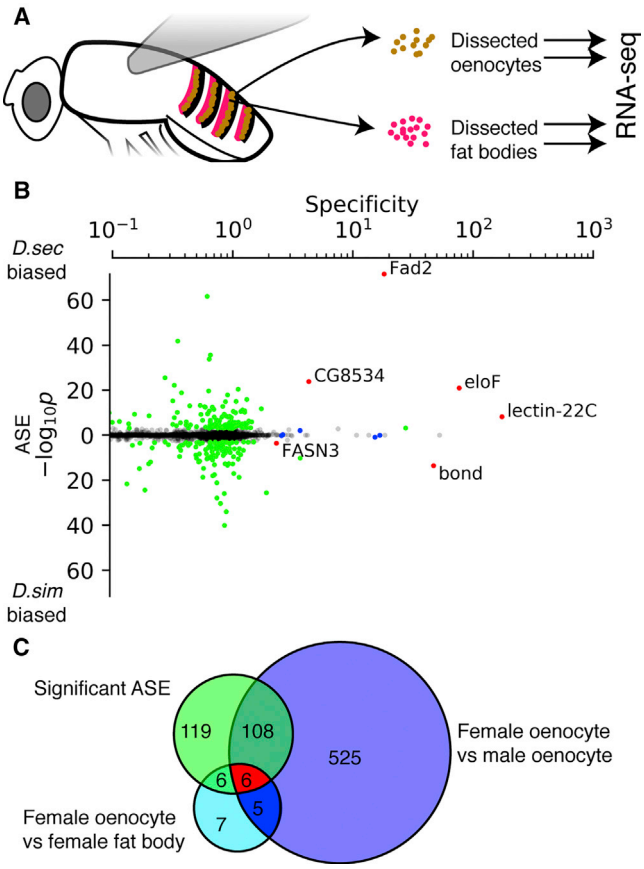

Figure 1. RNA-Seq of Oenocytes and Fat Bodies from Hybrid *D. simulans* × *D. sechellia* Flies Reveals a Strong *cis*-Regulatory Component of CHC Production

(A) We dissected oenocytes (brown) and fat bodies (pink) from hybrid *D. simulans* × *D. sechellia* males and females and performed RNA sequencing.

(B) Genes are plotted by specificity of expression to female oenocytes (x axis; mean of female oenocyte expression divided by maximum expression in female fat bodies, male oenocytes, and female oenocytes) and allele-specific expression p value (y axis). Green dots indicate genes with significant ASE compared to the distribution of reads in the female oenocytes, blue dots indicate those that have significantly higher expression in female oenocytes compared to female fat bodies and male oenocytes, and red dots indicate genes with both tissue-specific and species-specific expression.

(C) Overlap of genes with ASE in female oenocytes (green circle) and differential expression in female oenocytes compared to other tissues (blue and cyan circles).

See also Figure S1 and Table S1.

estimated the significance of each gene's ASE using a negative-binomial test [27] for deviation from the average fraction of *D. sechellia* reads in a given sample.

Even at a stringent cutoff, we identified 239 genes with significant (negative binomial q value < 0.001) ASE in female oenocytes. This is not surprising, because various *Drosophila* interspecific hybrids have also yielded large numbers of genes with strong ASE [28, 29]. Of the 239 significant genes, 27 were annotated with the Gene Ontology term "fatty acid biosynthetic process" (GO: 0006633). Therefore, we concluded that, even when combined with Gene Ontology (GO) annotations, ASE in female oenocytes was insufficient to identify a manageable number of CHC-related candidate genes.

We reasoned that, in addition to ASE, genes important to female CHC differences between *D. simulans* and *D. sechellia* would likely be expressed specifically in female oenocytes. To identify candidate genes, we looked for genes that had significantly higher expression in the female oenocytes compared to male oenocytes and to female fat bodies (Figures 1B and 1C; Sleuth q value < 0.001 for both comparisons) [30]. Only six genes passed these cutoffs (Table 1). Reassuringly, one of these was *desatF* (also known as *Fad2*), a fatty acid desaturase that is known to be expressed in *D. sechellia* female oenocytes, but not in males or in *D. simulans* [33]; another was *eloF*, which has been previously observed to lack expression in *D. simulans*, although its expression in *D. sechellia* has not been studied [34].

Among the six candidate genes, the only enriched GO molecular function terms were related to "fatty acid elongase activity" (GO: 0009922 and its parent GO terms), which describe the three genes *eloF*, *CG8534*, and *bond* (in all cases, we use the names of the *D. melanogaster* orthologs) [35]. All three of these have ELO family fatty acid elongase domains [36]. Both *eloF* and *CG8534* were *D. sechellia* biased, and *bond* was *D. simulans* biased. We further detected a weak signal for *FASN3*, a putative acyl transferase (Table 1). No other gene that is both oenocyte and species specific in its expression has an annotated GO term or protein domain that is clearly related to CHC production (Table 1).

Compared to the female oenocytes, the other three tissues we profiled (male oenocyte, male fat body, and female fat body) all had a much weaker signal of ASE among genes with sex- and tissue-specific expression (Figures S1A–S1D). Therefore, we chose to focus on changes in female CHC production that might drive speciation.

eloF Has Widespread Effects on the Hydrocarbon Profile of *D. sechellia* and *D. melanogaster*

To explore the role of our candidate genes on CHC profiles of these species, we performed gas chromatography coupled to mass spectrometry (GCMS). Consistent with previous measurements of hydrocarbon profiles of *Drosophila*, we found that wild-type *D. simulans* has more short-chain hydrocarbons than *D. sechellia* (Figure 2A) [37]. In particular, *D. sechellia* has almost no 23-carbon CHCs, and the predominant *D. simulans* hydrocarbon is 7-tricosene, a 23-carbon monoene. Indeed, there was only one hydrocarbon shorter than 26 carbons with a greater representation in *D. sechellia* than *D. simulans*, the 25-carbon pentacosadiene (~2-fold higher in *D. sechellia*). In contrast, all nine CHC peaks longer than 26 carbons were more abundant in *D. sechellia* than in *D. simulans.*

To explore the effects of our candidate genes on CHC profiles, we studied the phenotypic effects of their RNAi knockdowns in *D. melanogaster.* We did not pursue *desatF*, which already has a well-established role in *Drosophila* speciation [33, 38, 39]; *FASN3*, which is essential for viability [40]; or *lectin-22C*, which has relatively weak ASE and no obvious connection to CHC production. For the remaining three CHC-related candidates, we created RNAi knockdowns in *D. melanogaster* females for each of these genes specifically in oenocytes by crossing

Table 1. Genes with Female Oenocyte- and Species-Specific Expression

Gene	Female Oenocyte Specificity (Oenocyte/ Female Sleuth q Value)	% *D. sechellia* Reads in Female Oenocytes (Negative Binomial p Value)	GO Term(s)	Protein Domain(s)
eloF	76.6 (4.5e−9/2.3e−5)	98.75% (1.2e−21)	fatty acid elongase activity	ELO family
Fad2	18.2 (4.5e−9/2.5e−5)	95.5% (2.5e−72)	catalysis of an oxidation-reduction (redox) reaction in which hydrogen or electrons are transferred from each of two donors…	fatty acid desaturase type 1, conserved site; fatty acid desaturase domain; acyl-CoA desaturase
CG8534	4.3 (8.2e−4/3.6e−5)	93.75% (1.6e−24)	fatty acid elongase activity	ELO family
FASN3	2.31 (2.7e−4/3.8e−5)	60.5% (0.00026)	fatty acid synthase activity; hydrolase activity, acting on ester bonds	ketoacyl synthase (N-terminal, C-terminal, and C-terminal extension), acyl transferase, polyketide synthase, and alcohol dehydrogenase C-terminal
lectin-22C	173.9 (3.0e−4/2.0e−4)	85.75% (6.7e−09)	galactose binding	C-type lectin-like/link domain superfamily
bond	47.0 (7.8e−4/6.3e−5)	25.5% (2.8e−14)	fatty acid elongase activity	ELO family

Genes with significant tissue-specific (sleuth q value < 0.001 in comparisons both between the two female tissues and between the two oenocyte samples) and species-specific expression (negative binomial q value < 0.001). Specificity is the ratio of the mean expression in female oenocytes to the highest expression among male oenocytes, female fat bodies, and male fat bodies. Gene ontology (GO) terms are annotated molecular function terms (see Table S2 for citations). GO terms without experimental evidence are in italics. Protein domains are InterPro annotated protein domains and/or motifs as listed on FlyBase v2017_06 [31, 32]. See also Table S2. CoA, coenzyme A.

PromE(800)-gal4 males with upstream activating sequence (UAS)-short hairpin RNA (shRNA) females from the TRiP project [41, 42] and then screened the CHC profiles of the progeny by GCMS.

Of our three candidate genes (*CG8534*, *bond*, and *eloF*), we found that one (*CG8534*) was essential for viability, and knockdown of our second candidate (*bond*) in females led to ~60% increases in levels of pentacosadiene (a 25-carbon hydrocarbon) and ~60% decrease in levels of heptacosadiene (27 carbon; Figure S2A). However, other hydrocarbons were not significantly affected.

We observed the most pronounced effects for RNAi knockdown of our third candidate, *eloF*. We found that female flies with *eloF* knocked down had significantly fewer long-chain CHCs and more short-chain CHCs than wild-type flies (>3-fold change between CHCs with longer versus shorter than 26 carbons; Figure 2B), consistent with previous work [34]. Interestingly, *eloF* also had the strongest ASE among the six candidate genes (79-fold higher expression from *D. sechellia* alleles).

To examine the effect of *eloF* on CHCs in *D. sechellia*, we used CRISPR/Cas9 genome editing to create two independent lines of *D. sechellia* with *eloF* knocked out (Figure S3). Nearly all of the CHCs whose levels changed after *eloF* knockdown in *D. melanogaster* showed a similar difference in *D. sechellia* (Figure 2C). Thus, we conclude that the molecular substrates and products of *eloF* are similar between *D. melanogaster* and *D. sechellia.*

We noticed that there was a strong correlation between the CHC changes observed between the sister species *D. simulans* and *D. sechellia* and the changes between wild-type and *eloF* knockout *D. sechellia* females (Figure 2D). We found a similar result for *eloF*-depleted *D. melanogaster* females (Figure S2B). These correlations suggest that loss of *eloF* phenocopies the natural interspecific divergence in CHC profiles.

To visualize entire CHC profiles, we performed principal-components analysis, which showed that 94% of the total variation was captured by the first two components. The first principal component of variation separated *D. simulans* from both *D. melanogaster* and *D. sechellia* (Figure 2E). Although knockdown or knockout of *eloF* did not completely transform the profiles of either species to match *D. simulans*, it did shift the profiles significantly closer. Because the CHC profile after knockdown is much more similar to *D. simulans*, whose CHCs inhibit mating with *D. simulans* males, we reasoned that one or more of the CHCs produced by *eloF* may contribute to this inhibition of interspecies mating.

eloF Is Necessary for Inhibition of Interspecies Mating

To determine whether the change in *eloF* expression (and concomitant CHC changes) could be responsible for sexual isolation between the species, we performed mate choice assays by video recording and noting the time of various mating behaviors (Figures 3A–3C). We first tested whether *eloF* might drive the behavioral isolation of *D. simulans* and *D. sechellia*. As expected, *D. simulans* males courted wild-type *D. sechellia* females at a significantly lower rate than *D. simulans* females. Remarkably, *D. simulans* males courted *eloF- D. sechellia* females at the same rate as conspecific females (Figure 3D). We observed no significant difference in the courtship rate between the two independently generated *D. sechellia* knockout lines.

We then asked whether *eloF* might also mediate mate discrimination between *D. melanogaster* and *D. simulans.* As expected, when *D. simulans* males were presented with wild-type *D. melanogaster* females, they rarely proceeded to courtship (Figures 3D and S4A). However, when we knocked down *eloF* expression in *D. melanogaster* females using oenocyte-specific RNAi, males courted them at rates only slightly lower than conspecifics.

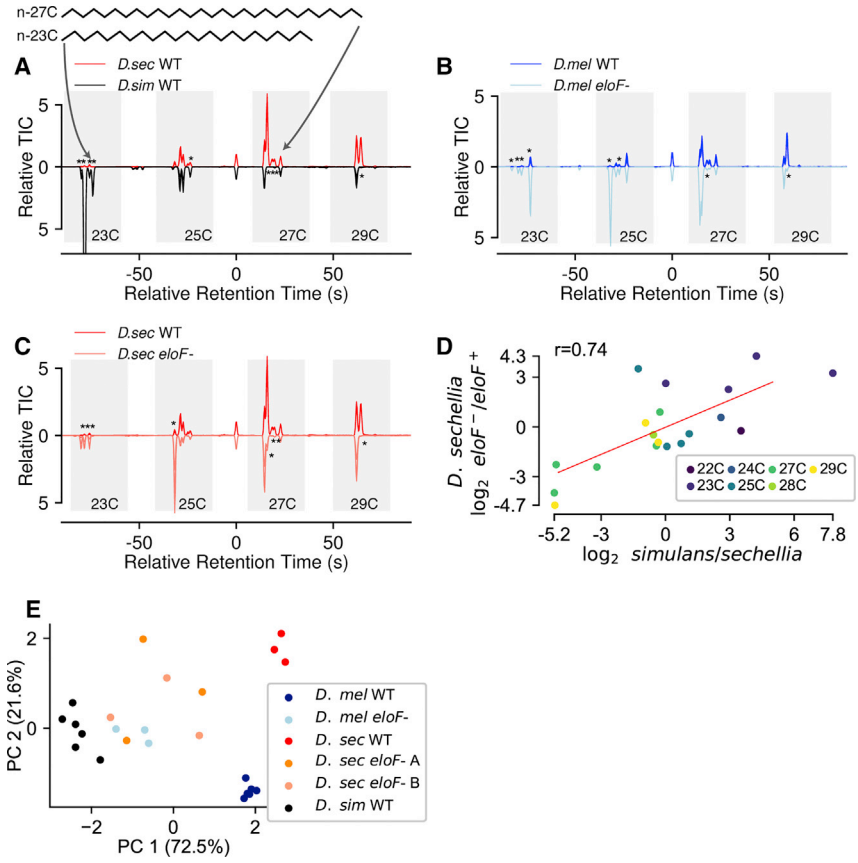

Figure 2. *eloF*– Flies Have an Overall Shorter CHC Complement

(A) Total ion chromatographs of the hydrocarbon profile of wild-type *D. sechellia* (top) and *D. simulans* (bottom). Retention time and abundance is relative to the n-hexacosane (26C) normalization peak. Grey regions indicate number of carbons in CHC backbone. CHCs with more than a 3-fold change are marked with asterisks at the location of the peak in the genotype with lower production.

(B and C) Total ion chromatographs of the hydrocarbon profile of wild-type (top) and *eloF*– (bottom) *D. melanogaster* (B) and *D. sechellia* (C).

(D) Average log2 fold changes of the measured compounds between *D. simulans* and *D. sechellia* versus log2 fold changes between wild-type and knockout of *eloF* in *D. sechellia*. Points are colored by the number of carbons in the backbone.

(E) Principal-components analysis of wild-type and *eloF*– *D. melanogaster*, *simulans*, and *sechellia*. Principal components were calculated for the wild-type data and then *eloF*– data were projected onto the same coordinates.

See also Figure S2.

The choice by males seems to be nearly binary. In the cases when *D. simulans* males did court *eloF*-bearing females, they did so approximately as quickly as they did for *D. simulans* females (Figures 3E and S4B). There was no significant difference in time between first contact between the flies and any of the steps in courtship.

Outlook and Future Work

In this study, we have found that RNA-seq in F1 hybrids is a rapid, efficient means of identifying genes potentially involved in phenotypic divergence. Neither comparisons of expression across tissues nor of ASE within a single tissue was able to sufficiently narrow the list of candidate genes (Figure 1C); however, the combination of these orthogonal filters, together with gene annotations, allowed us to focus on only three candidate genes. This can be compared with the most widely used alternative for studying the genetic basis of phenotypic divergence, QTL mapping. In QTL mapping, hundreds of progeny from genetic crosses must be genotyped and phenotyped, requiring significant effort even for rapidly reproducing species. Moreover, this effort leads to QTLs that typically span over a hundred genes, because resolution is limited by infrequent recombination events. Therefore, studies to test specific genes within those regions are often prohibitive. We envision that our approach of intersecting filters based only on RNA-seq in F1s may be widely applicable to other tissue-specific, sex-specific, stage-specific, or condition-specific traits that differ between interfertile populations or species. For example, we previously found that, in yeast, intersecting ASE with genes induced by a specific toxin pinpointed several genes whose *cis*-regulatory divergence contributed to toxin resistance [43]. This approach can also be combined with QTL mapping, though in many cases, this may not be necessary (as exemplified by the present study, where QTL data played no role in our selection of candidate genes).

Consistent with other recent observations [16], we found that CHC differences seem to be the major source of sexual isolation between *D. simulans* from both *D. sechellia* and *D. melanogaster*, and we also showed that ablating *eloF* alleviates nearly all of the isolation from both *D. sechellia* and *D. melanogaster*. The magnitude of this effect is comparable to the reduction in barriers between *D. simulans* males and *D. melanogaster* females by ablating oenocytes entirely, a much more radical intervention (*eloF* appears to represent ~85% of the barrier in this study, compared to ~100% in [41]).

We can hypothesize a parsimonious evolutionary scenario to explain our observations. *D. sechellia* and *D. melanogaster* both express *eloF* in female oenocytes; therefore, this is likely to be the ancestral state for these species, with the lower *eloF* expression in *D. simulans* being a derived change specific to this species. Our results, together with QTLs for CHC differences and mate discrimination between *D. simulans* and *D. sechellia* that contain *eloF* [23, 24], suggest that this dramatic *cis*-regulatory divergence may have led to the sexual isolation of *D. simulans*; however, further evidence, such as a reciprocal hemizygosity test [44], would be required to prove this.

This work raises several important questions. For example, because *eloF* affects so many CHCs, it is not clear which CHC(s) act as the discriminative signal. One candidate is the 27-carbon CHC 7,11-heptacosadiene, which is involved in male *D. melanogaster* and *D. simulans* preference [41], although

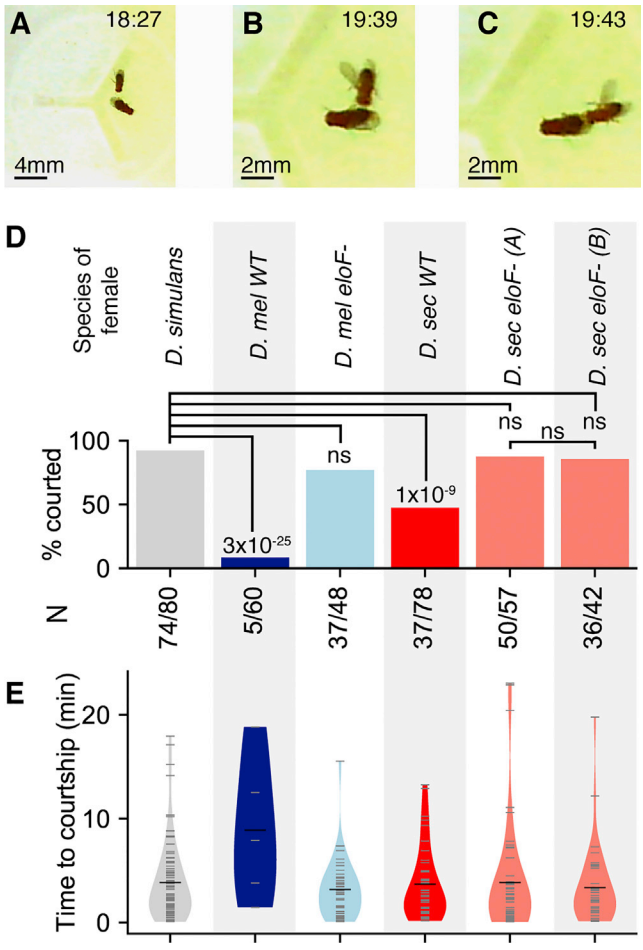

Figure 3. *D. simulans* Males Court Interspecific *eloF*– Females at Significantly Higher Rates

(A–C) We recorded between 42 and 80 pairs of single *D. simulans* males courting single females of each indicated genotype. We recorded the time between male's first tapping the female (A; ostensibly sampling the female CHCs) and either singing behavior (B) or licking of the female's posterior prior to copulation (C). N indicates the number of assays that proceeded to courtship (numerator) and the total number of assays recorded.

(D) Female flies bearing a functional copy of *eloF* (*D. melanogaster* wild-type [WT] and *D. sechellia* WT) were courted by *D. simulans* males at significantly lower rates than *D. simulans* conspecific females and interspecific females without *eloF*. We performed the indicated Fisher's exact tests for differences in courtship rate (as measured by rate of proceeding to precopulatory licking), with Bonferroni-corrected p values above each bar when significant.

(E) Violin plots of the delay between first contact between males and females and initiation of licking courtship behavior. Black lines indicate mean time to courtship. Gray ticks indicate the underlying data. Although the *D. simulans* males were slower to court *D. melanogaster* WT females, this represents only 5 cases of courtship (out of 60 trials), and no comparisons were significant by t test at even a nominal p = 0.05 cutoff.

See also Figure S4.

other CHCs could also contribute. More generally, the ecological relevance of sexual isolation measured in the lab remains unknown, considering that rates of hybridization in the wild can be strikingly higher than laboratory predictions [45].

Another open question regards the sequence change(s) that have led to the expression divergence of *eloF*. The transcription factor Doublesex has been implicated in the evolution of other *Drosophila* species' CHC profiles [33], but we have not found any divergent Doublesex binding sites near *eloF*. However, the noncoding region around *eloF* and *CG8534* contains 136 SNPs and 10 small indels, where *D. simulans* has a derived allele differing from both *D. sechellia* and *D. melanogaster* (thus matching the parsimonious evolutionary scenario described above) and several nonsynonymous changes in *eloF*.

Our finding that *D. simulans* males prefer mates lacking *eloF* suggests that male preferences have co-evolved with CHC profiles in *D. simulans*. An intriguing question for future work will be whether the gene(s) responsible for this co-evolved male preference could be identified with a similar cell-type-specific ASE approach as demonstrated here. The small population of neurons recently shown to be responsible for key differences in the male neural circuit that evaluates a species-specific CHCs [46] would be a logical focus for such a study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Fly rearing and generation
- METHOD DETAILS
  - RNA extraction and sequencing
  - Gas chromatography–mass spectrometry
  - Mating assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RNA-seq analysis
  - Mating Assays
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental information includes four figures and two tables and can be found with this article online at https://doi.org/10.1016/j.cub.2018.10.036.

ACKNOWLEDGMENTS

We thank Nirao Shah, Osama Ahmed, Mark Siegal, and the Fraser lab for helpful discussions and David Stern for helpful discussions and the *D. simulans* Tsimbazaza strain. The Bloomington Drosophila Stock Center (NIH P40OD018537) provided the RNAi strains and the Drosophila Species Stock Center at UCSD (NSF CSBR grant 1351502) provided the *D. sechellia* strains. GCMS was performed at the Vincent Coates Foundation Mass Spectrometry Laboratory at Stanford University, which is supported in part by NIH P30 CA124435. This work was supported by NIH grant 2R01GM097171-05A1 to H.B.F. and by grants from NSERC and CIHR to J.D.L. and J.J.K.

AUTHOR CONTRIBUTIONS

Conceptualization, P.A.C. and H.B.F.; Investigation, P.A.C., J.J.K., N.M.K., and D.B.; Writing – Original Draft, P.A.C. and H.B.F.; Writing – Review & Editing, P.A.C., J.J.K., N.M.K., D.A.P., J.D.L., and H.B.F.; Funding Acquisition, H.B.F. and J.D.L.; Resources, J.J.K. and J.D.L.; Supervision, J.D.L. and H.B.F.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 28, 2018
Revised: September 6, 2018
Accepted: October 12, 2018
Published: November 29, 2018

REFERENCES

1. Steiger, S., Franz, R., Eggert, A.-K., and Müller, J.K. (2008). The Coolidge effect, individual recognition and selection for distinctive cuticular signatures in a burying beetle. Proc. Biol. Sci. *275*, 1831–1838.
2. Buellesbach, J., Gadau, J., Beukeboom, L.W., Echinger, F., Raychoudhury, R., Werren, J.H., and Schmitt, T. (2013). Cuticular hydrocarbon divergence in the jewel wasp Nasonia: evolutionary shifts in chemical communication channels? J. Evol. Biol. *26*, 2467–2478.
3. Nunes, T.M., Oldroyd, B.P., Elias, L.G., Mateus, S., Turatti, I.C., and Lopes, N.P. (2017). Evolution of queen cuticular hydrocarbons and worker reproduction in stingless bees. Nat. Ecol. Evol *1*, 0185.
4. Coyne, J.A., and Orr, H.A. (1989). Patterns of speciation in Drosophila. Evolution *43*, 362–381.
5. Noor, M.A. (1995). Speciation driven by natural selection in Drosophila. Nature *375*, 674–675.
6. Coyne, J.A., and Orr, H.A. (2004). Speciation (Sinauer Associates Incorporated).
7. Servedio, M.R., and Noor, M.A.F. (2003). The role of reinforcement in speciation: theory and data. Annu. Rev. Ecol. Evol. Syst. *34*, 339–364.
8. Turelli, M., Lipkowitz, J.R., and Brandvain, Y. (2014). On the Coyne and Orr-igin of species: effects of intrinsic postzygotic isolation, ecological differentiation, X chromosome size, and sympatry on Drosophila speciation. Evolution *68*, 1176–1187.
9. Turissini, D.A., McGirr, J.A., Patel, S.S., David, J.R., and Matute, D.R. (2018). The rate of evolution of postmating-prezygotic reproductive isolation in Drosophila. Mol. Biol. Evol. *35*, 312–334.
10. Barbash, D.A. (2010). Ninety years of Drosophila melanogaster hybrids. Genetics *186*, 1–8.
11. Sokolowski, M.B. (2001). Drosophila: genetics meets behaviour. Nat. Rev. Genet. *2*, 879–890.
12. Spieth, H.T. (1952). Mating behavior within the genus Drosophila (Diptera). Bull. Am. Mus. Nat. Hist. *99*, 399–474.
13. Partridge, L., and Farquhar, M. (1981). Sexual activity reduces lifespan of male fruitflies. Nature *294*, 580–582.
14. Greenspan, R.J., and Ferveur, J.F. (2000). Courtship in Drosophila. Annu. Rev. Genet. *34*, 205–232.
15. Edward, D.A., and Chapman, T. (2011). The evolution and significance of male mate choice. Trends Ecol. Evol. *26*, 647–654.
16. Shahandeh, M.P., Pischedda, A., and Turner, T.L. (2018). Male mate choice via cuticular hydrocarbon pheromones drives reproductive isolation between Drosophila species. Evolution *72*, 123–135.
17. Garrigan, D., Kingan, S.B., Geneva, A.J., Andolfatto, P., Clark, A.G., Thornton, K.R., and Presgraves, D.C. (2012). Genome sequencing reveals complex speciation in the Drosophila simulans clade. Genome Res. *22*, 1499–1511.
18. Schrider, D.R., Ayroles, J., Matute, D.R., and Kern, A.D. (2018). Supervised machine learning reveals introgressed loci in the genomes of Drosophila simulans and D. sechellia. PLoS Genet. *14*, e1007341.
19. Kliman, R.M., Andolfatto, P., Coyne, J.A., Depaulis, F., Kreitman, M., Berry, A.J., McCarter, J., Wakeley, J., and Hey, J. (2000). The population genetics of the origin and divergence of the Drosophila simulans complex species. Genetics *156*, 1913–1931.
20. Matute, D.R., and Ayroles, J.F. (2014). Hybridization occurs between Drosophila simulans and D. sechellia in the Seychelles archipelago. J. Evol. Biol. *27*, 1057–1068.
21. Lachaise, D., David, J.R., Lemeunier, F., Tsacas, L., and Ashburner, M. (1986). The reproductive relationships of Drosophila sechellia with D. mauritiana, D. simulans, and D. melanogaster from the Afrotropical region. Evolution *40*, 262–271.
22. Coyne, J.A., Crittenden, A.P., and Mah, K. (1994). Genetics of a pheromonal difference contributing to reproductive isolation in Drosophila. Science *265*, 1461–1464.
23. Gleason, J.M., Jallon, J.-M., Rouault, J.-D., and Ritchie, M.G. (2005). Quantitative trait loci for cuticular hydrocarbons associated with sexual isolation between Drosophila simulans and D. sechellia. Genetics *171*, 1789–1798.
24. Gleason, J.M., James, R.A., Wicker-Thomas, C., and Ritchie, M.G. (2009). Identification of quantitative trait loci function through analysis of multiple cuticular hydrocarbons differing between Drosophila simulans and Drosophila sechellia females. Heredity (Edinb) *103*, 416–424.
25. Wittkopp, P.J., and Kalay, G. (2011). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. *13*, 59–69.
26. Lawrence, P.A., and Johnston, P. (1986). Observations on cell lineage of internal organs of Drosophila. J. Embryol. Exp. Morphol. *91*, 251–266.
27. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
28. Graze, R.M., McIntyre, L.M., Main, B.J., Wayne, M.L., and Nuzhdin, S.V. (2009). Regulatory divergence in Drosophila melanogaster and D. simulans, a genomewide analysis of allele-specific expression. Genetics *183*, 547–561.
29. Coolon, J.D., McManus, C.J., Stevenson, K.R., Graveley, B.R., and Wittkopp, P.J. (2014). Tempo and mode of regulatory evolution in Drosophila. Genome Res. *24*, 797–808.
30. Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods *14*, 687–690.
31. Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., et al. (2017). InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. *45* (D1), D190–D199.
32. Gramates, L.S., Marygold, S.J., Santos, G.D., Urbano, J.-M., Antonazzo, G., Matthews, B.B., Rey, A.J., Tabone, C.J., Crosby, M.A., Emmert, D.B., et al.; the FlyBase Consortium (2017). FlyBase at 25: looking to the future. Nucleic Acids Res. *45* (D1), D663–D671.
33. Shirangi, T.R., Dufour, H.D., Williams, T.M., and Carroll, S.B. (2009). Rapid evolution of sex pheromone-producing enzyme expression in Drosophila. PLoS Biol. *7*, e1000168.
34. Chertemps, T., Duportets, L., Labeur, C., Ueda, R., Takahashi, K., Saigo, K., and Wicker-Thomas, C. (2007). A female-biased expressed elongase involved in long-chain hydrocarbon biosynthesis and courtship behavior in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA *104*, 4273–4278.
35. Boyle, E.I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J.M., and Sherlock, G. (2004). GO:TermFinder–open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics *20*, 3710–3715.
36. Szafer-Glusman, E., Giansanti, M.G., Nishihama, R., Bolival, B., Jr., Pringle, J., Gatti, M., and Fuller, M.T. (2008). A role for very-long-chain fatty acids in furrow ingression during cytokinesis in Drosophila spermatocytes. Curr. Biol. *18*, 1426–1431.
37. Jallon, J.-M., and David, J.R. (1987). Variation in cuticular hydrocarbons among the eight species of the *Drosophila melanogaster* subgroup. Evolution *41*, 294–302.

38. Legendre, A., Miao, X.-X., Da Lage, J.-L., and Wicker-Thomas, C. (2008). Evolution of a desaturase involved in female pheromonal cuticular hydrocarbon biosynthesis and courtship behavior in Drosophila. Insect Biochem. Mol. Biol. *38*, 244–255.

39. Fang, S., Ting, C.-T., Lee, C.-R., Chu, K.-H., Wang, C.-C., and Tsaur, S.-C. (2009). Molecular evolution and functional diversification of fatty acid desaturases after recurrent gene duplication in Drosophila. Mol. Biol. Evol. *26*, 1447–1456.

40. Chung, H., and Carroll, S.B. (2015). Wax, sex and the origin of species: dual roles of insect cuticular hydrocarbons in adaptation and mating. BioEssays *37*, 822–830.

41. Billeter, J.-C., Atallah, J., Krupp, J.J., Millar, J.G., and Levine, J.D. (2009). Specialized cells tag sexual and species identity in Drosophila melanogaster. Nature *461*, 987–991.

42. Perkins, L.A., Holderbaum, L., Tao, R., Hu, Y., Sopko, R., McCall, K., Yang-Zhou, D., Flockhart, I., Binari, R., Shim, H.-S., et al. (2015). The transgenic RNAi project at Harvard Medical School: resources and validation. Genetics *201*, 843–852.

43. Naranjo, S., Smith, J.D., Artieri, C.G., Zhang, M., Zhou, Y., Palmer, M.E., and Fraser, H.B. (2015). Dissecting the genetic basis of a complex *cis*-regulatory adaptation. PLoS Genet. *11*, e1005751.

44. Stern, D.L. (2014). Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. Trends Genet. *30*, 547–554.

45. Coyne, J.A., Elwyn, S., and Rolán-Alvarez, E. (2005). Impact of experimental design on Drosophila sexual isolation studies: direct effects and comparison to field hybridization data. Evolution *59*, 2588–2601.

46. Seeholzer, L.F., Seppo, M., Stern, D.L., and Ruta, V. (2018). Evolution of a central neural circuit underlies Drosophila mate preferences. Nature *559*, 564–569.

47. Hu, T.T., Eisen, M.B., Thornton, K.R., and Andolfatto, P. (2013). A second-generation assembly of the Drosophila simulans genome provides new insights into patterns of lineage-specific divergence. Genome Res. *23*, 89–98.

48. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

49. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

50. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

51. Krupp, J.J., and Levine, J.D. (2010). Dissection of oenocytes from adult Drosophila melanogaster. J. Vis. Exp. 2242.

52. Chow, S.-C., Wang, H., and Shao, J. (2007). Sample Size Calculations in Clinical Research, Second Edition (CRC Press).

53. van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods *12*, 1061–1063.

54. Fontanillas, P., Landry, C.R., Wittkopp, P.J., Russ, C., Gruber, J.D., Nusbaum, C., and Hartl, D.L. (2010). Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. Mol. Ecol. *19* (*Suppl 1*), 212–227.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
RNeasy Micro Kit	QIAGEN	Cat# 74004
NextFLEX RNA-seq library preparation kit	BioO Scientific (now Perkin Elmer)	Cat# NOVA-5138-01
Deposited Data		
Raw and analyzed RNA-seq data	This paper	GEO: GSE114478
D. simulans v2.01 Genome	[47]	ftp://ftp.flybase.net/releases/FB2016_01/dsim_r2.01/dna/dsim-raw_scaffolds-r2.01.tar.gz
D. simulans gDNA reads	[47]	SRA: SRR520334
D. sechellia gDNA reads	[29]	SRA: SRR869587
Experimental Models: Organisms/Strains		
D. melanogaster: RNAi of *bond* y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8] = TRiP.HMS01154}attP2	Bloomington Drosophila Stock Center	BDSC:34676; FlyBase: FBtp0065394
D. melanogaster: RNAi of *eloF* y[1] sc[*] v[1]; P{y[+t7.7] v[+t1.8] = TRiP.HMS00609}attP2	Bloomington Drosophila Stock Center	BDSC:34393; FlyBase: FBtp0064874
D. melanogaster: RNAi of CG8534 y[1] v[1]; P{y[+t7.7] v[+t1.8] = TRiP.HMC03518}attP2	Bloomington Drosophila Stock Center	BDSC: 53299; FlyBase: FBtp0091013
D. melanogaster: Gal4 Driver PromE(800) Gal4/Tm3.5b	[41]	PromE(800)-Gal4
D. simulans: tsimbazaza strain	Stern Lab, Janelia Farms	Stock #195
D. sechellia: Wild-type	Drosophila Species Stock Center at UCSD (now at Cornell University)	Stock #14021-0248.25
D. sechellia: eloF- strain A	This paper	D. sec eloF- A
D. sechellia: eloF- strain B	This paper	D. sec eloF- B
Oligonucleotides		
Upstream Cas9 sense oligo: CTTCGCAGCGATCC ATGGGTCCCCA	This paper	N/A
Upstream Cas9 antisense oligo: AAACTGGGGAC CCATGGATCGCTGC	This paper	N/A
Downstream Cas9 Sense oligo: CTTCGATCCGCA TCCGTAGGTCAA	This paper	N/A
Downstream Cas9 Antisense oligo: AAACTTGACC TACGGATGCGGATC	This paper	N/A
Primer: RFP screening Forward: CTCCCAGCGAT CATTCATTT	This paper	N/A
Primer: RFP screening Reverse: GCTGCTACACTT GCCACAAA	This paper	N/A
Primer: eloF screening Forward: TCTGCAGGTTCT GATGGCAG	This paper	N/A
Primer: eloF screening Reverse: ACTGTGGAAAG GCAACACCA	This paper	N/A
Software and Algorithms		
STAR	[48]	https://github.com/alexdobin/STAR/releases
GATK	[49]	https://software.broadinstitute.org/gatk/
DESeq2	[27]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Kallisto	[50]	https://pachterlab.github.io/kallisto/about
Assorted analysis and glue scripts	This paper	https://doi.org/10.5281/zenodo.1436010
Other		
Video recordings of courtship assay	This paper	https://drive.google.com/a/stanford.edu/file/d/1m3K0vsW2qSqFCHtnrBj7ZVHNflOLaogl/view
Resource website for the RNA-seq data	This paper	https://combsfraser-oenocytes.appspot.com

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Hunter Fraser (hbfraser@stanford.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Fly rearing and generation

For RNAi flies, virgin females of the shRNA driver were isolated within 18 hr of eclosion, then kept isolated from males for 3 days on standard cornmeal media to ensure virgin status. We used Bloomington Stock IDs 34676 (*bond*), 53947 (*eloF),* and 53299 (*CG8534).* We combined approximately 25 UAS-shRNA females with approximately 10 PromE(800) Gal4 driver males [41]. As negative controls, we crossed PromE(800)-gal4 males with females of Bloomington stock #32186, which carries 10 copies of UAS-driven mCD8-tagged GFP. Adults were moved to fresh vials every 3 days to ensure separation of the parents and the Gal4+UAS offspring.

Knockout *D. sechellia* flies were created using CRISPR/Cas9 mediated editing. We designed guides to cut at the 55th nucleotide downstream of the ATG and the 114th nucleotide upstream of the stop codon of *GM23846* (the *D. sechellia* ortholog of *eloF)*. We used sense oligos CTTCGCAGCGATCCATGGGTCCCCA (gene 5′-ward cut site) and CTTCGATCCGCATCCGTAGGTCAA (gene 3′-ward cut site). Embryos were injected (WellGenetics, Taipei, Taiwan) with both guides and a dsDNA donor containing ~1000bp homology arms and RFP driven by a 3xP3 promoter and flanked by LoxP sites. Embryos were from the *D. sechellia* genome strain #14021-0248.25. Out of 200 injections over 2 rounds, 2 offspring showed positive RFP expression in the eyes, as expected for the 3xP3 promoter. As shown in Figure S3, we screened for presence of the inserted sequence using primers CTCCCAGCGATCATTCATTT and GCTGCTACACTTGCCACAAA (170bp product), and for absence of *eloF* using TCTGCAGGTTCTGATGGCAG and ACTGTG GAAAGGCAACACCA (306bp product).

All flies, either wild-type, RNAi, or CRISPR edited were separated by sex within 18 hr of eclosion, then kept isolated for 5-7 days to ensure virgin status. Any vials with larvae after 5 days were discarded. Since the PromE(800)-gal4 construct is balanced with Tm3.5b, we selected straight-winged flies as RNAi positive.

METHOD DETAILS

RNA extraction and sequencing

We mated *D. sechellia* males to *D. simulans* females and dissected both oenocytes and fat bodies from the progeny, pooling material from 20 individuals from each sex. Oenocyte and fat body dissections were performed as described in [51]: briefly, flies were pinned to a Sylgard dissection plate (Dow-Corning) and covered with chilled Shields and Sang M3 medium (Sigma). The oenocytes and fat body of 10-day-old *D. simulans*/*D. sechellia* hybrid flies were isolated separately from the dorsal abdominal segments of both adult male and female abdomens using a fine tungsten dissecting needle. Each tissue sample represented the pooled material collected from 20 flies. Hybrid flies were reared in a 12hr light:12 hr dark cycle and tissues dissected at equal time intervals across a 24hr period. Immediately following dissection tissues were placed into cell lysis buffer to aid in preserving the integrity of the RNA. Total RNA was isolated using the RNeasy Micro kit (QIAGEN). We collected two independent samples of each tissue from each sex, which in our experience is adequate for identifying strong patterns of allele specific expression.

We prepared libraries from the RNA using the NextFLEX RNA-seq library preparation kit (BioO Scientific, Austin, TX), and sequenced the libraries using 101bp paired end reads on an Illumina HiSeq 2000.

Gas chromatography–mass spectrometry

We performed GCMS by anesthetizing 5 females at 4°C for 3-5 min, then washing them for 5 min with 50μL of hexane spiked with 10mg/mL of n-hexane as a standard. Spectra were obtained using an Agilent (HP) 7890/5975 single quadrupole GC-MS instrument with a split ratio of 1:20, injector temperature of 280°C, and an oven temperature program of 35°C hold for 3.75min, 20°C/min ramp from 35°C to 320°C, and a 320°C hold for 7 min. We collected spectra for at least 3 sets of 5 flies for each genotype (but 5 spectra for *D. melanogaster* WT and 6 for *D. simulans*). Identities of different hydrocarbon peaks were inferred by inspecting the singly-ionized mass spectrum bin.

Mating assays

We performed mating assays by anesthetizing separate vials of males and females at 4°C for 3-5 min, then used a paintbrush to transfer one male and one female to each well of the mating chamber. The mating chamber was 3D printed from acrylic plastic and has 18 separate 2cm diameter × 5mm circular wells, with a removable clear plastic lid. We allowed flies to acclimate at room temperature and ambient light for 10-15 min, then recorded 30 m of video with bright lights, which we found were required for *D. simulans* males to initiate courtship. The mating light was a 75W, 14'' circular fluorescent bulb placed approximately 30cm above the mating chamber. Video of mating assays was recorded using a Dino-Lite digital microscope, then analyzed by two separate graders (PAC and NMK), who recorded the time of first contact by the male, the time of the male first following the female, the time of the first wing song by the male, and the time of first licking by the male of the female's abdomen [11]. Prior to analyzing videos, we estimated that sample sizes of 50 assays would be sufficient to identify a change from 90% successful mating to 55% successful mating with 90% power [52]. Graders were blinded to the fly identities in each video, which had a uniform, random 4-digit number as the file identifier. We noted the time of the first instance of various copulatory behaviors: tapping, male wing song, and licking. With the exception of licking, these behaviors are not subject to rejection by females (the mating chambers are small enough that females are effectively unable to escape, while tapping is very rapid and wing song does not involve contact), and thus primarily represent choice by the males.

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA-seq analysis

We used 2 independent samples of each tissue, each consisting of material from 20 flies.

In order to minimize the number of sequence mismatches between the strains we used and the reference genome, we created a corrected *D. simulans* genome sequence by using bowtie2 version 2.2.5 with arguments–very-sensitive to map genomic DNA reads from *D. simulans* and *D. sechellia* to the FlyBase 2.01 *D. simulans* reference genome [29, 47]. Polymorphisms were called using GATK (HaplotypeCaller–genotyping_mode DISCOVERY -fixMisencodedQuals -stand_emit_conf 10 -stand_call_conf 30) [49], then the ~34,000 SNPs that were fixed in both *D. simulans* and *D. sechellia* were replaced with the consensus sequence (this step was more important for creating a *simulans/sechellia* version of the *D. melanogaster* genome for Figures S1E and S1F).

RNA-seq reads were mapped to the *D. simulans* reference genome using STAR with arguments–outFilterMultimapNmax 1–outSAMattributes MD NH–clip5pNbases 6–sjdbGTFfile [47, 48]. Following the WASP pipeline, duplicate reads were discarded randomly, then filtered based on whether reads with the alleles swapped *in silico* to create artificial transcripts from the other species mapped to the same position [53]. Reads were assigned to a species only if both paired ends mapped unambiguously to one species, and allele-specific expression negative binomial p values were calculated from aligned read counts using DESeq2 with model ~Replicate + AlignsToSpecies [27]. Despite the use of a *D. simulans* reference genome, we found a majority of reads were assigned to *D. sechellia* (Table S1), possibly indicating a widespread species-specific bias in the strength of cis-regulatory elements, as has been observed in other interspecific *Drosophila* crosses [54]. Alternatively, the pattern could be explained if 3-4 of the 20 dissected flies in each sample were actually non-hybrid *D. sechellia*, though this seems unlikely given that > 99.1% of reads from the mitochondrial genome were assigned to *D. simulans*.

Default DESeq settings were used to correct for multiple hypothesis testing. Transcript abundances were estimated using kallisto with default arguments [50]. We used sleuth to identify differentially expressed genes between samples with matched sex and tissue type [30]. Since kallisto has not been extensively tested for suitability allele-specific expression, we opted to use the conservative mapping-based approach outlined above.

Mating Assays

Video was analyzed by two separate graders (PAC and NMK), who recorded the time of first contact by the male, the time of the male first following the female, the time of the first wing song by the male, and the time of first licking by the male of the female's abdomen [11]. Graders were blinded to the fly identities in each video. Rates of proceeding to pre-copulatory licking were compared using Fishers exact test. Time between tapping and courtship behaviors were compared using a t test.

DATA AND SOFTWARE AVAILABILITY

An interactive tool to explore the RNA-seq dataset is available at http://combsfraser-oenocytes.appspot.com/. The accession number for the RNA-seq data reported in this paper is GEO: GSE114478. Video data are available at https://drive.google.com/a/stanford.edu/file/d/1m3K0vsW2qSqFCHtnrBj7ZVHNflOLaogl/view. Analysis scripts are available at https://github.com/TheFraserLab/CombsOenocytes2018 (https://doi.org/10.5281/zenodo.1436010).