Intersecting experimental evolution and CRISPR screens to identify novel toxin resistance loci

Michele Marconcini[1], Steeve Cruchet[1], Srishti Goswami[2], Raghuvir Viswanatha[2], Matthew Butnaru[2,3], Joydeep De[4], Camilla Roselli[4], Dafni Hadjieconomou[4], Norbert Perrimon[2,3], Stephanie E. Mohr[2], Richard Benton[1,*]

[1]Center for Integrative Genomics
Faculty of Biology and Medicine
University of Lausanne
CH-1015
Lausanne
Switzerland

[2]Department of Genetics
Blavatnik Institute
Harvard Medical School
Boston
MA 02115
USA

[3]Howard Hughes Medical Institute
Boston
MA 02115
USA

[4]Institut du Cerveau-Paris Brain Institute
Sorbonne Université
Inserm
CNRS
Hôpital Pitié-Salpêtrière
75013
Paris
France

*Corresponding author:
E: richard.benton@unil.ch
T: ++41(0)21 692 3932

Short title: Orthogonal selection screens for toxin resistance genes

Abstract

Understanding toxin resistance in insects is key to appreciate niche adaptations, but remains challenging due to its often complex genetic basis. A well-known example is the specialized association of *Drosophila sechellia* with noni fruit (*Morinda citrifolia*), which is toxic to most other insects, including the closely-related drosophilids *D. simulans* and *D. melanogaster*. Noni toxicity is due to its high concentration of octanoic acid (OA), but the mechanisms that determine sensitivity or resistance to OA remain poorly understood. Here, we experimentally-evolved *D. simulans* with increased OA resistance, identifying multiple loci under selection. Cross-referencing these with a genome-wide, OA-resistance CRISPR screen in a *D. melanogaster* cell line highlighted two proteins: Kraken, a putative detoxification enzyme expressed in digestive and renal tissues, and Alkbh7, a mitochondrial protein linked to fatty acid metabolism. Both genes show elevated expression in *D. sechellia* and OA-resistant *D. simulans*. In *D. melanogaster*, *kraken* mutants are more OA-sensitive, while *Alkbh7* overexpression increased OA resistance. Importantly, mutation of these genes in *D. sechellia* reduced OA tolerance. Our identification of genes underlying OA resistance in laboratory and natural contexts demonstrates how complementary, cross-species selection approaches can provide insights into toxin susceptibility and adaptation mechanisms, with practical applications in the characterization of novel insecticides.

Introduction

How insects resist toxic chemicals in the environment is of both basic scientific interest – reflecting insects' arms race with plant food sources [1, 2] or venomous predators [3] – and applied interest, because humans have long used chemicals to combat insect disease vectors and agricultural pests [4]. The best-understood insecticides are those that bind to and modulate neuronal ion channels, where resistance arises through changes in these target proteins [5, 6]. For example, dichlorodiphenyltrichloroethane (DDT) and pyrethroids prevent closure of voltage-gated sodium channels, resulting in neuronal excitotoxicity and organismal paralysis; diverse insect species displaying resistance for these chemicals have independently-arising mutations in genes encoding these channels [7, 8]. Similarly, neonicotinoids bind to and hyperactivate nicotinic acetylcholine receptors, leading to neuronal death; mutations in various receptor subunits can confer resistance [9]. Even in such well-studied examples, however, it is thought that other neuronal targets exist [10] and/or that the chemicals impact receptors in non-neuronal tissues [11]. For many insecticides – as well as other environmental pollutants [12] – the mode of action is incompletely, or not at all, characterized [6].

Resistance observed in nature is only partially understood, reflecting the contribution of many genes beyond those encoding target proteins, which can underlie distinct defense mechanisms [13-15]. For example, thickening of the cuticle to reduce entry of insecticides into the body is widely considered an effective resistance mechanism [16, 17]. Cuticular changes have been linked to overexpression of proteins implicated in exoskeleton development (e.g., cuticle proteins, ABC transporters) but their precise roles, and how they are upregulated, remain unclear [17]. Increased expression of metabolic enzymes that degrade insecticides is another commonly observed adaptation in resistant strains, sometimes resulting from copy-number increases in the corresponding genes [18]. Duplicated metabolic genes can also undergo neofunctionalization to produce enzymes with altered specificity, as

illustrated by Cytochrome P450s in strains of the brown planthopper that recently evolved resistance to the neonicotinoid imidacloprid [19], as well as by more ancient evolution of glutathione S-transferases during transition of the *Scaptomyza* genus of Drosophilidae to herbivory to counter toxic isothiocyanates [20]. However, the lack of experimental resources in many insect species renders it difficult to characterize the causal contribution of specific genetic changes.

A compelling natural model clade study the evolution of toxin resistance is the trio of closely-related drosophilid species, *Drosophila melanogaster*, *Drosophila simulans* and *Drosophila sechellia* (Fig. 1A). *D. sechellia* is an extreme specialist on the toxic noni fruit of the *Morinda citrifolia* shrub [21, 22]. Amongst *D. sechellia*'s numerous adaptations [22-27], this species has evolved high resistance to octanoic acid (OA), an abundant noni chemical that is toxic to other drosophilids and more divergent insects [22, 28-33]. Previous pioneering studies of *D. sechellia*'s resistance to OA has been studied through genomic mapping-based approaches in *D. sechellia*/*D. simulans* hybrids, which revealed this trait to be polygenic, comprising at least five genomic regions [28, 34-36]. One of the main effect loci encompasses a cluster of 18 genes including members of the *Osiris* (*Osi*) family, which are thought to function in cuticle patterning [37, 38]. RNA interference (RNAi) knockdown of some *Osi* genes in *D. melanogaster* led to reduced tolerance of OA [39, 40]. However, genetic evidence for the contribution of these genes to OA resistance in *D. sechellia* is lacking.

Given the challenges of identifying candidate genes through genomic mapping-based strategies, in this work we took two complementary approaches to investigate the evolution of OA resistance in this drosophilid trio: experimental evolution in *D. sechellia*'s closest relative, *D. simulans*, through an Evolve-and-Resequence strategy [41], and CRISPR/Cas9-based genome-wide genetic screening in cultured *D. melanogaster* cells [42]. We show how the intersection of these approaches can identify novel genes contributing to OA susceptibility/resistance in these drosophilids, establishing these methods and this species clade as a powerful paradigm to investigate ecotoxicology.

Results

Experimental evolution of OA resistance in *D. simulans*

A previous study demonstrated that *D. simulans* strains can be experimentally evolved for higher OA resistance, enabling the identification of genomic regions (through microsatellite markers) responding to selection [43]. This success inspired us to take an Evolve-and-Resequence approach, which uses high-throughput sequencing to analyze the entire genome of populations before and after exposure to specific selective pressures, thereby facilitating the identification of adaptive loci from standing genetic variation in controlled and replicable settings [41, 44-47]. For studies of insecticide resistance, it has been predominantly used to examine changes in frequency of alleles of known genes rather than as a means to identify novel loci [48, 49].

The susceptibility/resistance of drosophilids to OA can be easily assessed by exposing flies to the toxin in a culture tube and counting the survivors after 24 h (Fig. 1B). In line with previous observations [34, 43], we found that a dose of 3 μl of pure OA (mixed with 1 ml glucose solution) leads to death of ~60% of *D. simulans* but has no effect on *D. sechellia* (Fig. 1C). By collecting the *D. simulans* survivors, we

150 reasoned that we could allow these to mate and establish the next generation (Fig.
151 1B), thus enriching genetic variants conferring higher resistance, similarly to [43].
152 Repeating the process over multiple generations and multiple replicates would allow
153 us to track phenotypic changes in resistance and investigate the associated
154 trajectories in allele frequencies.

155 Standing genetic variation should be present in the evolving populations for
156 natural selection to act upon [50, 51]. As the spontaneous mutation rate in drosophilids
157 is too low for artificial selection [52], we generated ten independent polymorphic base
158 populations of *D. simulans* from 93 isogenic strains [53] by round-robin crossing (Fig.
159 1D). These lines contain millions of SNPs, but we note that they were sampled from a
160 single wild population in California, therefore reflecting the genetic variation present in
161 that locality. The base populations therefore might not include all variants found across
162 the species' range, including those in regions where *D. sechellia* co-occurs. To
163 minimize bottleneck effects, each population was kept at a size of ~1000 flies. At each
164 generation, ~500 flies from each population were randomly collected and selected for
165 resistance to 3 µl OA. Over the course of 25 generations, all the populations showed
166 increased levels of survivability, suggesting parallel adaptive evolution (Fig. 1E). Given
167 the consistent responses across all populations, we also analyzed the phenotypic
168 changes grouping all replicates, finding that the mean survivability significantly
169 increased from ~43% at generation 0 to ~78% at generation 25 (Fig. 1F).

170 As we kept the same amount of OA throughout the 25 generations of selection,
171 the selective pressure lowered as the populations' survivability increased, evidenced
172 as a larger fraction of flies surviving and reproducing. Viewing the phenotype of
173 populations both individually (Fig. 1E) and collectively (Fig. 1F), it was apparent that
174 the change in phenotype plateaued at around generation 15. For continued evolution,
175 we therefore gradually increased the amount of OA during generations 26-31 (Fig.
176 1G), reaching a selection pressure comparable to the initial phase of the experiment
177 with 9 µl OA (~38% mean survivability across all populations) (Fig. 1G). We then
178 performed a second round of selection at this dose, observing an increase in
179 survivability across all populations, reaching ~70% by generation 50 (Fig. 1H). As the
180 populations adapted to the new condition, a second plateau started around generation
181 40 (Fig. 1H-I). Over 50 generations we were therefore able to evolve *D. simulans* to
182 have approximately 5-fold higher survivability to 9 µl OA compared to the unevolved
183 populations (Fig. 1J), though still falling short of the survival exhibited by *D. sechellia*
184 (Fig. 1J). These plateaus can be explained by the decreasing selective pressure
185 exerted by fixed OA volumes as an ever greater proportion of flies adapt and survive
186 the treatment. Determining whether an upper limit exists that would constrain the
187 evolution of a *D. sechellia*-like level of phenotypic resistance would require additional
188 selection rounds using progressively higher OA concentrations over an extended (and
189 unknown) number of generations. OA is only one component of noni fruit, although
190 probably the major toxic chemical [31]. We therefore tested three of our evolved
191 populations for resistance to noni fruit pulp, finding that they exhibited much higher
192 survival on this substrate compared to the same original populations maintained under
193 neutral conditions (i.e., no OA selection) (Fig. 1K).

194 Specialization of *D. sechellia* on toxic noni is associated with a reduced egg-
195 laying capacity [54, 55] and a shorter (albeit strain-specific) lifespan [26, 56, 57]. Our
196 success in evolving *D. simulans* with a partial *D. sechellia*-like resistance trait begs the
197 question as to whether these selected strains exhibit other phenotypic changes as a
198 trade-off for higher toxin resistance. We tested whether the evolved resistant
199 populations of *D. simulans* exhibit similar phenotypic changes in control conditions.

However, none of the three focal populations tested exhibited significant decreases in the number of eggs laid (Fig. 1L), developmental viability (Fig. 1M) or lifespan (Fig. 1N). Thus, at least within the timeframe of our experimental evolution experiment, *D. simulans* did not display obvious phenotypic trade-offs for the gain in OA resistance.

Dynamic directional selection of allele frequencies

We hypothesized that the observed phenotypic changes in *D. simulans* depended on the directional selection of alleles of consistent loci across the ten populations, while most of the genome evolves under neutrality. Allele frequencies at the selected loci should therefore consistently change across all populations, allowing us to statistically distinguish them from randomly evolving loci (Fig. 2A). We sequenced the genomes of the populations of evolved (generation 25 and generation 50) and unevolved (generation 0) flies, identifying 5,127,840 high-quality variants distributed across all chromosomes. Pairwise identity-by-state analysis indicated that as the number of generations increased, the proportions of shared alleles decreased, both between time points and between pairs of populations (left-to-right and top-to-bottom, respectively, in Fig. 2B).

To identify the selected genetic variants underlying the phenotypic adaptation, we analyzed the genome-wide allele frequency changes both between generations 0 and 25, and between generations 0 and 50, across all populations. The identification of significant changes in frequency is non-trivial, and many methods have been developed [58-60]. We therefore assessed the significance of frequency changes using two complementary approaches: the Cochran–Mantel–Haenszel (CMH) test [61, 62], a widely-used method that tests for allele frequency changes between two time points, and Bait-ER [60], a Bayesian method that tests for selection while explicitly modelling genetic drift. Both tests identified large portions of the genome under positive selection at generation 25 and 50 (Fig. 2C,D). The majority of prominent peaks were found by both methods and across time points (Fig. 2E).

Based on the analysis of genome-wide linkage disequilibrium decay (Fig. S1A), we grouped the significant SNPs identified by each of the two methods into genomic blocks according to their proximity (i.e., <50 kb). We retrieved all the candidate genes that fall within the block intervals, obtaining a total of 439 and 401 genes for the CMH and Bait-ER, respectively, at generation 25, and 1238 and 174 genes at generation 50 (Data S3). These genes likely include true targets of selection but also neighboring genes affected by hitchhiking. The increase in CMH candidate genes over time – in contrast to the decrease observed with Bait-ER – might reflect the CMH test's tendency to overestimate significance and to be insensitive to divergence among replicate populations.

The overlap of all significant genes between the two time points is <30%. This might reflect that a large fraction of the candidate genes results from noise caused by genetic drift and draft (hitchhiking) and/or reflect a shift in the phenotypic contribution of genes under the stronger selection (i.e., the higher dose of OA) during generations 27-50. Moreover, previous work on *Drosophila* using temporal sampling showed that adaptation involving complex traits can lead to non-overlapping sets of selected SNPs, in contrast with the simple expectation of continuous increases in allele frequency until fixation [41, 63]. Gene ontology analysis is complicated at this stage as most hits are likely hitchhikers of the true causal genes. It is also hard to compare currently to previous mapping studies in *D. simulans/D. sechellia* hybrids [34-36] or in the earlier *D. simulans* experimental evolution experiment [43], which only identified wide

genomic regions. The sole exception, a 170-kb window on 3R containing only 18 genes (including several *Osiris* genes) [64] did not emerge in our analysis.

A genome-wide, pooled CRISPR screen for OA susceptibility and resistance genes in *D. melanogaster* S2R+ cells

Given that these previous studies also contain extensive hitchhiking signals and substantial noise across large genomic regions, comparing their results with our analyses would likely yield numerous false positives and still leave a large pool of putative loci, offering limited benefit for fine-mapping. We therefore sought an orthogonal screening approach. Motivated by previous observations of OA toxicity for insect Sf9 cultured cells [33], we found that OA also leads to reduced viability of *D. melanogaster* S2R+ cells (Fig. 3A). This phenotype prompted us to perform a selection experiment in this cell line through exploitation of a genome-wide, pooled CRISPR/Cas9 mutagenesis approach (Fig. 3B) [42]. In brief, a library of sgRNA-encoding plasmids (6-8 sgRNAs/gene) was transfected into Cas9-expressing cells where site-specific plasmid integration predominantly limits targeting to a single gene per cell. Two replicate cultures of this pool of cells were either left untreated, or exposed to OA in the medium, at two different concentrations (Fig. 3B). After 26 days in culture, DNA was extracted from the cell pool, then sgRNA sequences were PCR amplified and sequenced to determine those that were enriched after OA selection (reflecting genes whose loss enhances cell viability, referred to hereafter as "susceptibility genes") or depleted (reflecting genes whose loss leads to lower OA resistance, hereafter "resistance genes").

To identify candidate OA susceptibility/resistance genes, we first filtered out genes essential for general cell survival [42] and genes that were not detectably expressed in S2R+ cells (using DGET [65]. (Removal of such genes are inevitable limitations of this cell-based screening approach; see Discussion). After comparing results from each condition and replicate, we identified the top 5% susceptibility genes (Fig. 3C) and top 5% resistance genes (Fig. 3D). Amongst the 87 susceptibility genes, there was no single stand-out hit, suggesting that – at least in S2R+ cells – OA does not have a single major cellular target (unlike, for example, insecticidal bacterial toxins that bind a specific cell surface receptor [66]). Gene-ontology analysis of these susceptibility genes revealed a significant enrichment of genes involved membrane trafficking (Fig. 3C). As OA has previously been hypothesized to disrupt cellular membranes [33, 67], the recovery of these hits suggests that general reduction in membrane transport reduces OA uptake into cells and/or renders membranes less susceptible to OA treatment. 22 OA resistance genes were identified, encoding diverse types of proteins (Fig. 3D), suggesting there are multiple mechanisms by which cells can resist the toxic effects of OA.

We next examined which of the genes identified in the S2R+ cell selection experiments might be relevant for evolution of OA resistance in whole animals by determining those also present within significant blocks of the experimental evolution analyses (Fig. 3E,F). Four susceptibility genes were identified: CG1620 (a predicted SANT-MYB domain transcription factor), CG4291 (a putative component of the splicing machinery), *Nph* (a chromatin remodeling factor [68]) and *Debcl* (a pro-apoptotic member of the Bcl-2 family [69]); these are all likely to have indirect roles in conferring susceptibility to OA given their broad functions. By contrast, two resistance genes have potentially direct roles in conferring tolerance to OA, *kraken* and CG14130. *kraken* encodes a putative serine hydrolase [70] and stood out both

because it was the top hit in the resistance gene screen (Fig. 3D), and because of its location in one of the most prominent peaks in the experimental evolution at generations 25 and 50 (Fig. 3F and Fig. 2C-D). CG14130 encodes the orthologue of mammalian *Alkbh7*, which is involved in regulating fatty acid metabolism [71, 72], and is located within a significant peak at generation 25 (Fig. 3F and Fig. 2C). Therefore, intersecting the experimental evolution and CRISPR screening results allowed us to narrow down tens or hundreds of potential candidates to two focal genes for downstream investigations.

Sequence, expression and evolutionary analyses of candidate OA resistance genes

To assess the candidacy of *kraken* and CG14130 (hereafter *Alkbh7*) in contributing to OA resistance in animals, we first examined their sequence and expression in *D. melanogaster*. Kraken belongs to a large family of serine hydrolases, which encompass lipases, esterases and proteases that use a nucleophilic serine residue in the active site for hydrolysis of substrates [70]. Consistent with the conservation of this serine in Kraken (Fig. 4A and Fig. S2A), an activity-based proteomic screen of the serine hydrolase superfamily in *D. melanogaster* provided evidence for *in vivo* catalytic activity of Kraken [73]. *kraken* is broadly expressed in adults, with particular enrichment in the gut and Malpighian tubules (Fig. 4B), similar to previous observations of *kraken* expression in the embryo by RNA *in situ* hybridization [70]. Using the Fly Cell Atlas [74], we examined sub-tissue adult expression of *kraken*: in the Malpighian tubules, *kraken* is expressed at highest levels in the principal cells (Fig. 4C), which are involved in ion transport and detoxification. In the gut, *kraken* transcripts are most robustly detected in various types of enterocytes, which secrete digestive enzymes [75] (Fig. 4C). However, Kraken lacks an N-terminal signal sequence [76], suggesting that the protein remains in the cytoplasm of these cells, unlike canonical digestive enzymes [75], and more similar to detoxification enzymes of the Cytochrome P450 and Glutathione S-transferase families [77]. To examine *kraken* expression *in situ*, we generated a *kraken* promoter-Gal4 driver line, which induced expression of a fluorescent reporter in cells of the proventriculus, the crop duct and the hindgut (Fig. 4D). No GFP expression was detected in Malpighian tubules (data not shown), which we suspect reflects the lack of necessary regulatory sequences in the driver transgene.

Alkbh7 is a member of the alpha-ketoglutarate-dependent hydroxylase (Alkbh) family (Fig. 4A), which are enzymes that act on diverse substrates in processes such as biosynthesis, metabolism and epigenetic regulation [78]. As in mammals [72], *D. melanogaster Alkbh7* displays broad tissue expression (Fig. 4B). Murine Alkbh7 is localized to the mitochondrial matrix [72]. This property appears to be conserved in the *Drosophila* protein, which has a mitochondrial targeting sequence (Fig. 4A and Fig. S2B) [79]; consistently, we identified Alkbh7 within *D. melanogaster* proteomes of the mitochondrial matrix [80, 81].

Although the sequences of both Kraken and Alkbh7 are overall highly similar across *D. melanogaster*, *D. simulans* and *D. sechellia* (Fig. S2A,B), we wondered whether we could detect signs of selection at these loci in *D. sechellia* employing population sequence datasets of *D. sechellia* and *D. simulans* [82]. Using the McDonald-Kreitman test [83] on the coding sequences of both genes, we found evidence for gene-level positive selection on *kraken* but not *Alkbh7* (Fig. S2C). Because selection might act pervasively at individual sites rather than along the entire

protein coding region, we also implemented a codon-based analysis using FUBAR (Fast, Unconstrained Bayesian AppRoximation) [84], which identified 1 and 4 codons under diversifying positive selection in *kraken* and *Alkbh7,* respectively (Fig. S2D). Although their functional relevance remains unclear, such mutations might reflect underlying evolutionary dynamics at the protein level.

We next compared the expression levels of these genes both in laboratory strains of *D. sechellia* and *D. simulans* (Fig. 4E), as well as in RNA-seq datasets of various strains of *D. sechellia*, *D. simulans* and *D. melanogaster* [85-87] (Fig. 4F, Fig. S3A,B). While there were some exceptions, overall both *kraken* and *Alkbh7* show higher expression levels in *D. sechellia* adults and larvae compared to the other species. Importantly, in our experimentally-evolved *D. simulans* populations, we also found that both *kraken* and *Alkbh7* were more highly expressed compared to unselected populations (Fig. 4G). The expression differences between species, and between unevolved and evolved *D. simulans*, reflect variation in basal levels of expression, as prior exposure to OA did not lead to changes in expression in either *D. sechellia* or *D. simulans* (Fig. S3C).

Functional contributions of *kraken* and *Alkbh7* to OA resistance

Given the expression differences observed in the naturally and experimentally evolved OA-resistant drosophilids, we asked whether artificially modulating gene expression would impact OA tolerance. We first performed ubiquitous RNAi of these genes in *D. melanogaster* (Fig. S4A) and assessed the effects on resistance to OA using a short-term, plate-based assay [28] (see Methods). $kraken^{RNAi}$ led to significantly diminished OA resistance compared to genetic controls, while $Alkbh7^{RNAi}$ did not have a discernable effect (Fig. 5A). We next generated deletion mutants for these genes in *D. melanogaster* (Fig. S4B). Similar to the RNAi experiments, *kraken* mutants, but not *Alkbh7* mutants, displayed reduced resistance to OA (Fig. 5B). All genotypes remained fully viable when no OA was added (Fig. S5).

Loss-of-function analyses might fail to uncover a role for *Alkbh7* in whole animals if there are redundant resistance mechanisms absent from S2R+ cells. We therefore took a complementary approach by overexpressing these genes in *D. melanogaster*. Strikingly, overexpression of *Alkbh7* resulted in survivability of *D. melanogaster* to levels close to those observed for *D. sechellia* under the same experimental conditions (Fig. 5C, Fig. S4C), indicating that this gene is sufficient to confer high-level OA resistance. By contrast, *kraken* overexpression did not have any detectable impact on OA resistance (Fig. 5C); a similar lack of effect was seen when overexpressing *D. sechellia kraken* (Fig. S2E). The absence of a phenotypic effect upon *kraken* overexpression is surprising, given its elevated expression in *D. sechellia* and in evolved *D. simulans* but this could reflect either insufficient overexpression levels (Extended Data Fig. 4c) and/or genotype-by-genotype interactions that are not recapitulated when manipulating a single locus in a *D. melanogaster* background.

A key question is whether these genes contribute to OA resistance of *D. sechellia*. Implementing CRISPR/Cas9-based genome-engineering in this species [22, 27], we generated and validated null mutants in these genes in *D. sechellia* (Fig. S4B) and tested their resistance to OA in the plate assay. Although we did not detect any significant difference to genetically-matched wild-type flies (Fig. 5D), this was anticipated given the polygenic nature of OA resistance in *D. sechellia* [28, 34-36], and the modest or absent phenotypes of equivalent mutants in *D. melanogaster* (Fig. 5B). We reasoned that more subtle differences might be detected by monitoring *D.*

400 *sechellia*'s survivability over longer periods of time, when repeated contact with OA
401 was enforced by blending it with the food. Indeed, we observed higher mortality rates
402 of both *kraken* and *Alkbh7* mutants compared to genetically-matched wild-type flies
403 over a period of 6 days (Fig. 5E). By contrast, both mutants survived equally well as
404 controls in the absence of OA over this time period (Fig. 5E). These data provide the
405 first evidence for the contribution of specific genes to OA resistance in *D. sechellia*.
406 However, future allelic swap studies with *D. simulans* or *D. melanogaster* would be
407 required to assess if and how they contribute to species-specific differences in OA
408 sensitivity (e.g., through structural and/or regulatory differences in the genes).
409
410 **Discussion**
411
412 Identifying specific loci underlying phenotypic evolution is challenging due to the
413 polygenicity of the vast majority of adaptive traits. Investigating the evolution of
414 chemical toxicity is an attractive model trait because of the relatively easy phenotypic
415 read-out. To gain insight into the genetic basis of susceptibility and resistance to a
416 natural toxin, OA, we have intersected the results of selection over vastly different
417 timescales: weeks in cell lines, years in experimentally-evolved laboratory strains, and
418 millennia during speciation.
419 Unlike insecticides that activate or antagonize broadly-expressed neuronal ion
420 channels to cause organismal death, the mode of action of OA appears to be
421 mechanistically very different, capable of inducing death at the level both of whole
422 organisms and of cultured cells. Beyond insect cell lines (our work and [33]), OA is
423 also toxic to budding yeast [88] and bacteria [89, 90], suggesting that this toxin acts
424 via a more fundamental cellular mechanism. Indeed, a genome-wide, knock-out
425 screen in yeast for genes required for tolerance of OA highlighted contributions of
426 several core cellular processes, such as vesicle trafficking, mitochondrial function or
427 chromatin regulation [88], reminiscent of our CRISPR screen results. Consistent with
428 this idea, the results of our experimental evolution and CRISPR screening, as well as
429 earlier mapping efforts in *D. sechellia*/*D. simulans* hybrids [28, 34-36] revealed multiple
430 genomic regions/genes contributing to the resistance to OA.
431 Considered separately, the experimental evolution and cell-based CRISPR
432 screens identified many candidates. That there was very limited overlap between the
433 screen results is unsurprising given their very different designs. Notably, the loss-of-
434 function, cell-based screen cannot identify genes that are not expressed in the cell
435 line, nor those essential for cell viability in the absence of OA. Nevertheless, both lists
436 are likely to contain multiple genes relevant to understanding OA susceptibility and
437 resistance, and hence mode(s) of action of OA. Here, we focused on two candidates,
438 *kraken* and *Alkbh7*, which were common hits of both screens. We note, however, that
439 neither of these genes were identified in previous studies of OA resistance [28, 34-
440 36], which could reflect several factors, such as differences in OA toxicity assay
441 conditions, developmental stages, species and molecular readouts (DNA or RNA), and
442 the evident multilayered, polygenic nature of this trait (discussed further in [91]).
443 The (partial) necessity, but insufficiency, of *kraken*, and lack of necessity but
444 sufficiency of *Alkbh7* to confer OA resistance implies fundamentally different roles of
445 these proteins. The expression of *kraken* in the gut and renal system supports a
446 hypothesis of this protein as a detoxification enzyme, although future work will be
447 necessary to determine whether Kraken has a direct role in OA degradation. The mild
448 loss-of-function phenotype of *kraken* in both *D. melanogaster* and *D. sechellia* likely
449 reflects substantial redundancy of defense mechanisms, including other potential

detoxification mechanisms, especially given the plethora of enzymes expressed in the fly gut and Malpighian tubules [75, 92]. Indeed, extremely few associations between such enzymes and substrates and/or physiological roles have previously been defined. One exception is the Cytochrome P450 gene *Cyp6g1*, which has a key role in Malpighian tubules in conferring resistance to DDT [77]. Our data on Kraken therefore provide one of the first links between a natural toxic chemical and a putative detoxification enzyme.

The mitochondrial Alkbh7 enzyme likely has a distinct role in OA resistance. Insights from mammals are informative: mice lacking *Alkbh7* are obese, with higher body fat levels, proposed to result from a defect in metabolism of fatty acids [72]. Molecularly, one study suggests that Alkbh7's metabolic role stems from its regulation of RNA editing to control mitochondrial protein translation [71], although it is unclear whether this is its only function. While the *Alkbh7* loss-of-function phenotype is mild in drosophilids, it is striking that artificially elevated expression in *D. melanogaster* is sufficient to produce *D. sechellia*-like levels of OA resistance. This result implies a potent effect of Alkbh7 on the mitochondrial metabolic network to modulate toxin resistance, perhaps through control of fatty acid metabolism; such ideas can be tested in future studies. Intriguingly, our implication of Alkbh7 in the niche adaptation of *D. sechellia* has a potential parallel in bats, where this gene displays evidence of positive selection during dietary diversification of different species [93].

Taken together, our work has demonstrated the power of orthogonal selection approaches in the laboratory to understand the evolution of toxin resistance, which might be relevant to understand a historic paradigm of this phenomenon [23], as well as offering clues as to how the natural toxin OA exerts its lethal effects. Moreover, the methodology – notably cell-based CRISPR screening, which can be extended to both gain-of-function approaches [94] and mosquito cell lines [95] – is applicable to explore the molecular mechanisms of susceptibility and resistance of insects to the wide-range of artificial toxic chemicals used in the environment [12]. Such knowledge will have application in the future characterization of novel types of insecticides.

Author contributions

M.M. and R.B. conceived the project. M.M. designed and performed the experimental evolution experiment, generated and analyzed all transgenic and mutant flies, and performed the comparative gene expression and molecular evolutionary analyses. S.C. generated the base populations and provided technical support for the experimental evolution experiment. M.M., R.B., R.V. and S.E.M. designed the cell screen. S.G. and R.V. tested and defined cell screen assay conditions. S.G. and M.B. performed the CRISPR/Cas9 screen and prepared DNA for sequencing. R.V. analyzed the screen data, and M.M., R.B., R.V., S.E.M and N.P. contributed to interpretation of screen results. J.D. and C.R. performed histological experiments, supervised by D.H. M.M. and R.B. wrote the paper with input from other authors. All authors approved the final version of the manuscript.

Acknowledgements

We thank Tadeusz Kawecki and Ana Marija Jakšić for advice on experimental evolution experiments, Tina Lenče and Jean-Yves Roignant for the gift of the *D. melanogaster Alkbh7* sgRNA transgenic line, the Bloomington *Drosophila* Stock Center (NIH P40OD018537) and the Vienna *Drosophila* Resource Center for *D.*

melanogaster stocks, Christian Schlötterer and Neda Barghi for the *D. simulans* lines, Blaise Tissot-Dit-Sanfin for maintenance of *M. citrifolia* plants and Ambra Masuzzo for providing noni fruits. Figure icons were created with Biorender.com. We are grateful to Tadeusz Kawecki, Philippe Reymond, Michael Shahandeh and members of the Benton laboratory for discussions throughout the project and comments on the manuscript. D.H. is funded by an ERC Starting Grant (101117267). N.P. is an Investigator of the Howard Hughes Medical Institute. Research in R.B.'s laboratory was supported by the University of Lausanne, an ERC Advanced Grant (833548) and the Swiss National Science Foundation (310030_219185).

Declaration of interests

The authors declare no competing interests.

Data availability

Source data is provided in the supplementary data files. All other data, new materials and code are available from the corresponding author upon reasonable request.

Materials and Methods

Drosophila strains and husbandry

Drosophila stocks were maintained on standard wheat flour–yeast–fruit–juice medium under a 12 h light:12 h dark cycle at 25°C. Wild-type strains, as well as mutant and transgenic lines used in this study are listed in Table S1. Experimental evolution was performed using a subset of published *D. simulans* isogenic strains [53] (Table S2).

Generation of mutant and transgenic flies

Loss-of-function mutants: we generated *D. melanogaster* mutants for *kraken* and *Alkbh7* using in vivo CRISPR/Cas9 mutagenesis [96] by crossing flies expressing Cas9 in the germline (*nanos-Cas9*) to those expressing sgRNAs targeting the desired gene. For *kraken*, we used *P{hsFLP}1,y^{1},w^{1118};P{HD_CFD02326}attP40* as a source of sgRNAs driven by *Act5C-GAL4*; for *Alkbh7*, we used the CG14130 sgRNA line described in [97]. Mutants were validated by genomic PCR and sequencing (Fig. S4B).

D. sechellia mutant flies were generated by WellGenetics Inc. using modified methods of [98]. For *kraken* (GM16772/LOC6617307) the upstream sequence CTGTAATGCTGGACGCGGAT[TGG] (PAM in square brackets) and the downstream sequence CCTGGTCACCCCGGATCGAG[TGG] were cloned separately into a U6 promoter plasmid. The 3xP3-RFP cassette, which contains a floxed 3xP3-RFP flanked by a 5' homology arm (999 bp: from -1043 nt to -45 nt relative to the *kraken* ATG) and a 3' homology arm (1063 bp: +1388 nt to +2450 nt) were cloned into *pUC57-Kan* as a donor template for repair. For *Alkbh7* (GM25300/LOC6605344) the upstream sequences ACTAATAAGAGTCGACCGGG[TGG], AACAGTCAAACAGCTGTTCC[TGG] and downstream sequences GTCCCAGTTTCAGGGTACGC[TGG], ATGGAAATGCGACGTGTCCT[GGG] were cloned separately into a U6 promoter plasmid. The 3xP3-RFP cassette, which contains a floxed 3xP3-RFP, 5' homology arm (1037 bp: -1121 nt to -85 nt relative to the *Alkbh7* ATG) and 3' homology arm (964 bp: +940 nt to +1903 nt) were cloned into *pUC57-Kan*. DNA plasmids encoding the sgRNAs and the corresponding homology-directed repair template were microinjected into embryos of *Dsec.07 pBAC(nos:Cas9,3P3-YFP)* [27]. In both cases, a single F1 line expressing the 3xP3-RFP selection marker was obtained, which was validated by genomic PCR and sequencing (Fig. S4B).

D. melanogaster and *D. sechellia* mutant flies were backcrossed for six generations to Canton-S and *Dsec*07, respectively.

Transgenic flies: to make *kraken-Gal4*, we amplified putative 5' regulatory regions for *D. melanogaster kraken* (encompassing the entire intergenic region between the last exon of the upstream gene and the first exon of *kraken*) from *w^{1118}* genomic DNA (all oligonucleotides are listed in Table S6), and subcloned the PCR product into *pGal4-attB* [99]. To make the *UAS-Dmel\kraken* and *UAS-Dsec\kraken* transgenes, we amplified the regions from the start to the stop codons from *w^{1118}* and *Dsec07* genomic DNA and subcloned them into a *pUASTattB* vector [100]. Transgenic flies were obtained by phiC31 integrase-mediated transgenesis into attP2 performed by BestGene Inc. and verified by sequencing.

568 **Toxin resistance assays**
569
570 *Tube assay*: resistance to OA for the experimental evolution setup was scored
571 adapting a previous protocol [34, 43]. In brief, flies were lightly anesthetized with CO_2
572 and ~50 animals were placed in standard cultured tubes (25 mm diameter × 95 mm
573 height, Milian SA) and provided with 1 ml of a 3% glucose solution on a tissue paper
574 (Kimtech, 7552). 3 μl of pure OA (Sigma-Aldrich, CAS 124-07-2, C2875) were directly
575 pipetted onto the tissue while the flies remained anesthetized. Direct contact with the
576 acid was scrupulously avoided until all flies recovered. The fraction of alive flies was
577 scored after 24 h. For the experimental evolution experiment, the surviving flies were
578 collected (see below).
579
580 *Plate assay*: to measure OA resistance in the genetically-manipulated *D.*
581 *melanogaster* and *D. sechellia* strains, we used a plate assay [28]. Compared to the
582 experiments in tubes, this assay offers greater temporal resolution and higher dynamic
583 range (in contrast to the experimental evolution experiment, we did not strive to
584 maintain a certain fraction of surviving flies), as well as greater statistical power from
585 tests of fewer animals. Ten female flies were lightly anesthetized with CO_2 and placed
586 in a Petri dish (60 mm diameter × 15 mm height) in which 2 μl of pure OA (or 30 μl for
587 *D. sechellia*) were spotted in the center of the underside of the lid. The flies were
588 allowed to recover (typically <5 min), and immobile/dead were scored every 5 min for
589 1 h. The survival curves were calculated and plotted using the R package survminer
590 (10.32614/CRAN.package.survminer). The statistical analysis was conducted fitting a
591 mixed effect Cox regression model as implemented in the R package coxme [101].
592 Genotype was included as a fixed effect, and replicate dishes were modelled as a
593 random effect.
594
595 *Noni resistance assay*: noni fruit was collected from *M. citrifolia* plants grown in the
596 University of Lausanne greenhouses. Five ripe (pale yellow) fruits were homogenized
597 together, and the pulp puree frozen at -20°C. The noni assay was performed with the
598 same setup as for tube assay described above, using 2 g of thawed noni puree placed
599 on the bottom of the tube. Noni contains 3.06 g OA/kg fruit [102]; assuming an
600 equivalent quantity in our fruits, 2 g noni pulp should contain 6.73 μl OA, approximately
601 double the amount used for the experimental evolution between generations 1 to 25,
602 and two thirds of the concentration during generations 30-50.
603
604 *Long-term assay*: To assess the long-term effects of OA (i.e., beyond 24 h), two-day-
605 old mated female flies were individually housed in vials containing food supplemented
606 with OA. OA was thoroughly mixed into Formula 4-24® Instant Drosophila Medium,
607 Blue (Carolina Biological Supply) to a final concentration of 0.9%. Fresh OA-containing
608 food was provided every two days, and mortality was monitored daily until all flies in
609 the experimental group died. Statistical analysis was run as for the plate assay.
610
611 *Fecundity* and *developmental viability*: 10 mated adult female flies (4-7 days old) were
612 allowed to lay eggs for 24 h on agar plates (2% agar, 2.5% sucrose in a 1:3 grape
613 juice:water mix). The total number of eggs on each plate was counted. From each
614 plate, 100 eggs were collected and transferred to standard fly food. Emerging adults
615 were counted and removed daily until no further flies emerged.
616

Lifespan: Two-day-old mated female flies were individually housed in vials containing standard fly food and monitored daily until death. Fresh food was provided every 2–3 days.

Experimental evolution

Selection procedure: to generate the base populations for experimental evolution, we used 93 isogenic lines (Table S2) of the *D. simulans* panel established by [53]. We first crossed each of the lines together following a round-robin design to ensure equal contribution of each line to the starting genetic pool; 5 males and 5 females from each cross were then combined in ten independent replicates (Fig. S1D). These ten populations were maintained in bottles (57 mm diameter ×103 mm height, Milian SA) of standard fly food at high population size (~1000 individuals per population) to minimize the effect of genetic drift. Each population was subdivided into smaller groups of ~250 individuals per bottle (i.e., four bottles per population) to reduce competition. At each generation, individuals from the four bottles were mixed to prevent bottle-specific genetic drift. The experimental evolution experiment was initiated 10 generations after the creation of these populations.

From each population, 500 individuals (2-7 day old) were randomly collected and further split into 10 groups of 50 mixed-sex flies (i.e., ~5000 flies total). Each group of 50 flies was then placed under selection with 3 μl of OA for 24 h (as described above); the survivors were collected and allowed to breed in food vials for 6-8 days, transferring them to fresh vials every 2-3 days. In the next generation, 500 flies per population were collected only from the last seeded vial, to maximize the probability of these flies being parented by the OA-resistant males and females and not offspring generated by females that mated before the selection procedure. These flies were then subjected to the same selection procedure. This process was repeated for 25 generations, and survivability was scored at each generation.

A second round of experimental evolution was performed by increasing the amount of OA to 9 μl between generations 31 and 50; this amount was determined empirically during generation 26-31 to re-establish the survivability close to the initial level (Fig. 1G). All successive generations were maintained following the same procedure.

As a control, the ten original populations were maintained in parallel in the same conditions as previously described (i.e. ~1000 individuals per population subdivided into smaller groups of ~250 individuals per bottle).

DNA sequencing and analysis: we pooled ~200 individual flies per population per time point (generations 0, 25 and 50). Genomic DNA (and total RNA for qPCR, as described below) extraction was performed using the Quick DNA/RNA Miniprep plus kit (Zymo Research, D7003) following the manufacturer's instructions. Samples were stored at -80°C.

We performed paired-end whole genome sequencing on an Illumina NovaSeq 6000 following Nextera DNA flex preparation, at an average coverage of >200×. Read quality was assessed with Fastqc [103] and adapters were trimmed with Trimmomatic [104] providing the Nextera adapters sequences. We aligned the trimmed reads to the *D. simulans* reference genome (NCBI GCF_016746395.2) using bwa-mem [105] and the results were sorted and compressed using samtools [106]. Duplicate reads were then removed with Picard MarkDuplicates (https://broadinstitute.github.io/picard/).

666 Unplaced scaffolds and alignments with mapping quality <20 were removed with
667 samtools. We then combined all the aligned sequences into a samtools pileup file and
668 split this into individual chromosomes. Regions with zero coverage were removed with
669 custom scripts. Sync files for each chromosome were generated with the
670 mpileup2sync.pl script implemented in Popoolation2 [107] and used as input files for
671 all the Evolve-and-Resequence analyses.
672 To explore the genetic variation of our samples we generated vcf files with
673 vcftools 0.1.17 [108] and binary files with plink1.9 [109, 110]. We calculated the
674 identity-by-state (IBS) matrix as implemented in plink1.9 (function -distance 'ibs').
675
676 *Evolve-and-Resequence analysis*: allele frequency changes between generation 0
677 and generations 25 or 50 were investigated via Popoolation2 and Bait-ER [60].
678 We performed a Cochran–Mantel–Haenszel (CMH) test via the cmh-test.pl
679 script implemented in Popoolation2 comparing base and evolved populations with
680 parameters --min-count 12 and --min-coverage 50. We estimated a plausible
681 significance threshold approximating the top 0.000005% of the simulated p-values
682 based on the effect of genetic drift alone by running genome-wide forward simulations
683 under neutrality in mimicree2 [111]. We used the same genomic information from the
684 original 93 lines as input and the average real survivability ratios observed throughout
685 the selection experiment as –selection-regime parameter and 500 as –population-size
686 parameter (Fig. S1B).
687 For Bait-ER, the input sync files were filtered to match --min-count 12 and --
688 min-coverage 50 and scaled to a uniform coverage of 100 as in [58] to avoid numerical
689 problems due to high coverage. We ran the algorithm on each chromosome
690 separately. Bait-ER models the effect of genetic drift directly using estimates of the
691 effective population size (N_e). We calculated N_e estimates for each chromosome in
692 each replicate population with the R package poolSeq [112] and used the median
693 value of each chromosome as "Population_size" parameter (Table S3). Other
694 parameters included "Sampling correction" = 0 (i.e., binomial) and "Prior parameters"
695 =0.001 (shape), 0.001 (rate). Data points were considered significant when the
696 absolute natural logarithm of the Bayes Factor exceeded the conservative threshold
697 of log(0.99/0.01) as in [60].
698 The linkage disequilibrium decay in the 10 evolved populations was analysed
699 across the whole genome using PopLDdecay [113]. Linkage disequilibrium maps of
700 focal regions were generated with LDBlockShow [114] using the Solid Spine method
701 based on D' statistics originally introduced in Haploview [115].
702 For the genome-wide screen, all the genes within selection blocks were
703 considered potential candidates. Based on the results on linkage disequilibrium decay,
704 a block was arbitrarily defined as a genomic interval where at least two significant
705 SNPs were found at a distance less than 50 kb. Consecutive SNPs respecting this
706 definition were grouped into larger blocks and visualized as circus plots using the R
707 package circlize [116]. The genes overlapping such blocks were extracted using
708 BEDtools intersect [117]. Orthologous genes in *D. melanogaster* were retrieved using
709 a combination of the OMA browser [118], DAVID [119, 120], BLAST [121], NCBI [122]
710 and FlyBase [123], and subsequent manual curation.
711
712

713 **Genome-wide CRISPR/Cas9 screen in *D. melanogaster* S2R+ cells**
714
715 *Library design, construction, and transfection*: design, synthesis and cloning of the
716 genome-wide sgRNA library are previously described [124]. To determine sub-lethal
717 doses of OA appropriate for screening, the effect of a gradient of OA supplementation
718 into media on attP+ Cas9+ cells was determined using CellTiter-Glo (Promega)
719 following the manufacturer's protocol; luminescence measurements were made using
720 a Spectramax Paradigm (Molecular Devices).
721 For the screen, attP+ Cas9+ *D. melanogaster* S2R+ cells (RRID:CVCL_UD30)
722 were transfected using Effectene and a 1:1 molar ratio of sgRNA library and pIntAC,
723 as described [124]. Briefly, transfected cells were distributed in 5 ml aliquots across
724 ten 100 mm culture dishes and incubated overnight, then 5 ml of fresh media was
725 added to each plate. After 4 days, cells were subjected to puromycin selection for 12
726 days, subculturing every 4 days, to establish a stable sgRNA-expressing cell library.
727 For the selection assay, 2.5×10^7 cells were plated onto four 15 cm diameter cell
728 culture dishes to ensure sufficient sgRNA coverage, with each sgRNA represented
729 approximately 1000 times [124].
730
731 *OA selection assay*: library-containing cells were left untreated or treated with either
732 0.075 µg/µl or 0.05 µg/µl OA (Sigma-Aldrich, CAS 124-07-2, C2875). Each replicate
733 consisted of four 15 cm diameter cell culture dishes seeded with cells, achieving
734 ~1000× coverage. Two replicate sets were included for each of the two OA treatment
735 conditions and one replicate for the untreated control. During the selection period, cells
736 were split 1-2 times per week to avoid overgrowth. An aliquot of cells from each
737 replicate was collected and stored for genomic prep following each split and at the end
738 of the assay. The final samples were collected after 26 days of OA or control treatment.
739
740 *Genomic DNA extraction and sequencing*: genomic DNA was extracted from samples
741 collected during the screen and at the final endpoint using a Zymo Quick-gDNA
742 Miniprep kit (Zymo Research, D3025). DNA fragments containing the sgRNA
743 sequences were amplified by PCR using what we calculated to be at least 1000
744 genomes per sgRNA as templates for each sample. PCR fragments were in-line
745 barcoded using a previously described approach [42, 124, 125]. Next-generation
746 sequencing (Illumina NextSeq) was performed at the Biopolymers Facility at Harvard
747 Medical School (RRID:SCR_007175). FASTQ sequence files were demultiplexed
748 using in-line barcodes as sequence tags as described previously [42, 124]. All further
749 processing was done using MAGeCK (v.0.5.4 and 0.5.9.4) [126] to generate readcount
750 files and perform robust rank aggregation (RRA) analysis and Maximum likelihood
751 estimation directly on the generated readcount files using default parameters. Plasmid
752 readcounts were used to normalize RRA values.
753 After removing essential genes, we considered as candidates the top 5% genes
754 overlapping across the two conditions that were over-represented or under-
755 represented in the OA-exposed cell populations compared to the control cells (i.e.,
756 representing "susceptibility" and "resistance" genes, respectively). The gene-set
757 enrichment analysis was performed with PANGEA 1.1 Beta using the *Drosophila* GO
758 subsets (GO slim) and a p-value cut-off of 0.05 [127].
759
760

761 **qPCR analysis**
762
763 Reverse transcription of total RNA (extracted as described above) was performed with
764 the PrimeScript first-strand cDNA synthesis kit (Takara, 6110A) following the
765 manufacturer's instructions. Samples were stored at -80°C.
766 Quantitative PCR experiments were conducted on a QuantStudio 6 Flex Real-
767 Time PCR System using the PowerUp™ SYBR™ Green Master Mix following the
768 manufacturer's instructions. Each sample consisted of at least three biological
769 replicates. We performed three technical replicates per reaction and removed
770 replicates where the standard deviation was equal or greater than 0.5. We determined
771 target specific amplification efficiencies for each gene (Table S4, Table S6). Relative
772 quantification and statistical analysis were performed with the software qbase+ v3.4
773 [128] using Calibrated Normalized Relative Quantities (CNRQs). Normalization was
774 performed on the geometric mean of the relative quantities for the three most
775 commonly-used reference targets: *Act5C*, *αTub84B* and *RpL32* [129]. We assessed
776 reference target stability by the geNorm expression stability value (M) and the
777 coefficient of variation of the normalized reference gene relative quantities (CV) (Table
778 S5) [128, 130].
779
780 **Population genetics analysis**
781
782 Publicly available WGS data for 41 *D. sechellia* and 20 *D. simulans* genomes were
783 downloaded from the Sequence Read Archive (PRJNA215932 and PRJNA395473)
784 [82, 131] with fasterq-dump [132]. FASTQ files were processed and aligned as
785 described in the "Experimental evolution" section. Aligned files were phased with
786 samtools phase [106] and the genes of interest were extracted with pilon [133] to find
787 variation among populations. CDS sequences were aligned based on amino acid
788 information using translatorX [134]. Gene sequences were aligned using Clustal
789 Omega [135]. McDonald-Kreitman test, Tajima's D and Fu and Li's statistics were
790 calculated in DNAsp 6 [136]. FUBAR analyses (Fast, Unconstrained Bayesian
791 AppRoximation) were performed in HyPhy 2.5 [137]. The consensus sequences of *D.*
792 *simulans* and *D. sechellia* were computed by SnapGene (www.snapgene.com) "export
793 consensus" with a >50% threshold. The sequence alignment and secondary structures
794 were depicted using ESPript 3 [138].
795
796 **Histology**
797
798 Intestinal tissues and Malpighian tubules were dissected in phosphate-buffered saline
799 (PBS) and fixed at room temperature for 45 min in PBS, 4% paraformaldehyde.
800 Staining of muscle was performed using a mushroom phalloidin AlexFluor647 probe
801 (Life Technologies, A22287, diluted 1:200 in PBS + 0.2% Triton X-100) for 45 min.
802 Samples were mounted in Vectashield with DAPI (Vector Labs, H-1200-10).
803 A Leica STELLARIS-DIVE confocal microscope was used to generate all
804 confocal images, using a water-immersive 40× objective or an oil-immersive 63×
805 objective. The images were acquired using both HyDs as well as PMTs tailored for the
806 fluorophores of each sample.
807
808 **Statistics and reproducibility**
809
810 Data were analysed and plotted in R 4.0.3 (R Core Team 2021).

Figure legends

Fig. 1. Experimental evolution of *D. simulans* with increased OA resistance.
A Phylogeny of the drosophilid trio studied in this work. Mya: million years ago.
B Schematic of octanoic acid (OA) resistance tube assay (see Materials and Methods) and selection procedure.
C Species differences in resistance of *D. simulans* 04 and *D. sechellia* 07 strains to 3 µl OA.
D Schematic illustrating the generation of the base populations for Evolve-and-Resequence (see Materials and Methods).
E Changes in survivability to 3 µl OA of the ten *D. simulans* populations over generations 0-25. Box plots represent the interquartile ranges of the survivability for all the replicates; blue lines shows the generalized additive model (GAM) regression; grey shading represents the 95% confidence interval for the mean responses at each time point.
F As in **E** for the 10 populations combined.
G OA resistance for the combined 10 populations with increasing OA volumes during generation 25-31.
H As in **E**, with 9 µl OA during generations 31-50.
I As in **H** for the 10 populations combined.
J Comparison of OA resistance of unevolved *D. simulans* (i.e., the original 10 outcrossed populations maintained on standard fly food; see Materials and Methods), evolved (generation 50 (G50)) *D. simulans*, and *D. sechellia*.
K-N Phenotypic comparison of unevolved and evolved (G50+) *D. simulans* populations 1-3 for **K** noni resistance (G54), **L** fecundity (G54), **M** developmental viability (G57), and **N** lifespan (G57). For **K-M**, significance was assessed using the unpaired t-test correcting for multiple testing. NS $p>0.05$, ***$p<0.001$. For **L-N**, tests were run on standard media. No significant differences between unevolved and evolved flies were detected using the Cox regression model. See Materials and Methods for details and Data S1 for raw data.

Fig. 2. Dynamic directional selection of allele frequencies during experimental evolution.
A Schematic of the analysis.
B Identity-by-state analysis. The heatmap depicts the proportion of shared alleles between populations, with values ranging from 0.8 (blue, lowest shared proportion) to 1 (red, identical alleles).
C,D Manhattan plot showing the probability of allele frequency changes between **C** generations 0 and 25, and **D** generations 0 and 50. Top panel shows the results of the CMH test; orange dots indicate significant variants at a threshold of $-\log_{10}(p) \geq 40$, a conservative approximation of the top 0.000005% simulated SNPs under neutrality (Fig. S1B). Bottom panel shows the absolute log-Bayes factors estimated by Bait-ER; orange dots indicate significant variants at the conservative threshold log(0.99/0.01). Raw data in Data S2.
E Circular plots illustrating overlap of predicted genomic regions under directional selection between the two statistical approaches at generation 25 (left) and generation 50 (middle), or between generation 25 and 50 (right). Raw data in Data S3.

861
862 **Fig. 3. A genome-wide pooled CRISPR screen for OA susceptibility and**
863 **resistance genes in *D. melanogaster* cultured cells.**
864 **A** Left: images of S2R+ cells cultured in the absence or presence of OA. Scale bar,
865 50 μm. Right: dose-dependent effects of OA on S2R+ cell viability, through
866 measurement of ATP levels with a CellTiter-Glo assay (see Materials and Methods).
867 **B** Schematic of CRISPR screen design.
868 **C** Top 5% "susceptibility" genes (see Materials and Methods and Data S4). Box plots
869 represent the interquartile ranges for all the replicates of the Z-scores obtained through
870 maximum likelihood estimation to test for selection. The inset (bottom-right) displays
871 the gene-ontology analysis of these genes.
872 **D** As in **C**, for the top 5% "resistance" genes.
873 **E** Venn diagram showing the overlap between all the genes identified in the
874 experimental evolution at G25 and G50, and genes associated with susceptibility and
875 resistance in the CRISPR screen.
876 **F** Subportion of the top panel (CMH test) of the Manhattan plot from Fig. 2C for
877 chromosome arms 2L and 3L, with the red arrows highlighting the regions magnified
878 on the right showing the significant variants and linkage disequilibrium blocks
879 (triangles) for *kraken* and *Alkbh7* (CG14130).
880
881
882 **Fig. 4. Sequence and expression analyses of *kraken* and *Alkbh7*.**
883 **A** AlphaFold protein models for *D. melanogaster* Kraken (AF-O18391-F1-model_v4)
884 and Alkbh7 (AF-Q9VTP1-F1-model_v4) [139, 140], highlighting a conserved putative
885 catalytic serine in the active site in Kraken, and the mitochondrial targeting sequence
886 in Alkbh7 (predicted by MitoFates [79]). See also Fig. S2A,B.
887 **B** Tissue-specific expression of *D. melanogaster kraken* and *Alkbh7* from bulk RNA-
888 sequencing data from the Fly Atlas 2.0 [141].
889 **C** tSNE plots illustrating *D. melanogaster kraken* expression in single-cell
890 transcriptomes of the Malpighian tubules and gut from the Fly Cell Atlas (Stringent 10x
891 datasets) [74].
892 **D** Left: whole gut of an adult female expressing GFP (green, endogenous fluorescence
893 signal) under the control of a *kraken* promoter Gal4 driver (genotype: *kraken-*
894 *Gal4/UAS-FB1.1*, aged 4-5 days); muscles are labelled by phalloidin (red) to reveal
895 tissue morphology. Imaging was performed with a 40× objective; scale bars, 200 μm.
896 Right: higher magnification images of the regions indicated with dashed boxes.
897 Imaging was performed with a 63× objective; scale bars, 100 μm.
898 **E** Expression levels of the *kraken* and *Alkbh7* in the indicate strains of adult female *D.*
899 *sechellia* and *D. simulans* measured by qPCR (Data S6). Expression is represented
900 as calibrated and normalized relative quantities (CNRQ). Significance was assessed
901 using the unpaired t-test correcting for multiple testing and represented using letter
902 codes.
903 **F** Expression levels of *kraken* and *Alkbh7* in the indicated strains and species at adult
904 or larval stages as indicated from published RNA-seq data [85-87]. Significance was
905 assessed using the unpaired t-test correcting for multiple testing and represented
906 using letter codes when multiple replicates were present in the original data.
907 **G** Expression levels of the candidate genes at generation 0, 25 and 50 of the
908 experimentally-evolved *D. simulans* populations measured by qPCR. Significance was
909 assessed using the paired t-test correcting for multiple testing and represented using
910 letter codes.

Fig. 5. Functional validation of the contribution of *kraken* and *Alkbh7* to OA resistance.

A-C Survival curves of adult flies of the indicated genotypes tested in the plate assay with 2 µl OA for *kraken* and *Alkbh7* RNAi (**A**), mutants (**B**) and overexpression (**C**). Each replicate consisted of 10 2-7 day-old female flies. N replicates ≥6, n flies ≥60. Shading indicates 95% confidence intervals. Genotypes: **A** *act-Gal4/+;UAS-kraken^{RNAi}/+, UAS-kraken^{RNAi}/+, act-Gal4/+, act-Gal4/+;UAS-Alkbh7^{RNAi}/+, UAS-Alkbh7^{RNAi}/+, act-Gal4/+* **B** *kraken^{1}*, *Alkbh7^{1}, Canton-S* (wild-type genetic background control for both mutants), **C** *act-Gal4/+;UAS-kraken$^{o/e}$/+*, *UAS-kraken$^{o/e}$/+*, *act-Gal4/+*, *act-Gal4/+;UAS-Alkbh7$^{o/e}$/+*, *UAS-Alkbh7$^{o/e}$/+*, *act-Gal4/+*.

Significance is based on the resulting mixed effect Cox regression model. NS $p>0.05$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$. The same data for *D. sechellia* are shown in both plots in **C**.

D Survival curves of wild-type (*Dsec07*), and *kraken* (*Dsec\kraken^{RFP}*) and *Alkbh7* (*Dsec\Alkbh7^{RFP}*) mutant *D. sechellia* in the plate assay with 30 µl OA. Statistics as in **A**.

E Long-term survivability of the same genotypes as in **D** (n = 60 per condition) in control conditions (no OA) (left) and food supplemented with 0.9% OA (right). Statistics as in **A**.

Raw data in Data S7.

933 **Supplementary Information**
934
935 **Supplementary Tables**
936
937 **Table S1. *Drosophila* strains.**
938

Genotype	Reference/Source
D. sechellia 07 (*Dsec07*)	*Drosophila* Species Stock Center (DSSC):14021-0248.07
D. simulans 04 (*Dsim04*)	DSSC:14021-0251.004
D. melanogaster Canton-S *(DmelCS)*	
y[1] w[*]; P{Act5C-GAL4-w}E1/CyO (act-Gal4)	RRID:BDSC_25374
y[1] w[1118] (used as background for act-Gal4)	RRID:BDSC_6598
y[1] w[67c23] (background of Alkbh7$^{o/e}$)	RRID:BDSC_6599
y[1] w[67c23]; P{y[+mDint2] w[+mC]=EPgy2}EY05855 (*Alkbh7$^{o/e}$*)	RRID:BDSC_16679
y,w[1118];P{attP,y[+],w[3']} (background of UAS-kraken^{RNAi} and UAS-Alkbh7^{RNAi})	VDRC:v60100
P{KK100356}VIE-260B (UAS-kraken^{RNAi})	VDRC:v105604
P{KK102727}VIE-260B (UAS-Alkbh7^{RNAi})	VDRC:v109669
y[1] M{w[+mC]=nanos-Cas9.P}ZH-2A w[*] (nanos-cas9)	RRID:BDSC_54591
P{hsFLP}1, y1 w1118; P{HD_CFD02326}attP40/CyO-GFP (sgRNA line for Dmel\kraken CRISPR mutant)	VDRC:v342579
CG14130 cris (gRNA 1+2)/CyO *(sgRNA line for Dmel\Alkbh7 CRISPR mutant)*	[97]
y[1] w[67c23];;UAS-kraken (kraken$^{o/e}$)	This study
y[1] w[67c23];;UAS-Dsec\kraken (Dsec\kraken$^{o/e}$)	This study
y[1] w[67c23]; P{y[+t7.7]=CaryP}attP2 (background of kraken$^{o/e}$)	This study
Dmel\kraken-Gal4	This study
UAS-FB1.1 (GFP reporter)	[142]
Dmel\kraken^{1}	This study
Dmel\Alkbh7^{1}	This study
Dsec\kraken^{RFP}	This study
Dsec\Alkbh7^{RFP}	This study

939
940
941 **Table S2. *D. simulans* lines used to generate the base populations.**
942 *- see separate Excel file.*
943
944

945 **Table S3. Effective population size estimates from the experimentally-evolved**
946 **populations.**
947

	2L	2R	3L	3R	4	X
pop1	80.65681	101.5796	107.7145	100.5748	369.3642	107.3616
pop2	135.7243	56.29977	76.40593	91.41878	333.638	53.66758
pop3	74.08474	119.4491	95.20849	128.4216	193.5982	99.40552
pop4	83.05221	89.62312	112.3929	116.7979	171.2875	112.1581
pop5	123.0386	106.8179	107.448	81.31624	364.4287	83.55315
pop6	98.8018	77.03804	60.71772	147.1788	124.9405	75.48775
pop7	119.4038	101.4515	96.22689	114.8849	431.7489	114.6792
pop8	136.6513	176.4308	70.8519	130.9822	639.7962	78.70108
pop9	95.58733	101.7351	122.8352	78.80166	109.3757	69.50865
pop10	141.862	101.6789	107.512	83.0109	376.8733	71.34583
mean	108.8863	103.2104	95.73134	107.3388	311.5051	86.58684
median	109.1028	101.6293	101.8374	107.7299	349.0333	81.12711

948
949 **Table S4. qPCR primer efficiency.**
950

Gene	**Efficiency**	**SE**
kraken	1.998	0.041
Alkbh7	1.977	0.016
Act5C	1.976	0.022
αTub84B	2.027	0.023
RpL32	2.053	0.031

951
952 **Table S5. geNorm analysis of reference target stability.**
953 *- see separate Excel file.*
954
955 **Table S6. Oligonucleotides.**
956 *- see separate Excel file.*
957
958

Supplementary Figure legends

Fig. S1. Linkage disequilibrium decay, and comparison of simulated and observed p-value distributions in the experimentally evolved populations.
A Linkage disequilibrium decay at generations 25 (left) and 50 (right). Black curves represent the trajectory of r2 values (y-axis left) as genomic distances increase (x-axis). Red curves represent the same but for D' values (y-axis right).
B Violin plots showing the distribution of p-values from the CMH test, comparing allele frequency changes between generation 25 and 0 (left), and generation 50 and 0 (right). Blue violins represent p-values derived from forward evolution simulations, while orange violins correspond to p-values from experimentally evolved populations. Red dashed lines mark the top 0.000005% of simulated p-values, serving as the significance threshold for the experimental data.

Fig. S2. Sequence and molecular evolutionary analyses of *kraken* and *Alkbh7*.
A,B Protein sequence alignment and secondary structure of Kraken (**A**) and Alkbh7 (**B**) from the reference sequence of *D. melanogaster* (NP_477265.1 and NP_648511.2) and the consensus sequences of *D. simulans* and *D. sechellia*. Sequence data from PRJNA215932 and PRJNA395473 (see Materials and Methods). Black triangles indicate the residues where signs of selection were detected by the FUBAR analysis (see **D**).
C Results from the McDonald-Kreitman test. NI = neutrality index; alpha = proportion of substitutions driven by positive selection; p-value = p-value resulting from a two-tailed Fisher's exact test.
D FUBAR analysis to detect pervasive selection as implemented in HyPhy. position = codon position; alpha = mean posterior synonymous substitution rate; beta = mean posterior nonsynonymous substitution rate; posterior prob = posterior probability of positive selection; codons = alternative codons, Freq Dsec = frequency of the alternative codons in *D. sechellia* in the same order as codons; Freq Dsim = frequency of the alternative codons in *D. simulans* in the same order as codons.
E Survival curves of *D. melanogaster* overexpressing *Dmel\kraken* and *Dsec\kraken* exposed to 2 µl OA. Each replicate consisted of 10 2-7 day-old female flies. Genotypes: *act-Gal4/+;UAS-Dmel\kraken/+*, *act-Gal4/+;UAS-Dsec\kraken/+*. N replicates ≥6, n flies ≥60. Shading indicates 95% confidence intervals. Significance is based on the resulting mixed effect Cox regression model. NS p>0.05. Raw data in Data S5.

Fig. S3. Additional expression analyses of *kraken* and *Alkbh7*.
A Bar plots showing the expression levels of *kraken* and *Alkbh7* across various adult *Drosophila* tissues and species. Data from [143].
B Bar plots showing the expression levels of *kraken* and *Alkbh7* across various larval *Drosophila* tissues and species. Data from [85] ("medium diet" condition).
C Bar plots showing the effect of exposure to 9 µl OA (in the tube assay) on *kraken* and *Alkbh7* expression in *D. sechellia* and three evolved (G50) *D. simulans* populations.

Fig. S4. Validation of RNAi and overexpression lines and generation of mutants for *kraken* and *Alkbh7*.

A Quantitative RT-PCR analysis of RNAi-mediated knockdown of *kraken* and *Alkbh7* in *D. melanogaster*. Genotypes as in Fig. 5A.

B Schematics of the CRISPR/Cas9 design to generate *D. melanogaster* and *D. sechellia* mutants for *kraken* and *Alkbh7*. Validation was performed by PCR and Sanger sequencing. The *Dsec\kraken*^{RFP} mutant comprised a 1431 bp deletion (-44 to +1,387 nt relative to the ATG), replaced with the 3xP3-RFP cassette. The *Dsec\Alkbh7*^{RFP} mutant was a 1023 bp deletion (-84 nt to +939 nt relative to the ATG), replaced with the 3xP3-RFP cassette.

C Quantitative RT-PCR analysis of Gal4/UAS-mediated overexpression of *D. melanogaster kraken* and *Alkbh7*. Genotypes as in Fig. 5C.

Fig. S5. Control survival curves in the absence of OA.

Survival curves of adult flies of the indicated genotypes tested in the plate assay without OA. Each replicate consisted of 10 2-7 day-old female flies. N replicates ≥5, n flies ≥50. All genotypes display full viability, rendering datapoints invisible beneath the topmost displayed genotype (*UAS-kraken*^{RNAi}). Genotypes as in Fig. 5A-B.

Data S1. Source data for Fig. 1
Data S2. Experimental evolution data available at:
https://drive.google.com/file/d/1Gy3k9eYvQ8xQMjIsTCqWevRaO_b1__JY/view?usp=sharing
Data S3. Source data for Fig. 2
Data S4. Source data for Fig. 3
Data S5. Source data for Fig. S2
Data S6. Source data for Fig. 4
Data S7. Source data for Fig. 5

References

1. Erb, M., and Reymond, P. (2019). Molecular Interactions Between Plants and Insect Herbivores. Annu Rev Plant Biol *70*, 527-557.
2. Jones, A.C., Felton, G.W., and Tumlinson, J.H. (2022). The dual function of elicitors and effectors from insects: reviewing the 'arms race' against plant defenses. Plant Mol Biol *109*, 427-445.
3. Smith, J.J., Herzig, V., King, G.F., and Alewood, P.F. (2013). The insecticidal potential of venom peptides. Cell Mol Life Sci *70*, 3665-3693.
4. Umetsu, N., and Shirai, Y. (2020). Development of novel pesticides in the 21st century. J Pestic Sci *45*, 54-74.
5. Raisch, T., and Raunser, S. (2023). The modes of action of ion-channel-targeting neurotoxic insecticides: lessons from structural biology. Nat. Struct. Mol. Biol. *30*, 1411-1427.
6. Sparks, T.C., Crossthwaite, A.J., Nauen, R., Banba, S., Cordova, D., Earley, F., Ebbinghaus-Kintscher, U., Fujioka, S., Hirao, A., Karmon, D., et al. (2020). Insecticides, biologics and nematicides: Updates to IRAC's mode of action classification - a tool for resistance management. Pesticide biochemistry and physiology *167*, 104587.
7. Busvine, J.R. (1951). Mechanism of resistance to insecticide in houseflies. Nature *168*, 193-195.
8. Dong, K. (2007). Insect sodium channels and insecticide resistance. Invert Neurosci *7*, 17-30.
9. Matsuda, K., Ihara, M., and Sattelle, D.B. (2020). Neonicotinoid Insecticides: Molecular Targets, Resistance, and Toxicity. Annu Rev Pharmacol Toxicol *60*, 241-255.
10. Soderlund, D.M., and Bloomquist, J.R. (1989). Neurotoxic actions of pyrethroid insecticides. Annu. Rev. Entomol. *34*, 77-96.
11. Di Prisco, G., Cavaliere, V., Annoscia, D., Varricchio, P., Caprio, E., Nazzi, F., Gargiulo, G., and Pennacchio, F. (2013). Neonicotinoid clothianidin adversely affects insect immunity and promotes replication of a viral pathogen in honey bees. Proceedings of the National Academy of Sciences of the United States of America *110*, 18466-18471.
12. Gandara, L., Jacoby, R., Laurent, F., Spatuzzi, M., Vlachopoulos, N., Borst, N.O., Ekmen, G., Potel, C.M., Garrido-Rodriguez, M., Bohmert, A.L., et al. (2024). Pervasive sublethal effects of agrochemicals on insects at environmentally relevant concentrations. Science *386*, 446-453.
13. Clarkson, C.S., Temple, H.J., and Miles, A. (2018). The genomics of insecticide resistance: insights from recent studies in African malaria vectors. Curr Opin Insect Sci *27*, 111-115.
14. Lucas, E.R., Nagi, S.C., Egyir-Yawson, A., Essandoh, J., Dadzie, S., Chabi, J., Djogbenou, L.S., Medjigbodo, A.A., Edi, C.V., Ketoh, G.K., et al. (2023). Genome-wide association studies reveal novel loci associated with pyrethroid and organophosphate resistance in *Anopheles gambiae* and *Anopheles coluzzii*. Nat Commun *14*, 4946.
15. Catteruccia, F. (2020). Malaria-carrying mosquitoes get a leg up on insecticides. Nature *577*, 319-320.
16. Balabanidou, V., Kefi, M., Aivaliotis, M., Koidou, V., Girotti, J.R., Mijailovsky, S.J., Juarez, M.P., Papadogiorgaki, E., Chalepakis, G., Kampouraki, A., et al.

(2019). Mosquitoes cloak their legs to resist insecticides. Proc Biol Sci *286*, 20191091.

17. Balabanidou, V., Grigoraki, L., and Vontas, J. (2018). Insect cuticle: a critical determinant of insecticide resistance. Curr Opin Insect Sci *27*, 68-74.
18. Weetman, D., Djogbenou, L.S., and Lucas, E. (2018). Copy number variation (CNV) and insecticide resistance in mosquitoes: evolving knowledge or an evolving problem? Curr Opin Insect Sci *27*, 82-88.
19. Zimmer, C.T., Garrood, W.T., Singh, K.S., Randall, E., Lueke, B., Gutbrod, O., Matthiesen, S., Kohler, M., Nauen, R., Davies, T.G.E., et al. (2018). Neofunctionalization of Duplicated P450 Genes Drives the Evolution of Insecticide Resistance in the Brown Planthopper. Curr Biol *28*, 268-274 e265.
20. Gloss, A.D., Vassao, D.G., Hailey, A.L., Nelson Dittrich, A.C., Schramm, K., Reichelt, M., Rast, T.J., Weichsel, A., Cravens, M.G., Gershenzon, J., et al. (2014). Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the drosophilidae. Mol. Biol. Evol. *31*, 2441-2456.
21. Jones, C.D. (2005). The genetics of adaptation in *Drosophila sechellia*. Genetica *123*, 137-145.
22. Auer, T.O., Shahandeh, M.P., and Benton, R. (2021). *Drosophila sechellia*: A Genetic Model for Behavioral Evolution and Neuroecology. Annu Rev Genet *55*, 527-554.
23. R'Kha, S., Capy, P., and David, J.R. (1991). Host-plant specialization in the *Drosophila melanogaster* species complex: a physiological, behavioral, and genetical analysis. Proceedings of the National Academy of Sciences of the United States of America *88*, 1835-1839.
24. Dekker, T., Ibba, I., Siju, K.P., Stensmyr, M.C., and Hansson, B.S. (2006). Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila melanogaster* sibling, *D. sechellia*. Curr Biol *16*, 101-109.
25. Alvarez-Ocana, R., Shahandeh, M.P., Ray, V., Auer, T.O., Gompel, N., and Benton, R. (2023). Odor-regulated oviposition behavior in an ecological specialist. Nat Commun *14*, 3041.
26. Shahandeh, M.P., Abuin, L., Decker, L.L., Cergneux, J., Koch, R., Nagoshi, E., and Benton, R. (2024). Circadian plasticity evolves via regulatory changes in a neuropeptide gene. Nature *635*, 951-959.
27. Auer, T.O., Khallaf, M.A., Silbering, A.F., Zappia, G., Ellis, K., Alvarez-Ocana, R., Arguello, J.R., Hansson, B.S., Jefferis, G.S.X.E., Caron, S.J.C., et al. (2020). Olfactory receptor and circuit evolution promote host specialization. Nature *579*, 402-408.
28. Jones, C.D. (1998). The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. Genetics *149*, 1899-1908.
29. Legal, L., Moulin, B., and Jallon, J.M. (1999). The Relation between Structures and Toxicity of Oxygenated Aliphatic Compounds Homologous to the Insecticide Octanoic Acid and the Chemotaxis of Two Species of *Drosophila*. Pesticide biochemistry and physiology *65*, 90-101.
30. Markow, T.A. (2019). Host use and host shifts in *Drosophila*. Curr Opin Insect Sci *31*, 139-145.
31. Legal, L., Chappe, B., and Jallon, J.M. (1994). Molecular basis of *Morinda citrifolia* (L.): Toxicity on *Drosophila*. J Chem Ecol *20*, 1931-1943.
32. Kaczmarek, A., Katarzyna Wronska, A., and Irena Bogus, M. (2024). Octanoic acid kills Lucilia sericata (Diptera: Calliphoridae) by affecting two major

defence systems: cuticular free fatty acids and immunocompetent cells. J Invertebr Pathol *206*, 108165.

33. Kaczmarek, A., Wronska, A.K., Kazek, M., and Bogus, M.I. (2022). Octanoic Acid-An Insecticidal Metabolite of *Conidiobolus coronatus* (Entomopthorales) That Affects Two Majors Antifungal Protection Systems in *Galleria mellonella* (Lepidoptera): Cuticular Lipids and Hemocytes. International journal of molecular sciences *23*.
34. Amlou, M., Pla, E., Moreteau, B., and David, J.R. (1997). Genetic analysis by interspecific crosses of the tolerance of *Drosophila sechellia* to major aliphatic acids of its host plant. Genet Sel Evol *29*, 511-522.
35. Amlou, M., Moreteau, B., and David, J.R. (1998). Larval tolerance in the *Drosophila melanogaster* species complex toward the two toxic acids of the *D. sechellia* host plant. Hereditas *129*, 7-14.
36. Huang, Y., and Erezyilmaz, D. (2015). The Genetics of Resistance to *Morinda* Fruit Toxin During the Postembryonic Stages in *Drosophila sechellia*. G3 (Bethesda) *5*, 1973-1981.
37. Sun, Z., Inagaki, S., Miyoshi, K., Saito, K., and Hayashi, S. (2024). Osiris gene family defines the cuticle nanopatterns of *Drosophila*. Genetics *227*.
38. Ando, T., Sekine, S., Inagaki, S., Misaki, K., Badel, L., Moriya, H., Sami, M.M., Itakura, Y., Chihara, T., Kazama, H., et al. (2019). Nanopore Formation in the Cuticle of an Insect Olfactory Sensillum. Curr Biol *29*, 1512-1520 e1516.
39. Andrade Lopez, J.M., Lanno, S.M., Auerbach, J.M., Moskowitz, E.C., Sligar, L.A., Wittkopp, P.J., and Coolon, J.D. (2017). Genetic basis of octanoic acid resistance in *Drosophila sechellia*: functional analysis of a fine-mapped region. Mol. Ecol. *26*, 1148-1160.
40. Lanno, S.M., Shimshak, S.J., Peyser, R.D., Linde, S.C., and Coolon, J.D. (2019). Investigating the role of *Osiris* genes in *Drosophila sechellia* larval resistance to a host plant toxin. Ecology and evolution *9*, 1922-1933.
41. Schlotterer, C., Kofler, R., Versace, E., Tobler, R., and Franssen, S.U. (2016). Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity (Edinb) *116*, 248.
42. Viswanatha, R., Li, Z., Hu, Y., and Perrimon, N. (2018). Pooled genome-wide CRISPR screening for basal and context-specific fitness gene essentiality in *Drosophila* cells. Elife *7*.
43. Colson, I. (2004). *Drosophila simulans*' response to laboratory selection for tolerance to a toxic food source used by its sister species *D. sechellia*. Evolutionary Ecology *18*, 15-28.
44. Barghi, N., Tobler, R., Nolte, V., Jaksic, A.M., Mallard, F., Otte, K.A., Dolezal, M., Taus, T., Kofler, R., and Schlotterer, C. (2019). Genetic redundancy fuels polygenic adaptation in *Drosophila*. PLOS Biol. *17*, e3000128.
45. Turner, T.L., and Miller, P.M. (2012). Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. Genetics *191*, 633-642.
46. Kelly, J.K., and Hughes, K.A. (2019). Pervasive Linked Selection and Intermediate-Frequency Alleles Are Implicated in an Evolve-and-Resequencing Experiment of *Drosophila simulans*. Genetics *211*, 943-961.
47. Shahrestani, P., King, E., Ramezan, R., Phillips, M., Riddle, M., Thornburg, M., Greenspan, Z., Estrella, Y., Garcia, K., Chowdhury, P., et al. (2021). The molecular architecture of *Drosophila melanogaster* defense against *Beauveria*

bassiana explored through evolve and resequence and quantitative trait locus mapping. G3 (Bethesda) *11*.

48. Zoh, M.G., Bonneville, J.M., Tutagata, J., Laporte, F., Fodjo, B.K., Mouhamadou, C.S., Sadia, C.G., McBeath, J., Schmitt, F., Horstmann, S., et al. (2021). Experimental evolution supports the potential of neonicotinoid-pyrethroid combination for managing insecticide resistance in malaria vectors. Sci Rep *11*, 19501.
49. Sadia, C.G., Bonneville, J.M., Zoh, M.G., Fodjo, B.K., Kouadio, F.A., Oyou, S.K., Koudou, B.G., Adepo-Gourene, B.A., Reynaud, S., David, J.P., et al. (2024). The impact of agrochemical pollutant mixtures on the selection of insecticide resistance in the malaria vector *Anopheles gambiae*: insights from experimental evolution and transcriptomics. Malar J *23*, 69.
50. Long, A., Liti, G., Luptak, A., and Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nat. Rev. Genet. *16*, 567-582.
51. Teotonio, H., Chelo, I.M., Bradic, M., Rose, M.R., and Long, A.D. (2009). Experimental evolution reveals natural selection on standing genetic variation. Nat Genet *41*, 251-257.
52. Keightley, P.D., Ness, R.W., Halligan, D.L., and Haddrill, P.R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. Genetics *196*, 313-320.
53. Signor, S.A., New, F.N., and Nuzhdin, S. (2018). A Large Panel of *Drosophila simulans* Reveals an Abundance of Common Variants. Genome Biol. Evol. *10*, 189-206.
54. R'Kha, S., Moreteau, B., Coyne, J.A., and David, J.R. (1997). Evolution of a lesser fitness trait: egg production in the specialist *Drosophila sechellia*. Genet Res *69*, 17-23.
55. Markow, T.A., Beall, S., and Matzkin, L.M. (2009). Egg size, embryonic development time and ovoviviparity in *Drosophila* species. J Evol Biol *22*, 430-434.
56. Abe, M., Kamiyama, T., Izumi, Y., Qian, Q., Yoshihashi, Y., Degawa, Y., Watanabe, K., Hattori, Y., Uemura, T., and Niwa, R. (2022). Shortened lifespan induced by a high-glucose diet is associated with intestinal immune dysfunction in *Drosophila sechellia*. J Exp Biol *225*.
57. Watada, M., Hayashi, Y., Watanabe, K., Mizutani, S., Mure, A., Hattori, Y., and Uemura, T. (2020). Divergence of *Drosophila* species: Longevity and reproduction under different nutrient balances. Genes Cells *25*, 626-636.
58. Vlachos, C., Burny, C., Pelizzola, M., Borges, R., Futschik, A., Kofler, R., and Schlotterer, C. (2019). Benchmarking software tools for detecting and quantifying selection in evolve and resequencing studies. Genome Biol *20*, 169.
59. Kojima, Y., Matsumoto, H., and Kiryu, H. (2020). Estimation of population genetic parameters using an EM algorithm and sequence data from experimental evolution populations. Bioinformatics *36*, 221-231.
60. Barata, C., Borges, R., and Kosiol, C. (2023). Bait-ER: A Bayesian method to detect targets of selection in Evolve-and-Resequence experiments. J Evol Biol *36*, 29-44.
61. Cochran, W.G. (1954). Some Methods for Strengthening the Common Chi-Squared Tests. Biometrics *10*, 417-451.

1237 62. Mantel, N., and Haenszel, W. (1959). Statistical aspects of the analysis of
1238 data from retrospective studies of disease. J Natl Cancer Inst *22*, 719-748.
1239 63. Orozco-terWengel, P., Kapun, M., Nolte, V., Kofler, R., Flatt, T., and
1240 Schlotterer, C. (2012). Adaptation of *Drosophila* to a novel laboratory
1241 environment reveals temporally heterogeneous trajectories of selected alleles.
1242 Mol. Ecol. *21*, 4931-4941.
1243 64. Hungate, E.A., Earley, E.J., Boussy, I.A., Turissini, D.A., Ting, C.T., Moran,
1244 J.R., Wu, M.L., Wu, C.I., and Jones, C.D. (2013). A locus in *Drosophila*
1245 *sechellia* affecting tolerance of a host plant toxin. Genetics *195*, 1063-1075.
1246 65. Hu, Y., Comjean, A., Perrimon, N., and Mohr, S.E. (2017). The *Drosophila*
1247 Gene Expression Tool (DGET) for expression analyses. BMC Bioinf. *18*, 98.
1248 66. Xu, Y., Viswanatha, R., Sitsel, O., Roderer, D., Zhao, H., Ashwood, C.,
1249 Voelcker, C., Tian, S., Raunser, S., Perrimon, N., et al. (2022). CRISPR
1250 screens in *Drosophila* cells identify Vsg as a Tc toxin receptor. Nature *610*,
1251 349-355.
1252 67. Borrull, A., Lopez-Martinez, G., Poblet, M., Cordero-Otero, R., and Rozes, N.
1253 (2015). New insights into the toxicity mechanism of octanoic and decanoic
1254 acids on *Saccharomyces cerevisiae*. Yeast *32*, 451-460.
1255 68. Emelyanov, A.V., Rabbani, J., Mehta, M., Vershilova, E., Keogh, M.C., and
1256 Fyodorov, D.V. (2014). *Drosophila* TAP/p32 is a core histone chaperone that
1257 cooperates with NAP-1, NLP, and nucleophosmin in sperm chromatin
1258 remodeling during fertilization. Genes Dev *28*, 2027-2040.
1259 69. Galindo, K.A., Lu, W.J., Park, J.H., and Abrams, J.M. (2009). The Bax/Bak
1260 ortholog in *Drosophila*, Debcl, exerts limited control over programmed cell
1261 death. Development *136*, 275-283.
1262 70. Edwin Chan, H.Y., Harris, S.J., and O'Kane, C.J. (1998). Identification and
1263 characterization of kraken, a gene encoding a putative hydrolytic enzyme in
1264 *Drosophila melanogaster*. Gene *222*, 195-201.
1265 71. Zhang, L.S., Xiong, Q.P., Pena Perez, S., Liu, C., Wei, J., Le, C., Zhang, L.,
1266 Harada, B.T., Dai, Q., Feng, X., et al. (2021). ALKBH7-mediated
1267 demethylation regulates mitochondrial polycistronic RNA processing. Nat Cell
1268 Biol *23*, 684-691.
1269 72. Solberg, A., Robertson, A.B., Aronsen, J.M., Rognmo, O., Sjaastad, I.,
1270 Wisloff, U., and Klungland, A. (2013). Deletion of mouse Alkbh7 leads to
1271 obesity. J Mol Cell Biol *5*, 194-203.
1272 73. Kumar, K., Mhetre, A., Ratnaparkhi, G.S., and Kamat, S.S. (2021). A
1273 Superfamily-wide Activity Atlas of Serine Hydrolases in *Drosophila*
1274 *melanogaster*. Biochemistry *60*, 1312-1324.
1275 74. Li, H., Janssens, J., De Waegeneer, M., Kolluru, S.S., Davie, K., Gardeux, V.,
1276 Saelens, W., David, F.P.A., Brbic, M., Spanier, K., et al. (2022). Fly Cell Atlas:
1277 A single-nucleus transcriptomic atlas of the adult fruit fly. Science *375*,
1278 eabk2432.
1279 75. Lemaitre, B., and Miguel-Aliaga, I. (2013). The digestive tract of *Drosophila*
1280 *melanogaster*. Annu Rev Genet *47*, 377-404.
1281 76. Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gislason, M.H., Pihl,
1282 S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H.
1283 (2022). SignalP 6.0 predicts all five types of signal peptides using protein
1284 language models. Nat. Biotechnol. *40*, 1023-1025.

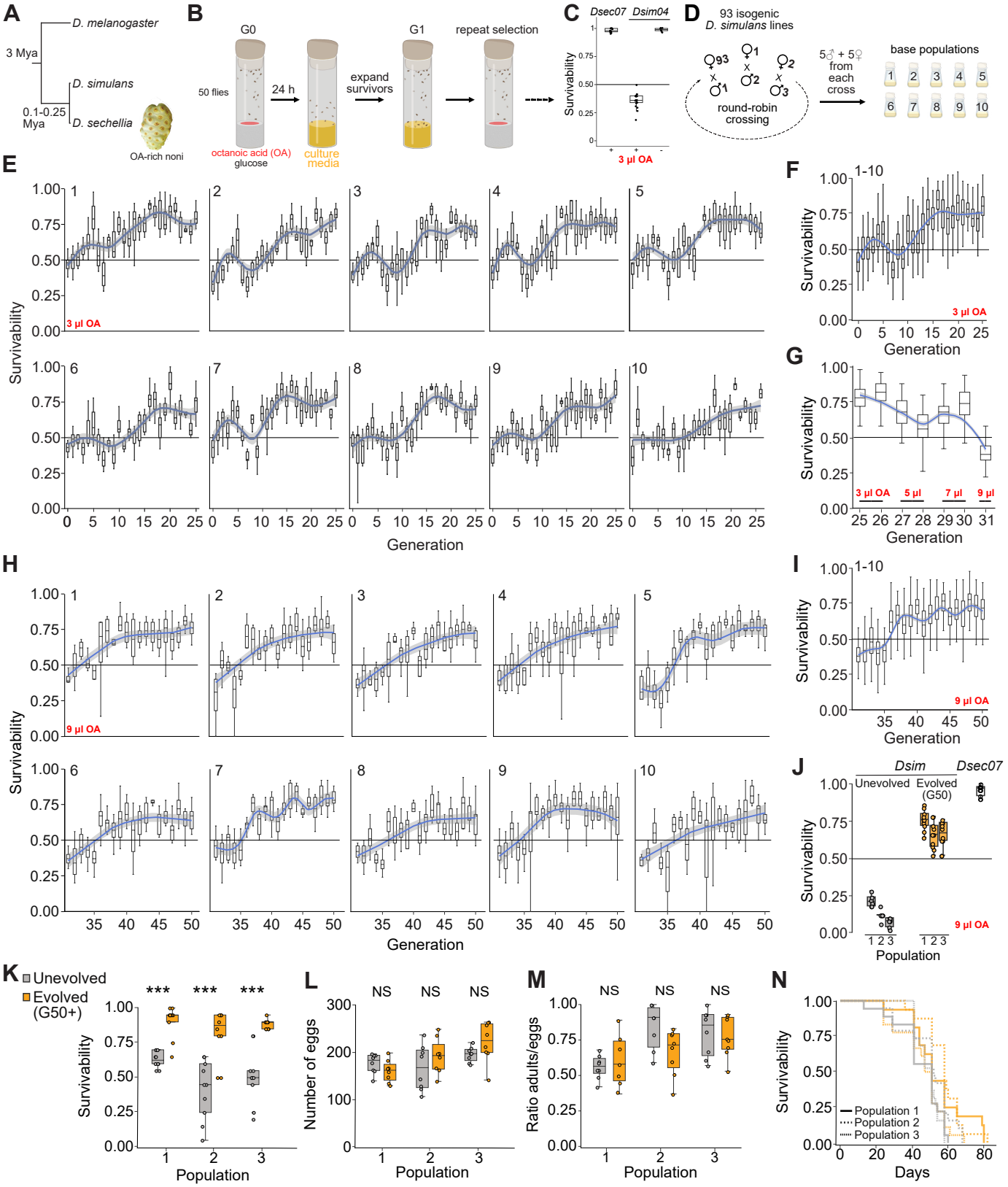
77. Yang, J., McCart, C., Woods, D.J., Terhzaz, S., Greenwood, K.G., ffrench-Constant, R.H., and Dow, J.A. (2007). A *Drosophila* systems approach to xenobiotic metabolism. Physiol Genomics *30*, 223-231.
78. Xu, B., Liu, D., Wang, Z., Tian, R., and Zuo, Y. (2021). Multi-substrate selectivity based on key loops and non-homologous domains: new insight into ALKBH family. Cell Mol Life Sci *78*, 129-141.
79. Fukasawa, Y., Tsuji, J., Fu, S.C., Tomii, K., Horton, P., and Imai, K. (2015). MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. Molecular & cellular proteomics : MCP *14*, 1113-1126.
80. Chen, C.L., Hu, Y., Udeshi, N.D., Lau, T.Y., Wirtz-Peitz, F., He, L., Ting, A.Y., Carr, S.A., and Perrimon, N. (2015). Proteomic mapping in live *Drosophila* tissues using an engineered ascorbate peroxidase. Proceedings of the National Academy of Sciences of the United States of America *112*, 12093-12098.
81. Sen, A., and Cox, R.T. (2022). Loss of *Drosophila* Clueless differentially affects the mitochondrial proteome compared to loss of Sod2 and Pink1. Frontiers in physiology *13*, 1004099.
82. 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.
83. McDonald, J.H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. Nature *351*, 652-654.
84. Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L., and Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol. Biol. Evol. *30*, 1196-1205.
85. Watanabe, K., Kanaoka, Y., Mizutani, S., Uchiyama, H., Yajima, S., Watada, M., Uemura, T., and Hattori, Y. (2019). Interspecies Comparative Analyses Reveal Distinct Carbohydrate-Responsive Systems among *Drosophila* Species. Cell Rep *28*, 2594-2607 e2597.
86. Ma, S., Avanesov, A.S., Porter, E., Lee, B.C., Mariotti, M., Zemskaya, N., Guigo, R., Moskalev, A.A., and Gladyshev, V.N. (2018). Comparative transcriptomics across 14 *Drosophila* species reveals signatures of longevity. Aging Cell *17*, e12740.
87. Kalra, S., Lanno, S., Sanchez, G., and Coolon, J.D. (2024). cis- and trans-regulatory contributions to a hierarchy of factors influencing gene expression variation. Communications biology *7*, 1563.
88. Mota, M.N., Matos, M., Bahri, N., and Sa-Correia, I. (2024). Shared and more specific genetic determinants and pathways underlying yeast tolerance to acetic, butyric, and octanoic acids. Microb Cell Fact *23*, 71.
89. Liu, P., Chernyshov, A., Najdi, T., Fu, Y., Dickerson, J., Sandmeyer, S., and Jarboe, L. (2013). Membrane stress caused by octanoic acid in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol *97*, 3239-3251.
90. National Center for Biotechnology Information PubChem Compound Summary for CID 379, Octanoic Acid. Retrieved January 19, 2024.
91. Marconcini, M., Fragnière, C., Masuzzo, A., and Benton, R. (2025). Genome-wide association studies identify new candidate genes and tissues underlying resistance to a natural toxin in drosophilids. bioRxiv.
92. Dow, J.A.T., Simons, M., and Romero, M.F. (2022). *Drosophila melanogaster*: a simple genetic model of kidney structure, function and disease. Nat Rev Nephrol *18*, 417-434.

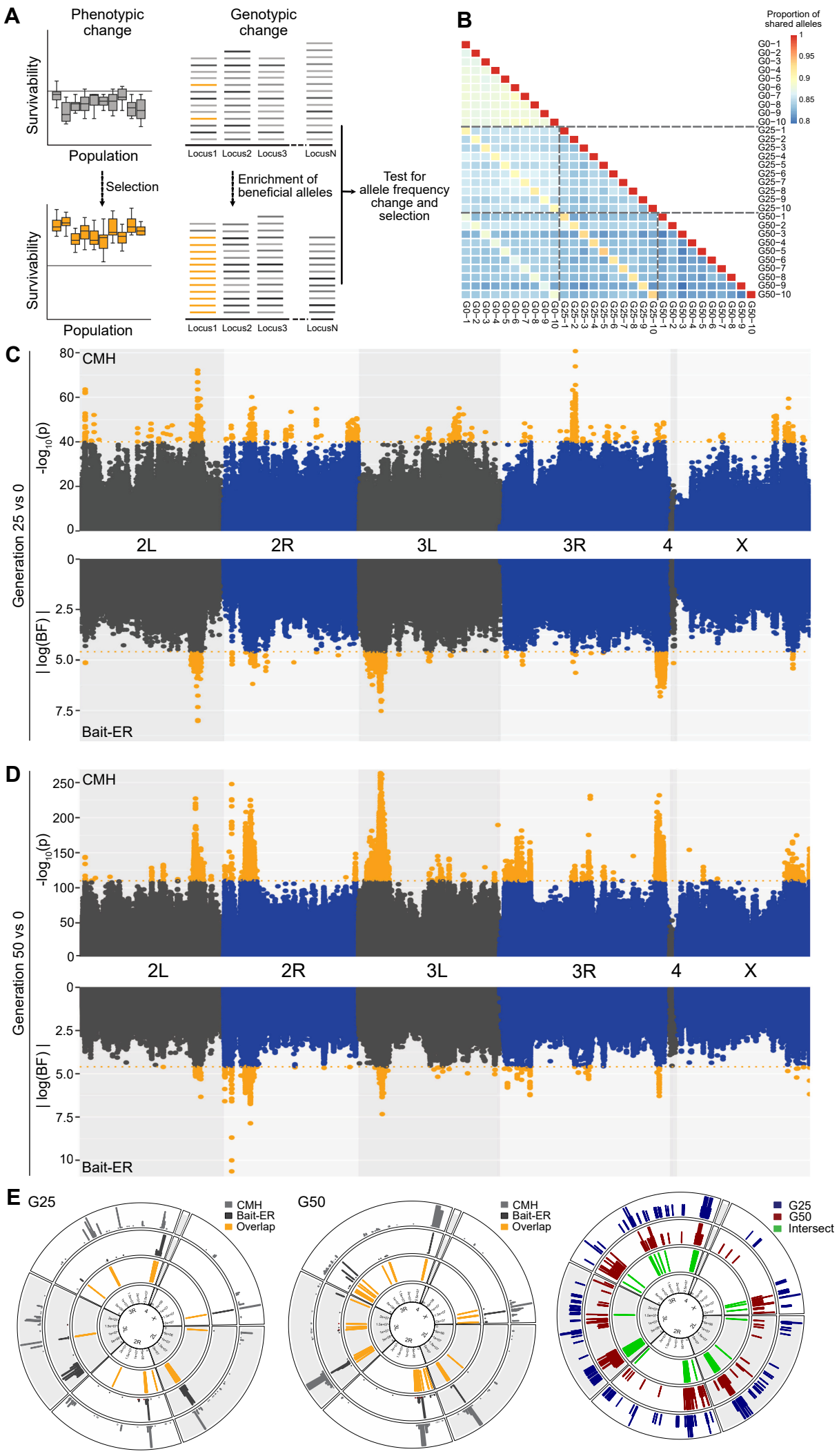
1335 93. Gutierrez-Guerrero, Y.T., Ibarra-Laclette, E., Martinez Del Rio, C., Barrera-
1336 Redondo, J., Rebollar, E.A., Ortega, J., Leon-Paniagua, L., Urrutia, A.,
1337 Aguirre-Planter, E., and Eguiarte, L.E. (2020). Genomic consequences of
1338 dietary diversification and parallel evolution due to nectarivory in leaf-nosed
1339 bats. Gigascience *9*.
1340 94. Xia, B., Viswanatha, R., Hu, Y., Mohr, S.E., and Perrimon, N. (2023). Pooled
1341 genome-wide CRISPR activation screening for rapamycin resistance genes in
1342 *Drosophila* cells. Elife *12*.
1343 95. Viswanatha, R., Mameli, E., Rodiger, J., Merckaert, P., Feitosa-Suntheimer,
1344 F., Colpitts, T.M., Mohr, S.E., Hu, Y., and Perrimon, N. (2021). Bioinformatic
1345 and cell-based tools for pooled CRISPR knockout screening in mosquitos. Nat
1346 Commun *12*, 6825.
1347 96. Port, F., Strein, C., Stricker, M., Rauscher, B., Heigwer, F., Zhou, J.,
1348 Beyersdorffer, C., Frei, J., Hess, A., Kern, K., et al. (2020). A large-scale
1349 resource for tissue-specific CRISPR mutagenesis in *Drosophila*. Elife *9*.
1350 97. Lenče, T. (2021). PhD Thesis: The role of m6A modification on mRNA
1351 processing in *Drosophila melanogaster*. Volume PhD. (Johannes Gutenberg
1352 University Mainz).
1353 98. Kondo, S., and Ueda, R. (2013). Highly Improved Gene Targeting by
1354 Germline-Specific Cas9 Expression in *Drosophila*. Genetics.
1355 99. Croset, V., Rytz, R., Cummins, S.F., Budd, A., Brawand, D., Kaessmann, H.,
1356 Gibson, T.J., and Benton, R. (2010). Ancient protostome origin of
1357 chemosensory ionotropic glutamate receptors and the evolution of insect taste
1358 and olfaction. PLOS Genet. *6*, e1001064.
1359 100. Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An
1360 optimized transgenesis system for *Drosophila* using germ-line-specific phiC31
1361 integrases. Proceedings of the National Academy of Sciences of the United
1362 States of America *104*, 3312-3317.
1363 101. Therneau, T.M. (2024). coxme: Mixed Effects Cox Models. R package version
1364 *2.2-22*.
1365 102. Pino, J.A., Marquez, E., and Castro, D. (2009). Volatile and non-volatile acids
1366 of noni (*Morinda citrifolia* L.) fruit. J Sci Food Agric *89*, 1247-1249.
1367 103. Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal
1368 likelihood estimation of semiparametric generalized linear models. Journal of
1369 the Royal Statistical Society: Series B (Statistical Methodology) *73*, 3-36.
1370 104. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible
1371 trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.
1372 105. Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly
1373 Contigs with BWA-MEM. arXiv.
1374 106. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O.,
1375 Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve
1376 years of SAMtools and BCFtools. Gigascience *10*.
1377 107. Kofler, R., Pandey, R.V., and Schlotterer, C. (2011). PoPoolation2: identifying
1378 differentiation between populations using sequencing of pooled DNA samples
1379 (Pool-Seq). Bioinformatics *27*, 3435-3436.
1380 108. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A.,
1381 Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The
1382 variant call format and VCFtools. Bioinformatics *27*, 2156-2158.
1383 109. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D.,
1384 Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set

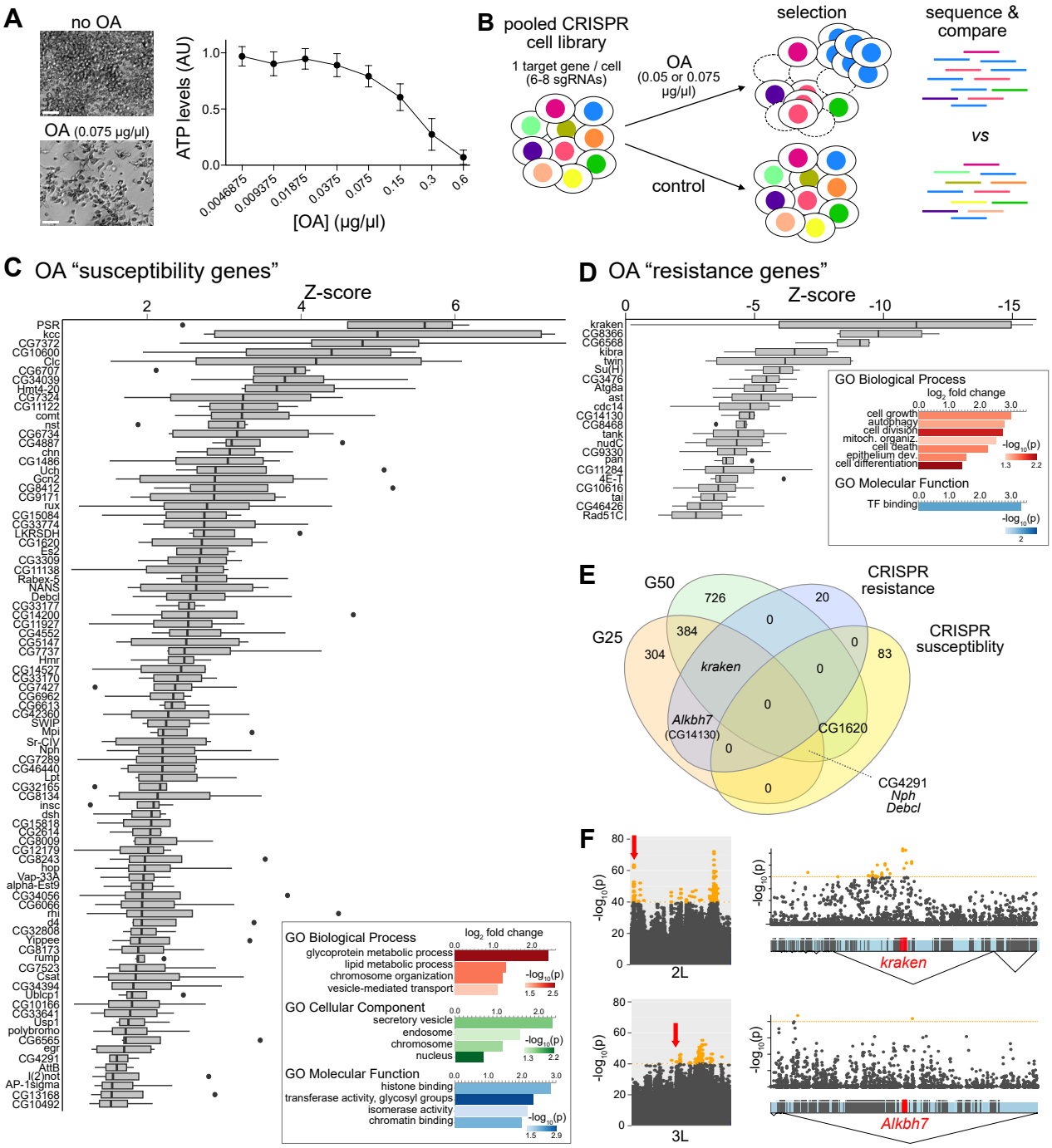
1385 for whole-genome association and population-based linkage analyses. Am J
1386 Hum Genet *81*, 559-575.
1387 110. Purcell, S. (2020). PLINK v1.90b6.21.
1388 111. Vlachos, C., and Kofler, R. (2018). MimicrEE2: Genome-wide forward
1389 simulations of Evolve and Resequencing studies. PLoS Comput Biol *14*,
1390 e1006413.
1391 112. Taus, T., Futschik, A., and Schlotterer, C. (2017). Quantifying Selection with
1392 Pool-Seq Time Series Data. Mol. Biol. Evol. *34*, 3023-3034.
1393 113. Zhang, C., Dong, S.S., Xu, J.Y., He, W.M., and Yang, T.L. (2019).
1394 PopLDdecay: a fast and effective tool for linkage disequilibrium decay
1395 analysis based on variant call format files. Bioinformatics *35*, 1786-1788.
1396 114. Dong, S.S., He, W.M., Ji, J.J., Zhang, C., Guo, Y., and Yang, T.L. (2021).
1397 LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium
1398 and haplotype blocks based on variant call format files. Briefings in
1399 bioinformatics *22*.
1400 115. Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis
1401 and visualization of LD and haplotype maps. Bioinformatics *21*, 263-265.
1402 116. Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize
1403 Implements and enhances circular visualization in R. Bioinformatics *30*, 2811-
1404 2812.
1405 117. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for
1406 comparing genomic features. Bioinformatics *26*, 841-842.
1407 118. Altenhoff, A.M., Train, C.M., Gilbert, K.J., Mediratta, I., Mendes de Farias, T.,
1408 Moi, D., Nevers, Y., Radoykova, H.S., Rossier, V., Warwick Vesztrocy, A., et
1409 al. (2021). OMA orthology in 2021: website overhaul, conserved isoforms,
1410 ancestral gene order and more. Nucleic Acids Res. *49*, D373-D379.
1411 119. Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C.,
1412 Imamichi, T., and Chang, W. (2022). DAVID: a web server for functional
1413 enrichment analysis and functional annotation of gene lists (2021 update).
1414 Nucleic Acids Res. *50*, W216-W221.
1415 120. Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and
1416 integrative analysis of large gene lists using DAVID bioinformatics resources.
1417 Nat Protoc *4*, 44-57.
1418 121. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer,
1419 K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC
1420 Bioinf. *10*, 421.
1421 122. Sayers, E.W., Beck, J., Bolton, E.E., Brister, J.R., Chan, J., Connor, R.,
1422 Feldgarden, M., Fine, A.M., Funk, K., Hoffman, J., et al. (2025). Database
1423 resources of the National Center for Biotechnology Information in 2025.
1424 Nucleic Acids Res. *53*, D20-D29.
1425 123. Ozturk-Colak, A., Marygold, S.J., Antonazzo, G., Attrill, H., Goutte-Gattat, D.,
1426 Jenkins, V.K., Matthews, B.B., Millburn, G., Dos Santos, G., Tabone, C.J., et
1427 al. (2024). FlyBase: updates to the *Drosophila* genes and genomes database.
1428 Genetics *227*.
1429 124. Viswanatha, R., Entwisle, S., Hu, Y., Kim, A.R., Reap, K., Butnaru, M., Qadiri,
1430 M., Mohr, S.E., and Perrimon, N. (2025). Higher resolution pooled genome-
1431 wide CRISPR knockout screening in *Drosophila* cells using integration and
1432 anti-CRISPR (IntAC). Nat Commun *16*, 6498.

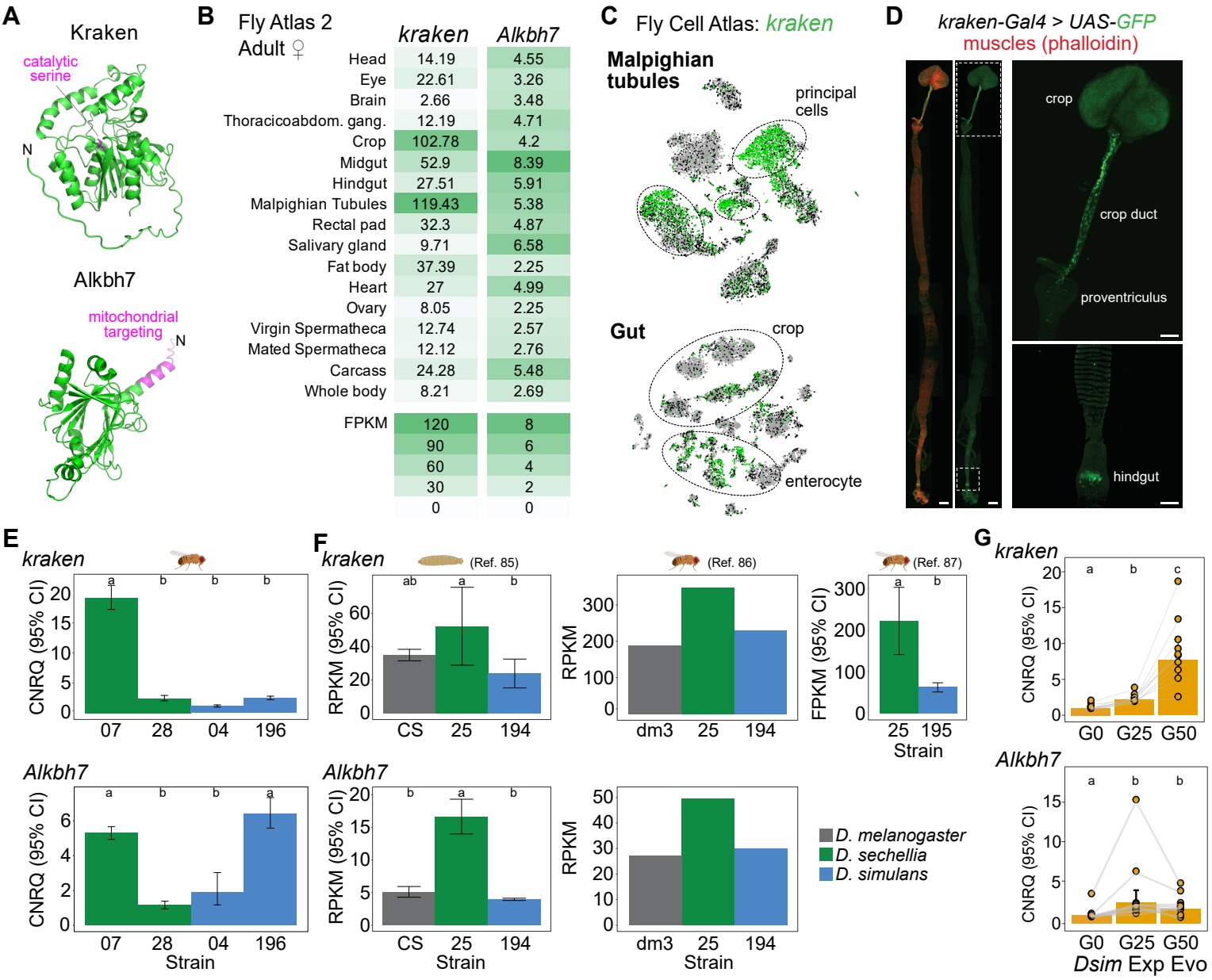
125. Viswanatha, R., Brathwaite, R., Hu, Y., Li, Z., Rodiger, J., Merckaert, P., Chung, V., Mohr, S.E., and Perrimon, N. (2019). Pooled CRISPR Screens in *Drosophila* Cells. Current protocols in molecular biology *129*, e111.
126. Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol *15*, 554.
127. Hu, Y., Comjean, A., Attrill, H., Antonazzo, G., Thurmond, J., Chen, W., Li, F., Chao, T., Mohr, S.E., Brown, N.H., et al. (2023). PANGEA: a new gene set enrichment tool for *Drosophila* and common research organisms. Nucleic Acids Res. *51*, W419-W426.
128. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol *8*, R19.
129. Lu, J., Yang, C., Zhang, Y., and Pan, H. (2018). Selection of Reference Genes for the Normalization of RT-qPCR Data in Gene Expression Studies in Insects: A Systematic Review. Frontiers in physiology *9*, 1560.
130. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol *3*, RESEARCH0034.
131. Rogers, R.L., Cridland, J.M., Shao, L., Hu, T.T., Andolfatto, P., and Thornton, K.R. (2014). Landscape of standing variation for tandem duplications in *Drosophila yakuba* and *Drosophila simulans*. Mol. Biol. Evol. *31*, 1750-1766.
132. Leinonen, R., Akhtar, R., Birney, E., Bonfield, J., Bower, L., Corbett, M., Cheng, Y., Demiralp, F., Faruque, N., Goodgame, N., et al. (2010). Improvements to services at the European Nucleotide Archive. Nucleic Acids Res. *38*, D39-45.
133. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLOS ONE *9*, e112963.
134. Abascal, F., Zardoya, R., and Telford, M.J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. *38*, W7-13.
135. Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A.R.N., Lopez, R., and Butcher, S. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. Nucleic Acids Res. *52*, W521-W525.
136. Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., and Sanchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol. Biol. Evol. *34*, 3299-3302.
137. Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A., et al. (2020). HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol. Biol. Evol. *37*, 295-299.
138. Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320-324.

139. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. *50*, D439-D444.
140. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.
141. Krause, S.A., Overend, G., Dow, J.A.T., and Leader, D.P. (2022). FlyAtlas 2 in 2022: enhancements to the *Drosophila melanogaster* expression atlas. Nucleic Acids Res. *50*, D1010-D1015.
142. Hadjieconomou, D., Rotkopf, S., Alexandre, C., Bell, D.M., Dickson, B.J., and Salecker, I. (2011). Flybow: genetic multicolor cell labeling for neural circuit analysis in *Drosophila melanogaster*. Nat Methods *8*, 260-266.
143. Bontonou, G., Saint-Leandre, B., Kafle, T., Baticle, T., Hassan, A., Sanchez-Alcaniz, J.A., and Arguello, J.R. (2024). Evolution of chemosensory tissues and cells across ecologically diverse Drosophilids. Nat Commun *15*, 1047.

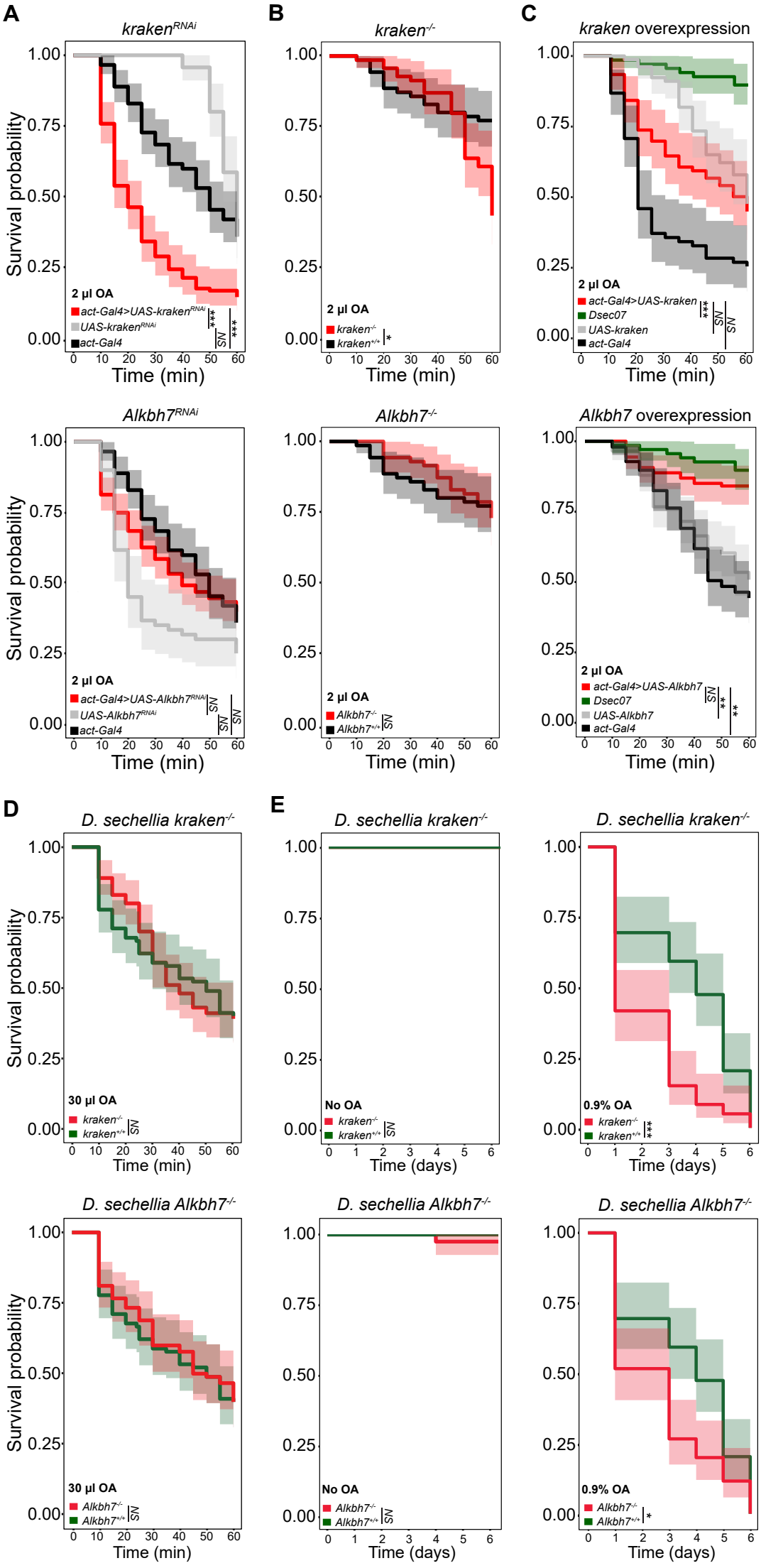

A
D. melanogaster
3 Mya
D. simulans
0.1-0.25 Mya
D. sechellia
OA-rich noni
B
G0
50 flies
24 h
octanoic acid (OA)
glucose
culture media
expand survivors
G1
repeat selection
C
Dsec07
Dsim04
Survivability
3 µl OA
D
93 isogenic D. simulans lines
round-robin crossing
5♂ + 5♀ from each cross
base populations
E
Survivability
3 µl OA
Generation
F
1-10
Survivability
3 µl OA
Generation
G
Survivability
3 µl OA
5 µl
7 µl
9 µl
Generation
H
Survivability
9 µl OA
Generation
I
1-10
Survivability
9 µl OA
Generation
J
Dsim
Dsec07
Unevolved
Evolved (G50)
Survivability
9 µl OA
Population
K
Unevolved
Evolved (G50+)
***
Survivability
Population
L
NS
Number of eggs
Population
M
NS
Ratio adults/eggs
Population
N
Survivability
Population 1
Population 2
Population 3
Days


A
Phenotypic change
Genotypic change
Survivability
Population
Selection
Enrichment of beneficial alleles
Locus1 Locus2 Locus3 LocusN
Test for allele frequency change and selection
B
Proportion of shared alleles
C
CMH
Generation 25 vs 0
-log10(p)
| log(BF) |
Bait-ER
2L 2R 3L 3R 4 X
D
CMH
Generation 50 vs 0
Bait-ER
E
G25
G50
CMH
Bait-ER
Overlap
G25
G50
Intersect

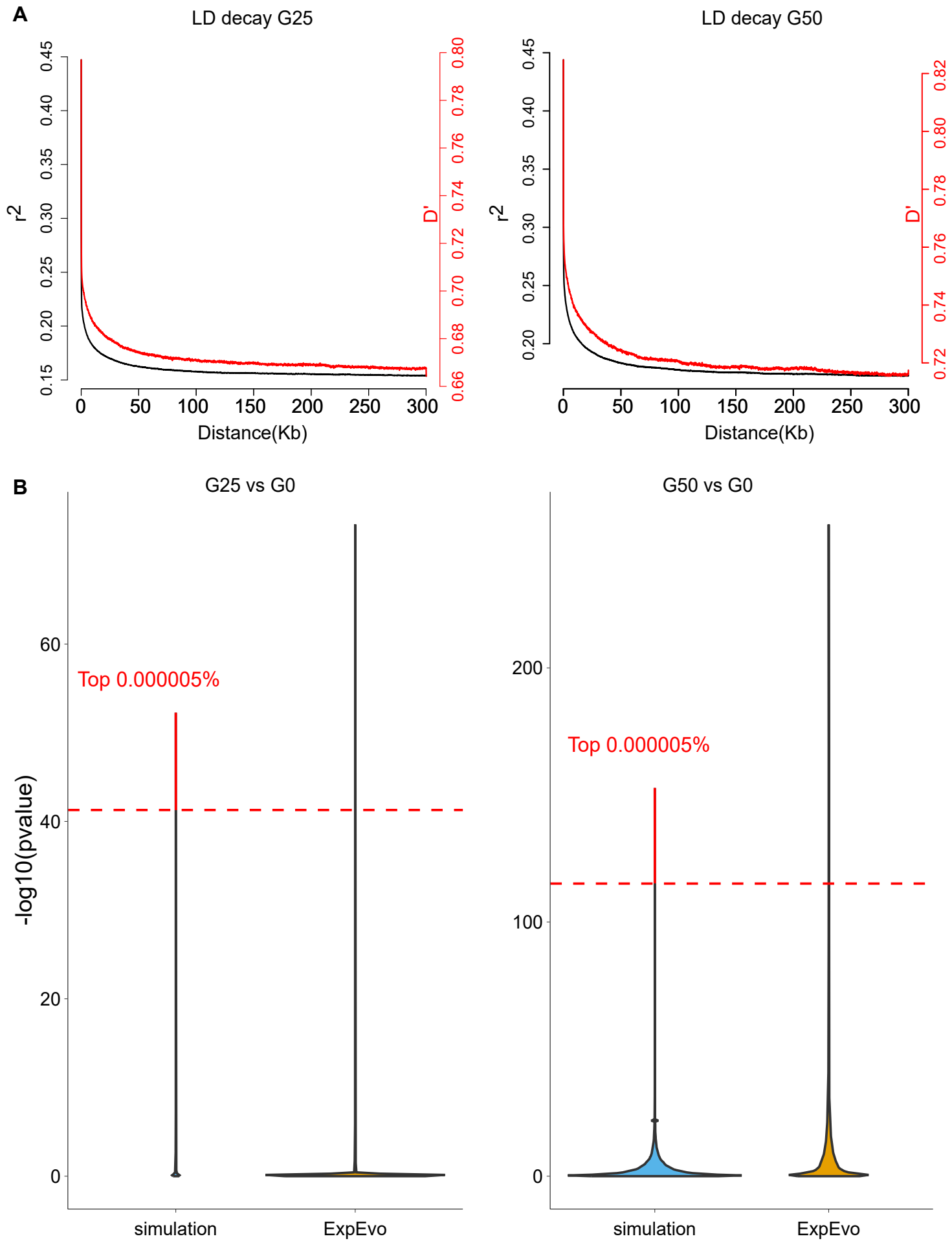

A
no OA
OA (0.075 µg/µl)
ATP levels (AU)
1.0
0.5
0.0
0.0046875
0.009375
0.01875
0.0375
0.075
0.15
0.3
0.6
[OA] (µg/µl)
B
pooled CRISPR cell library
1 target gene / cell (6-8 sgRNAs)
OA (0.05 or 0.075 µg/µl)
control
selection
sequence & compare
vs
C
OA "susceptibility genes"
Z-score
2
4
6
PSR
kcc
CG7372
CG10600
Clc
CG6707
CG34039
Hmt4-20
CG7324
CG11122
comt
nst
CG6734
CG4887
chn
CG1486
Uch
Gcn2
CG8412
CG9171
nux
CG15084
CG33774
LKRSDH
CG1620
Es2
CG3309
CG11138
Rabex-5
NANS
Debcl
CG33177
CG14200
CG11927
CG4552
CG5147
CG7737
Hmr
CG14527
CG33170
CG7427
CG6962
CG6613
CG42360
SWIP
Mpi
Sr-CIV
Nph
CG7289
CG46440
Lpt
CG32165
CG8134
insc
dsh
CG15818
CG2614
CG8009
CG12179
CG8243
hop
Vap-33A
alpha-Est9
CG34056
CG6066
rhi
d4
CG32808
Yippee
CG8173
rump
CG7523
Csat
CG34394
Ubiquitin-conjugating enzyme 1
CG10166
CG33641
Usp1
polybromo
CG6565
egr
CG4291
AttB
l(2)not
AP-1sigma
CG13168
CG10492
GO Biological Process
log2 fold change
0.0
1.0
2.0
glycoprotein metabolic process
lipid metabolic process
chromosome organization
vesicle-mediated transport
-log10(p)
1.5
2.5
GO Cellular Component
secretory vesicle
endosome
chromosome
nucleus
1.3
2.2
GO Molecular Function
3.0
histone binding
transferase activity, glycosyl groups
isomerase activity
chromatin binding
2.9
D
OA "resistance genes"
Z-score
0
-5
-10
-15
kraken
CG8366
CG6568
kibra
twin
Su(tr)
CG3476
Atg8a
ast
cdc14
CG14130
CG8468
tank
nudC
CG9330
pan
CG11284
4E-T
CG10616
tai
CG46426
Rad51C
GO Biological Process
log2 fold change
cell growth
autophagy
cell division
mitoch. organiz.
cell death
epithelium dev.
cell differentiation
-log10(p)
1.3
2.2
GO Molecular Function
TF binding
2
E
G50
G25
CRISPR resistance
CRISPR susceptiblity
726
20
384
304
83
0
kraken
Alkbh7 (CG14130)
CG1620
CG4291
Nph
Debcl
F
-log10(p)
80
60
40
20
0
2L
3L
kraken
Alkbh7

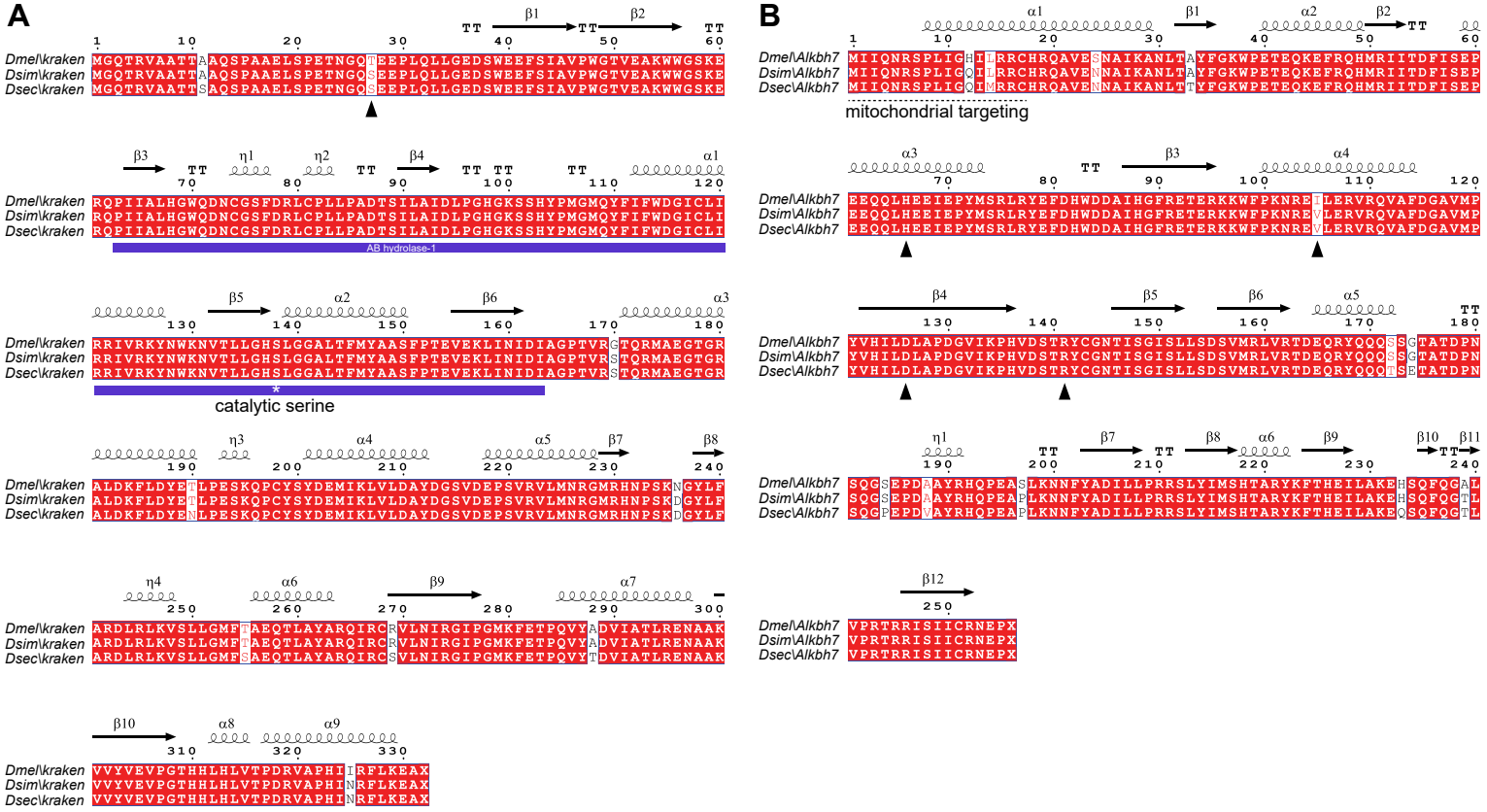

A
Kraken
catalytic serine
N
Alkbh7
mitochondrial targeting
N
B
Fly Atlas 2
Adult ♀
kraken
Alkbh7
Head 14.19 4.55
Eye 22.61 3.26
Brain 2.66 3.48
Thoracicoabdom. gang. 12.19 4.71
Crop 102.78 4.2
Midgut 52.9 8.39
Hindgut 27.51 5.91
Malpighian Tubules 119.43 5.38
Rectal pad 32.3 4.87
Salivary gland 9.71 6.58
Fat body 37.39 2.25
Heart 27 4.99
Ovary 8.05 2.25
Virgin Spermatheca 12.74 2.57
Mated Spermatheca 12.12 2.76
Carcass 24.28 5.48
Whole body 8.21 2.69
FPKM 120 8
90 6
60 4
30 2
0 0
C
Fly Cell Atlas: kraken
Malpighian tubules
principal cells
Gut
crop
enterocyte
D
kraken-Gal4 > UAS-GFP
muscles (phalloidin)
crop
crop duct
proventriculus
hindgut
E
kraken
CNRQ (95% CI)
07 28 04 196
Alkbh7
CNRQ (95% CI)
07 28 04 196
Strain
F
kraken
(Ref. 85)
RPKM (95% CI)
CS 25 194
(Ref. 86)
RPKM
dm3 25 194
(Ref. 87)
FPKM (95% CI)
25 195
Strain
Alkbh7
RPKM (95% CI)
CS 25 194
Strain
RPKM
dm3 25 194
Strain
D. melanogaster
D. sechellia
D. simulans
G
kraken
CNRQ (95% CI)
G0 G25 G50
Alkbh7
CNRQ (95% CI)
G0 G25 G50
Dsim Exp Evo

Marconcini, Figure 5



Marconcini, Figure S1





C

McDonald & Kreitman Test

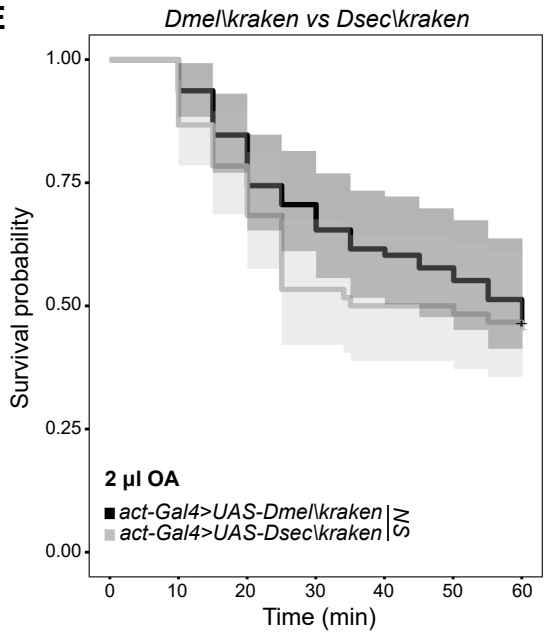
	NI	alpha	p-value
kraken	0.083	0.917	0.018*
Alkbh7	0.544	0.456	0.477

D

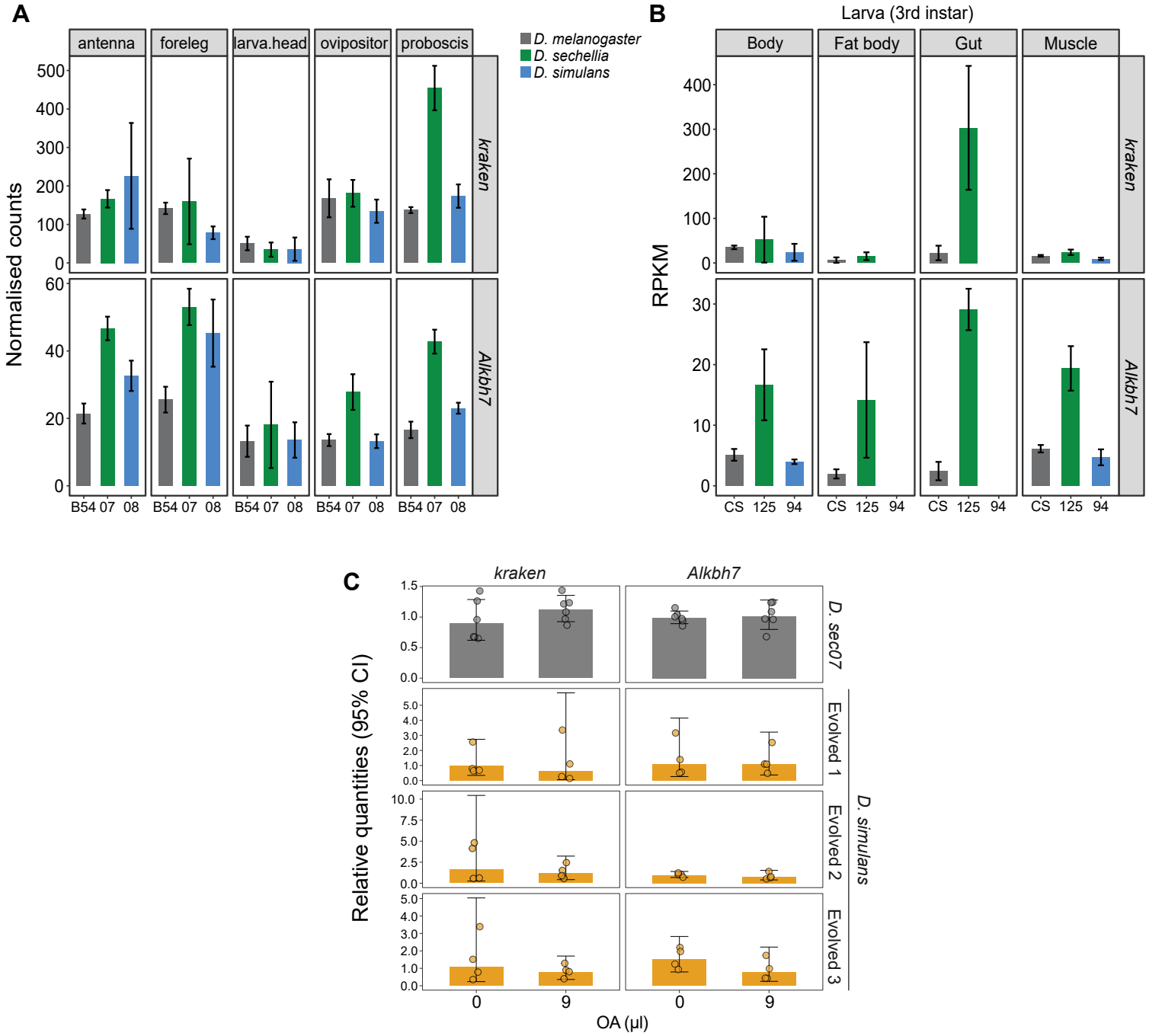
HyPhy FUBAR

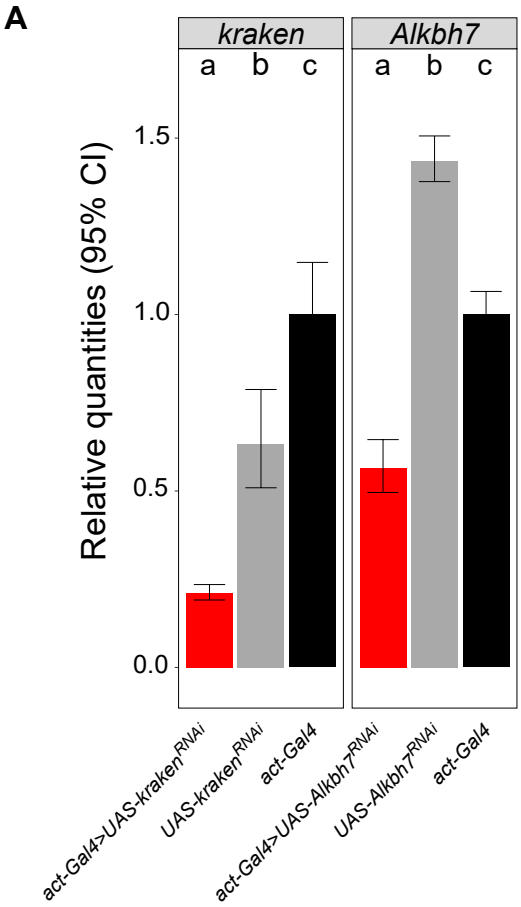
	position	alpha	beta	post. prob.	codons	Freq *Dsec*	Freq *Dsim*
kraken	27	2.503	40.230	0.979	S / A / T	1 / 0 / 0	0.658 / 0.237 / 0.105
Alkbh7	66	0.768	9.440	0.977	H / Y	1 / 0	0.806 / 0.194
	105	0.665	7.158	0.957	V / I / L	1 / 0 / 0	0.839 / 0.097 / 0.064
	126	0.783	13.505	0.989	D / E	1 / 0	0.774 / 0.226
	141	0.723	9.841	0.982	R / K	1 / 0	0.871 / 0.129

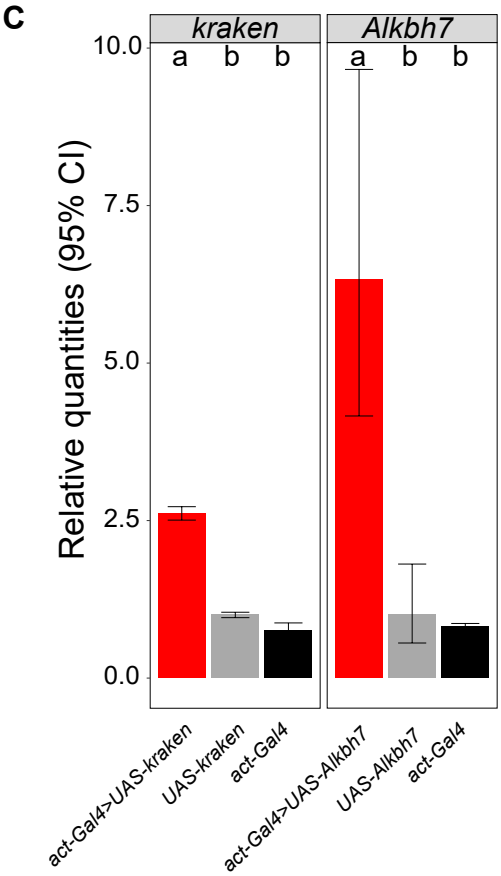
E

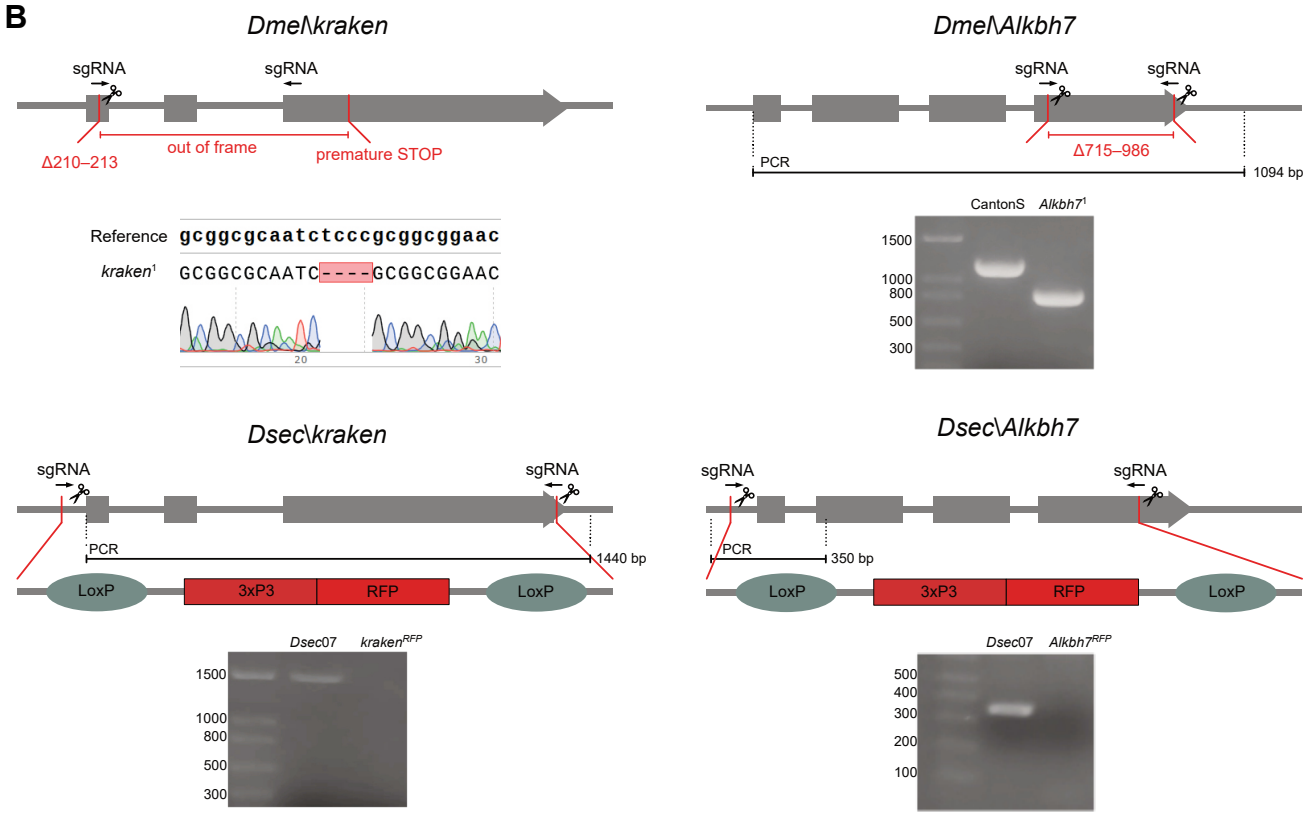


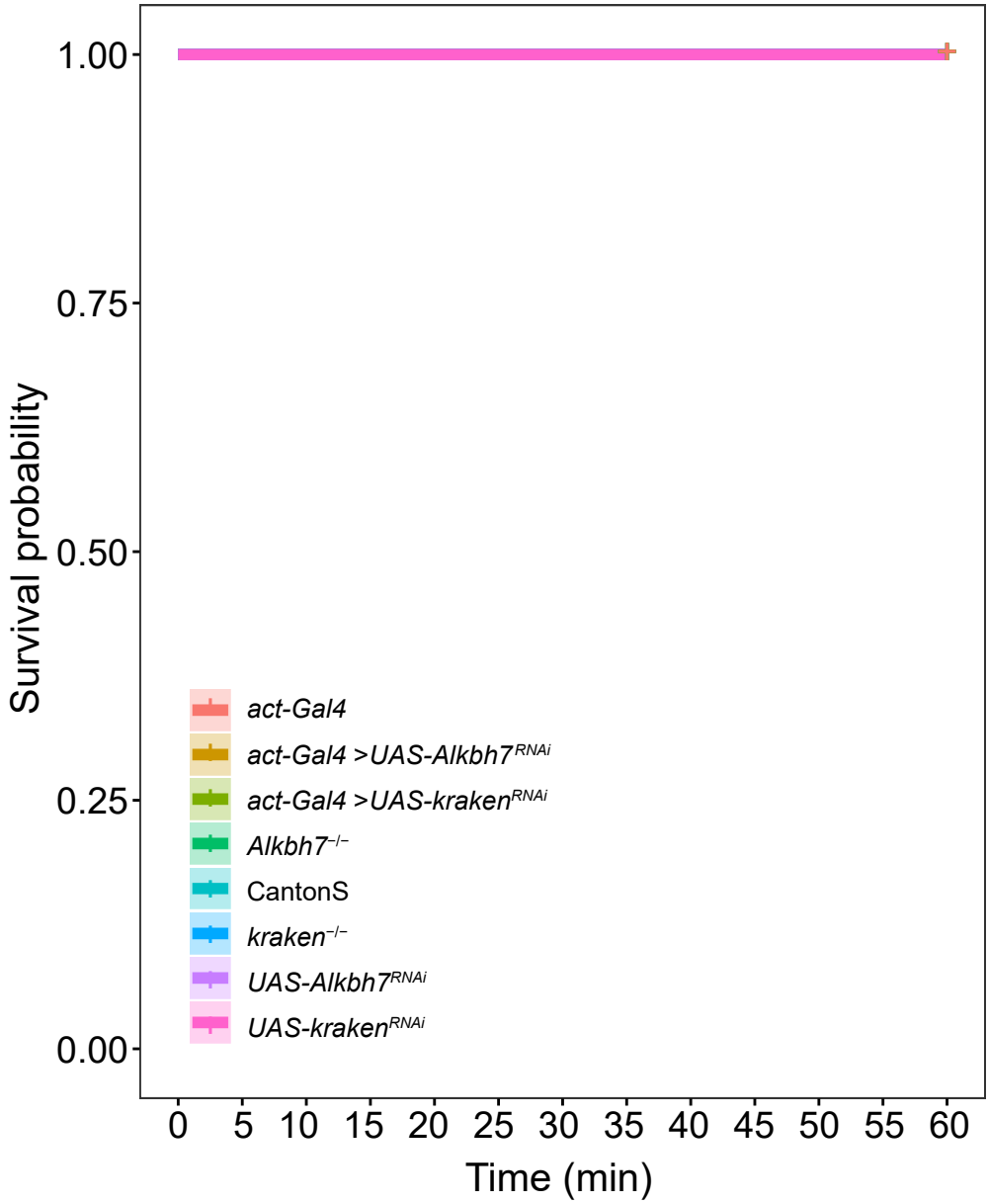
Marconcini, Figure S3




A
kraken
Alkbh7
a b c
a b c
Relative quantities (95% CI)
1.5
1.0
0.5
0.0
act-Gal4>UAS-kraken^RNAi
UAS-kraken^RNAi
act-Gal4
act-Gal4>UAS-Alkbh7^RNAi
UAS-Alkbh7^RNAi
act-Gal4


C
kraken
Alkbh7
a b b
a b b
Relative quantities (95% CI)
10.0
7.5
5.0
2.5
0.0
act-Gal4>UAS-kraken
UAS-kraken
act-Gal4
act-Gal4>UAS-Alkbh7
UAS-Alkbh7
act-Gal4


B
Dme\kraken
sgRNA
sgRNA
Δ210–213
out of frame
premature STOP
Reference gcggcgcaatctcccgcggcggaac
kraken¹ GCGGCGCAATC----GCGGCGGAAC
Dme\Alkbh7
sgRNA
sgRNA
PCR
Δ715–986
1094 bp
CantonS
Alkbh7¹
1500
1000
800
500
300
Dsec\kraken
sgRNA
sgRNA
PCR
1440 bp
LoxP
3xP3
RFP
LoxP
Dsec07
kraken^RFP
1500
1000
800
500
300
Dsec\Alkbh7
sgRNA
sgRNA
PCR
350 bp
LoxP
3xP3
RFP
LoxP
Dsec07
Alkbh7^RFP
500
400
300
200
100


1.00
0.75
0.50
0.25
0.00
Survival probability
0
5
10
15
20
25
30
35
40
45
50
55
60
Time (min)
act-Gal4
act-Gal4 >UAS-Alkbh7RNAi
act-Gal4 >UAS-krakenRNAi
Alkbh7−/−
CantonS
kraken−/−
UAS-Alkbh7RNAi
UAS-krakenRNAi