# 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 and for pest management, but remains challenging due to complex genetics. To investigate how *Drosophila sechellia* evolved resistance to toxic octanoic acid (OA) of noni fruit, its exclusive habitat, we experimentally-evolved the closely-related *Drosophila simulans* with increased OA resistance, identifying multiple loci under selection. Cross-referencing these with a genome-wide, OA-resistance CRISPR screen in *Drosophila melanogaster* cells highlighted two proteins: Kraken, a putative detoxification enzyme 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, mutations of these genes in *D. sechellia* reduced OA tolerance. Our identification of genes underlying OA resistance in lab and natural contexts demonstrates how complementary, cross-species selection approaches can elucidate toxin adaptation mechanisms.

## 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, 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 (*12-14*). For example, thickening of the cuticle to reduce entry of insecticides into the body is widely considered an effective resistance mechanism (*15, 16*). 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 (*16*). 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 (*17*). Duplicated metabolic genes can also undergo neofunctionalisation 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 (*18*), 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 (*19*). 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 to 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 (*20, 21*). Amongst *D. sechellia*'s numerous adaptations (*21-26*), this species has evolved high resistance to octanoic acid (OA), an abundant noni chemical that is toxic to other drosophilids and more divergent insects (*21, 27-32*). Previous 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 (*27, 33-35*). The main effect locus encompasses a cluster of 18 genes including members of the *Osiris* (*Osi*) family, which are thought to function in cuticle patterning (*36, 37*). RNA interference (RNAi) knockdown of some *Osi* genes in *D. melanogaster* led to reduced tolerance of OA (*38, 39*); however, recent analyses indicate a general decrease in viability of *Osi* null mutant *D. melanogaster*, even in the absence of OA (*36*). Moreover, 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 (*40*), and CRISPR/Cas9-based genome-wide genetic screening in cultured *D. melanogaster* cells (*41*). We show how the intersection of these approaches can identify novel genes underlying OA resistance in *D. sechellia*, establishing these methods and this drosophilid clade as a powerful paradigm to investigate the evolution and mechanisms of insecticide resistance.

## Results

### Experimental evolution of OA resistance in *D. simulans*

The Evolve-and-Resequence approach 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 (*40, 42-45*). 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 (*46, 47*).

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 (*33, 48*), 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 reasoned that we could allow these to mate and establish the next generation (Fig. 1B), thus enriching genetic variants conferring higher resistance. Repeating the process over multiple generations and multiple replicates would allow us to track phenotypic changes in resistance and investigate the associated trajectories in allele frequencies.

Standing genetic variation should be present in the evolving populations for natural selection to act upon (*49, 50*). As the spontaneous mutation rate in drosophilids

is too low for artificial selection (*51*), we generated ten independent polymorphic base populations of *D. simulans* from 93 isogenic strains (*52*) by round-robin crossing (Fig. 1D). To minimize bottleneck effects, each population was kept at a size of ~1000 flies. At each generation, ~500 flies from each population were randomly collected and selected for resistance to 3 µl OA. Over the course of 25 generations, all the populations showed increased levels of survivability, suggesting parallel adaptive evolution (Fig. 1E). Given the consistent responses across all populations, we also analyzed the phenotypic changes grouping all replicates, finding that the mean survivability significantly increased from ~43% at generation 0 to ~78% at generation 25 (Fig. 1F).

As we kept the same amount of OA throughout the 25 generations of selection, the selective pressure lowered as the populations' survivability increased, evidenced as a larger fraction of flies surviving and reproducing. Viewing the phenotype of populations both individually (Fig. 1E) and collectively (Fig. 1F), it was apparent that the change in phenotype plateaued at around generation 15. For continued evolution, we therefore gradually increased the amount of OA during generations 26-31 (Fig. 1G), reaching a selection pressure comparable to the initial phase of the experiment with 9 µl OA (~38% mean survivability across all populations) (Fig. 1G). We then performed a second round of selection at this dose, observing an increase in survivability across all populations, reaching ~70% by generation 50 (Fig. 1H). As the populations adapted to the new condition, a second plateau started around generation 40 (Fig. 1H-I). Over 50 generations we were therefore able to evolve *D. simulans* to have approximately 5-fold higher survivability to 9 µl OA compared to the unevolved populations (Fig. 1J), though still falling short of the survival exhibited by *D. sechellia* (Fig. 1J). OA is only one component of noni fruit, although probably the major toxic chemical (*30*). We therefore tested three of our evolved populations for resistance to noni fruit pulp, finding that they exhibited much higher survival on this substrate compared to the same original populations maintained under neutral conditions (i.e., no OA selection) (Fig. 1K).

Specialization of *D. sechellia* on toxic noni is associated with a reduced egg-laying capacity (*53, 54*) and a shorter (albeit strain-specific) lifespan (*25, 55, 56*). Our success in evolving *D. simulans* with a partial *D. sechellia*-like resistance trait begs the question as to whether these selected strains exhibit other phenotypic changes as a trade-off for higher toxin resistance. We tested whether the evolved resistant 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 (*57-59*). We therefore assessed the significance of frequency changes using two complementary approaches: the Cochran–Mantel–Haenszel (CMH) test (*60, 61*), a widely-used method that tests for allele frequency changes between two time points, and Bait-ER (*59*), 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 (*40, 62*).

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

Although the experimental evolution experiment pointed to a number of genomic regions relevant for OA resistance, they contained too many genes to enable facile choice of candidates for functional follow-up. We therefore sought an orthogonal screening approach. Motivated by previous observations of OA toxicity for insect Sf9 cultured cells (*32*), we found that OA also leads to reduced viability of *D. melanogaster* S2R+ cells (Fig. 3A). This phenotype allowed us to perform a selection experiment in this cell line through exploitation of a genome-wide, pooled CRISPR/Cas9 mutagenesis approach (Fig. 3B) (*41*). 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 essential genes (*41*) and genes that were not detectably expressed in S2R+ cells (using DGET (*63*)). 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 (*64*)). 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 (*32, 65*), 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 (*66*)) and *Debcl* (a pro-apoptotic member of the Bcl-2 family (*67*)); 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 (*68*) 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 (*69, 70*), 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 (*68*). 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 (*71*). *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 (*68*).

Using the Fly Cell Atlas (*72*), 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 (*73*) (Fig. 4C). However, Kraken lacks an N-terminal signal sequence (*74*), suggesting that the protein remains in the cytoplasm of these cells, unlike canonical digestive enzymes (*73*), and more similar to detoxification enzymes of the Cytochrome P450 and Glutathione S-transferase families (*75*). 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 (*76*). As in mammals (*70*), *D. melanogaster Alkbh7* displays broad tissue expression (Fig. 4B). Murine Alkbh7 is localized to the mitochondrial matrix (*70*). This property appears to be conserved in the *Drosophila* protein, which has a mitochondrial targeting sequence (Fig. 4A and Fig. S2B) (*77*); consistently, we identified Alkbh7 within *D. melanogaster* proteomes of the mitochondrial matrix (*78, 79*).

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* (*80*). Using the McDonald-Kreitman test (*81*) 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) (*82*), 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* (*83-85*) (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.*

350 *melanogaster* (Fig. S4A) and assessed the effects on resistance to OA using a short-
351 term, plate-based assay (*27*) (see Methods). *kraken*$^{RNAi}$ led to significantly diminished
352 OA resistance compared to genetic controls, while *Alkbh7*$^{RNAi}$ did not have a
353 discernable effect (Fig. 5A). We next generated deletion mutants for these genes in
354 *D. melanogaster* (Fig. S4B). Similar to the RNAi experiments, *kraken* mutants, but not
355 *Alkbh7* mutants, displayed reduced resistance to OA (Fig. 5B). All genotypes remained
356 fully viable when no OA was added (Fig. S5).
357 Loss-of-function analyses might fail to uncover a role for *Alkbh7* in whole
358 animals if there are redundant resistance mechanisms absent from S2R+ cells. We
359 therefore took a complementary approach by overexpressing these genes in *D.*
360 *melanogaster*. Strikingly, overexpression of *Alkbh7* resulted in survivability of *D.*
361 *melanogaster* to levels close to those observed for *D. sechellia* under the same
362 experimental conditions (Fig. 5C, Fig. S4C), indicating that this gene is sufficient to
363 confer high-level OA resistance. By contrast, *kraken* overexpression did not have any
364 detectable impact on OA resistance (Fig. 5C); a similar lack of effect was seen when
365 overexpressing *D. sechellia kraken* (Fig. S2E). The absence of a phenotypic effect
366 upon *kraken* overexpression is surprising, given its elevated expression in *D. sechellia*
367 and in evolved *D. simulans* but this could reflect either insufficient overexpression
368 levels (Extended Data Fig. 4c) and/or genotype-by-genotype interactions that are not
369 recapitulated when manipulating a single locus in a *D. melanogaster* background.
370 A key question is whether these genes contribute to the natural OA resistance
371 of *D. sechellia*. Implementing CRISPR/Cas9-based genome-engineering in this
372 species (*21, 26*), we generated and validated null mutants in these genes in *D.*
373 *sechellia* (Fig. S4B) and tested their resistance to OA in the plate assay. Although we
374 did not detect any significant difference to genetically-matched wild-type flies (Fig. 5D),
375 this was anticipated given the polygenic nature of OA resistance in *D. sechellia* (*27,*
376 *33-35*), and the modest or absent phenotypes of equivalent mutants in *D.*
377 *melanogaster* (Fig. 5B). We reasoned that more subtle differences might be detected
378 by monitoring *D. sechellia*'s survivability over longer periods of time, when repeated
379 contact with OA was enforced by blending it with the food. Indeed, we observed higher
380 mortality rates of both *kraken* and *Alkbh7* mutants compared to genetically-matched
381 wild-type flies over a period of 6 days (Fig. 5E). By contrast, both mutants survived
382 equally well as controls in the absence of OA over this time period (Fig. 5E). These
383 data provide the first evidence for the contribution of specific genes to the natural OA
384 resistance of *D. sechellia*.
385
386 **Discussion**
387
388 Identifying specific loci underlying phenotypic evolution is challenging due to the
389 polygenicity of the vast majority of adaptive traits. Investigating the evolution of
390 chemical toxicity is an attractive model trait because of the relatively easy phenotypic
391 read-out. To gain insight into the genetic basis of resistance to a natural insecticide,
392 OA, we have intersected the results of selection over vastly different timescales:
393 weeks in cell lines, years in experimentally-evolved laboratory strains, and millennia
394 during speciation.
395 Unlike insecticides that activate or antagonize broadly-expressed neuronal ion
396 channels to cause organismal death, the mode of action of OA appears to be
397 mechanistically very different, capable of inducing death at the level both of whole
398 organisms and of cultured cells. Beyond insect cell lines (our work and (*32*)), OA is
399 also toxic to budding yeast (*86*) and bacteria (*87, 88*), suggesting that this toxin acts

via a more fundamental cellular mechanism. Indeed, a genome-wide knock-out screen in yeast for genes required for tolerance of OA highlighted contributions of several core cellular processes, such as vesicle trafficking, mitochondrial function or chromatin regulation (*86*), reminiscent of our CRISPR screen results. Consistent with this idea, the results of our experimental evolution and CRISPR screening, as well as earlier mapping efforts in *D. sechellia*/*D. simulans* hybrids (*27, 33-35*) revealed multiple genomic regions/genes contributing to the resistance to OA.

Considered separately, the experimental evolution and cell-based CRISPR screens identified many candidates. That there was very limited overlap between the screen results is unsurprising given their very different designs. While both lists are likely to contain multiple genes relevant to understanding OA resistance (and hence mode(s) of action of OA), here we prioritized two common hits, which point to new defense mechanisms underlying OA resistance.

Although both Kraken and Alkbh7 are enzymes, the (partial) necessity, but insufficiency, of *kraken*, and lack of necessity but sufficiency of *Alkbh7* to confer OA resistance implies fundamentally different roles of these proteins. The expression of *kraken* in the gut and renal system supports a hypothesis of this protein as a detoxification enzyme, although future work will be necessary to determine whether Kraken has a direct role in OA degradation. The mild loss-of-function phenotype of *kraken* in both *D. melanogaster* and *D. sechellia* likely 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 (*73, 89*). 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 (*75*). 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 likely informative: mice lacking *Alkbh7* are obese, with higher body fat levels, proposed to result from a defect in metabolism of fatty acids (*70*). Molecularly, one study suggests that Alkbh7's metabolic role stems from its regulation of RNA editing to control mitochondrial protein translation (*69*), 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, which can be explored 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 (*90*).

Taken together, our work has demonstrated the power of orthogonal selection approaches in the laboratory to understand the evolution of toxin resistance in a historic natural paradigm of this phenomenon (*22*). Moreover, the methodology and concepts are now applicable to explore the molecular mechanisms of susceptibility and resistance of insects to the wide-range of artificial toxic chemicals used in the environment, encompassing those used to specifically target pest insects, as well as those that indirectly impair the health of insect species (*91*). Such knowledge will have application in the future design and deployment 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 (*52*) (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 (*92*) 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 (*93*). Mutants were validated by genomic PCR and sequencing (Fig. S4B).

*D. sechellia* mutant flies were generated by WellGenetics Inc. using modified methods of (*94*). 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)* (*26*). 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* (*95*). 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 (*96*). Transgenic flies were obtained by phiC31 integrase-mediated transgenesis into attP2 performed by BestGene Inc. and verified by sequencing.

540 **Toxin resistance assays**
541
542 *Tube assay*: resistance to OA for the experimental evolution setup was scored
543 adapting a previous protocol (*33, 48*). In brief, flies were lightly anesthetized with $CO_2$
544 and ~50 animals were placed in standard cultured tubes (25 mm diameter × 95 mm
545 height, Milian SA) and provided with 1 ml of a 3% glucose solution on a tissue paper
546 (Kimtech, 7552). 3 μl of pure OA (Sigma-Aldrich, CAS 124-07-2, C2875) were directly
547 pipetted onto the tissue while the flies remained anesthetized. Direct contact with the
548 acid was scrupulously avoided until all flies recovered. The fraction of alive flies was
549 scored after 24 h. For the experimental evolution experiment, the surviving flies were
550 collected (see below).
551
552 *Plate assay*: to measure OA resistance in the genetically-manipulated *D.*
553 *melanogaster* and *D. sechellia* strains, we used a plate assay (*27*). Compared to the
554 experiments in tubes, this assay offers greater temporal resolution and higher dynamic
555 range (in contrast to the experimental evolution experiment, we did not strive to
556 maintain a certain fraction of surviving flies), as well as greater statistical power from
557 tests of fewer animals. Ten female flies were lightly anesthetized with $CO_2$ and placed
558 in a Petri dish (60 mm diameter × 15 mm height) in which 2 μl of pure OA (or 30 μl for
559 *D. sechellia*) were spotted in the center of the underside of the lid. The flies were
560 allowed to recover (typically <5 min), and immobile/dead were scored every 5 min for
561 1 h. The survival curves were calculated and plotted using the R package survminer
562 (10.32614/CRAN.package.survminer). The statistical analysis was conducted fitting a
563 mixed effect Cox regression model as implemented in the R package coxme (*97*).
564 Genotype was included as a fixed effect, and replicate dishes were modelled as a
565 random effect.
566
567 *Noni resistance assay*: noni fruit was collected from *M. citrifolia* plants grown in the
568 University of Lausanne greenhouses. Five ripe (pale yellow) fruits were homogenized
569 together, and the pulp puree frozen at -20°C. The noni assay was performed with the
570 same setup as for tube assay described above, using 2 g of thawed noni puree placed
571 on the bottom of the tube.
572
573 *Long-term assay*: To assess the long-term effects of OA (i.e., beyond 24 h), two-day-
574 old mated female flies were individually housed in vials containing food supplemented
575 with OA. OA was thoroughly mixed into Formula 4-24® Instant Drosophila Medium,
576 Blue (Carolina Biological Supply) to a final concentration of 0.9%. Fresh OA-containing
577 food was provided every two days, and mortality was monitored daily until all flies in
578 the experimental group died. Statistical analysis was run as for the plate assay.
579
580 *Fecundity* and *developmental viability*: 10 mated adult female flies (4-7 days old) were
581 allowed to lay eggs for 24 h on agar plates (2% agar, 2.5% sucrose in a 1:3 grape
582 juice:water mix). The total number of eggs on each plate was counted. From each
583 plate, 100 eggs were collected and transferred to standard fly food. Emerging adults
584 were counted and removed daily until no further flies emerged.
585
586 *Lifespan*: Two-day-old mated female flies were individually housed in vials containing
587 standard fly food and monitored daily until death. Fresh food was provided every 2–3
588 days.
589

590 **Experimental evolution**
591
592 *Selection procedure*: to generate the base populations for experimental evolution, we
593 used 93 isogenic lines (Table S2) of the *D. simulans* panel established by (*52*). We
594 first crossed each of the lines together following a round-robin design to ensure equal
595 contribution of each line to the starting genetic pool; 5 males and 5 females from each
596 cross were then combined in ten independent replicates (Fig. S1D). These ten
597 populations were maintained in bottles (57 mm diameter ×103 mm height, Milian SA)
598 of standard fly food at high population size (~1000 individuals per population) to
599 minimize the effect of genetic drift. Each population was subdivided into smaller
600 groups of ~250 individuals per bottle (i.e., four bottles per population) to reduce
601 competition. At each generation, individuals from the four bottles were mixed to
602 prevent bottle-specific genetic drift. The experimental evolution experiment was
603 initiated 10 generations after the creation of these populations.
604 From each population, 500 individuals (2-7 day old) were randomly collected
605 and further split into 10 groups of 50 mixed-sex flies (i.e., ~5000 flies total). Each group
606 of 50 flies was then placed under selection with 3 μl of OA for 24 h (as described
607 above); the survivors were collected and allowed to breed in food vials for 6-8 days,
608 transferring them to fresh vials every 2-3 days. In the next generation, 500 flies per
609 population were collected only from the last seeded vial, to maximize the probability
610 of these flies being parented by the OA-resistant males and females and not offspring
611 generated by females that mated before the selection procedure. These flies were
612 then subjected to the same selection procedure. This process was repeated for 25
613 generations, and survivability was scored at each generation.
614 A second round of experimental evolution was performed by increasing the
615 amount of OA to 9 μl between generations 31 and 50; this amount was determined
616 empirically during generation 26-31 to re-establish the survivability close to the initial
617 level (Fig. 1G). All successive generations were maintained following the same
618 procedure.
619 As a control, the ten original populations were maintained in parallel in the same
620 conditions as previously described (i.e. ~1000 individuals per population subdivided
621 into smaller groups of ~250 individuals per bottle).
622
623 *DNA sequencing and analysis:* we pooled ~200 individual flies per population per time
624 point (generations 0, 25 and 50). Genomic DNA (and total RNA for qPCR, as described
625 below) extraction was performed using the Quick DNA/RNA Miniprep plus kit (Zymo
626 Research, D7003) following the manufacturer's instructions. Samples were stored at
627 -80°C.
628 We performed paired-end whole genome sequencing on an Illumina NovaSeq
629 6000 following Nextera DNA flex preparation, at an average coverage of >200×. Read
630 quality was assessed with Fastqc (*98*) and adapters were trimmed with Trimmomatic
631 (*99*) providing the Nextera adapters sequences. We aligned the trimmed reads to the
632 *D. simulans* reference genome (NCBI GCF_016746395.2) using bwa-mem (*100*) and
633 the results were sorted and compressed using samtools (*101*). Duplicate reads were
634 then removed with Picard MarkDuplicates (https://broadinstitute.github.io/picard/).
635 Unplaced scaffolds and alignments with mapping quality <20 were removed with
636 samtools. We then combined all the aligned sequences into a samtools pileup file and
637 split this into individual chromosomes. Regions with zero coverage were removed with
638 custom scripts. Sync files for each chromosome were generated with the

639 mpileup2sync.pl script implemented in Popoolation2 (*102*) and used as input files for
640 all the Evolve-and-Resequence analyses.
641 To explore the genetic variation of our samples we generated vcf files with
642 vcftools 0.1.17 (*103*) and binary files with plink1.9 (*104, 105*). We calculated the
643 identity-by-state (IBS) matrix as implemented in plink1.9 (function -distance 'ibs').
644
645 *Evolve-and-Resequence analysis*: allele frequency changes between generation 0
646 and generations 25 or 50 were investigated via Popoolation2 and Bait-ER (*59*).
647 We performed a Cochran–Mantel–Haenszel (CMH) test via the cmh-test.pl
648 script implemented in Popoolation2 comparing base and evolved populations with
649 parameters --min-count 12 and --min-coverage 50. We estimated a plausible
650 significance threshold approximating the top 0.000005% of the simulated p-values
651 based on the effect of genetic drift alone by running genome-wide forward simulations
652 under neutrality in mimicree2 (*106*). We used the same genomic information from the
653 original 93 lines as input and the average real survivability ratios observed throughout
654 the selection experiment as –selection-regime parameter and 500 as –population-size
655 parameter (Fig. S1B).
656 For Bait-ER, the input sync files were filtered to match --min-count 12 and --
657 min-coverage 50 and scaled to a uniform coverage of 100 as in (*57*) to avoid numerical
658 problems due to high coverage. We ran the algorithm on each chromosome
659 separately. Bait-ER models the effect of genetic drift directly using estimates of the
660 effective population size ($N_e$). We calculated $N_e$ estimates for each chromosome in
661 each replicate population with the R package poolSeq (*107*) and used the median
662 value of each chromosome as "Population_size" parameter (Table S3). Other
663 parameters included "Sampling correction" = 0 (i.e., binomial) and "Prior parameters"
664 =0.001 (shape), 0.001 (rate). Data points were considered significant when the
665 absolute natural logarithm of the Bayes Factor exceeded the conservative threshold
666 of log(0.99/0.01) as in (*59*).
667 The linkage disequilibrium decay in the 10 evolved populations was analysed
668 across the whole genome using PopLDdecay (*108*). Linkage disequilibrium maps of
669 focal regions were generated with LDBlockShow (*109*) using the Solid Spine method
670 based on D' statistics originally introduced in Haploview (*110*).
671 For the genome-wide screen, all the genes within selection blocks were
672 considered potential candidates. Based on the results on linkage disequilibrium decay,
673 a block was arbitrarily defined as a genomic interval where at least two significant
674 SNPs were found at a distance less than 50 kb. Consecutive SNPs respecting this
675 definition were grouped into larger blocks and visualized as circus plots using the R
676 package circlize (*111*). The genes overlapping such blocks were extracted using
677 BEDtools intersect (*112*). Orthologous genes in *D. melanogaster* were retrieved using
678 a combination of the OMA browser (*113*), DAVID (*114, 115*), BLAST (*116*), NCBI (*117*)
679 and FlyBase (*118*), and subsequent manual curation.
680
681 **Genome-wide CRISPR/Cas9 screen in *D. melanogaster* S2R+ cells**
682
683 *Library design, construction, and transfection*: design, synthesis and cloning of the
684 genome-wide sgRNA library are previously described (*119*). To determine sub-lethal
685 doses of OA appropriate for screening, the effect of a gradient of OA supplementation
686 into media on attP+ Cas9+ cells was determined using CellTiter-Glo (Promega)
687 following the manufacturer's protocol; luminescence measurements were made using
688 a Spectramax Paradigm (Molecular Devices).

For the screen, attP+ Cas9+ *D. melanogaster* S2R+ cells (RRID:CVCL_UD30) were transfected using Effectene and a 1:1 molar ratio of sgRNA library and pIntAC, as described (*119*). Briefly, transfected cells were distributed in 5 ml aliquots across ten 100 mm culture dishes and incubated overnight, then 5 ml of fresh media was added to each plate. After 4 days, cells were subjected to puromycin selection for 12 days, subculturing every 4 days, to establish a stable sgRNA-expressing cell library. For the selection assay, $2.5 \times 10^7$ cells were plated onto four 15 cm diameter cell culture dishes to ensure sufficient sgRNA coverage, with each sgRNA represented approximately 1000 times (*119*).

*OA selection assay*: library-containing cells were left untreated or treated with either 0.075 µg/µl or 0.05 µg/µl OA (Sigma-Aldrich, CAS 124-07-2, C2875). Each replicate consisted of four 15 cm diameter cell culture dishes seeded with cells, achieving ~1000× coverage. Two replicate sets were included for each of the two OA treatment conditions and one replicate for the untreated control. During the selection period, cells were split 1-2 times per week to avoid overgrowth. An aliquot of cells from each replicate was collected and stored for genomic prep following each split and at the end of the assay. The final samples were collected after 26 days of OA or control treatment.

*Genomic DNA extraction and sequencing*: genomic DNA was extracted from samples collected during the screen and at the final endpoint using a Zymo Quick-gDNA Miniprep kit (Zymo Research, D3025). DNA fragments containing the sgRNA sequences were amplified by PCR using what we calculated to be at least 1000 genomes per sgRNA as templates for each sample. PCR fragments were in-line barcoded using a previously described approach (*41, 119, 120*). Next-generation sequencing (Illumina NextSeq) was performed at the Biopolymers Facility at Harvard Medical School (RRID:SCR_007175). FASTQ sequence files were demultiplexed using in-line barcodes as sequence tags as described previously (*41, 119*). All further processing was done using MAGeCK (v.0.5.4 and 0.5.9.4) (*121*) to generate readcount files and perform robust rank aggregation (RRA) analysis and Maximum likelihood estimation directly on the generated readcount files using default parameters. Plasmid readcounts were used to normalize RRA values.

After removing essential genes, we considered as candidates the top 5% genes overlapping across the two conditions that were over-represented or under-represented in the OA-exposed cell populations compared to the control cells (i.e., representing “susceptibility” and “resistance” genes, respectively). The gene-set enrichment analysis was performed with PANGEA 1.1 Beta using the *Drosophila* GO subsets (GO slim) and a p-value cut-off of 0.05 (*122*).

**qPCR analysis**

Reverse transcription of total RNA (extracted as described above) was performed with the PrimeScript first-strand cDNA synthesis kit (Takara, 6110A) following the manufacturer’s instructions. Samples were stored at -80°C.

Quantitative PCR experiments were conducted on a QuantStudio 6 Flex Real-Time PCR System using the PowerUp™ SYBR™ Green Master Mix following the manufacturer’s instructions. Each sample consisted of at least three biological replicates. We performed three technical replicates per reaction and removed replicates where the standard deviation was equal or greater than 0.5. We determined target specific amplification efficiencies for each gene (Table S4, Table S6). Relative

quantification and statistical analysis were performed with the software qbase+ v3.4 (*123*) using Calibrated Normalized Relative Quantities (CNRQs). Normalization was performed on the geometric mean of the relative quantities for the three most commonly-used reference targets: *Act5C*, *αTub84B* and *RpL32* (*124*). We assessed reference target stability by the geNorm expression stability value (M) and the coefficient of variation of the normalized reference gene relative quantities (CV) (Table S5) (*123, 125*).

**Population genetics analysis**

Publicly available WGS data for 41 *D. sechellia* and 20 *D. simulans* genomes were downloaded from the Sequence Read Archive (PRJNA215932 and PRJNA395473) (*80, 126*) with fasterq-dump (*127*). FASTQ files were processed and aligned as described in the "Experimental evolution" section. Aligned files were phased with samtools phase (*101*) and the genes of interest were extracted with pilon (*128*) to find variation among populations. CDS sequences were aligned based on amino acid information using translatorX (*129*). Gene sequences were aligned using Clustal Omega (*130*). McDonald-Kreitman test, Tajima's D and Fu and Li's statistics were calculated in DNAsp 6 (*131*). FUBAR analyses (Fast, Unconstrained Bayesian AppRoximation) were performed in HyPhy 2.5 (*132*). The consensus sequences of *D. simulans* and *D. sechellia* were computed by SnapGene (www.snapgene.com) "export consensus" with a >50% threshold. The sequence alignment and secondary structures were depicted using ESPript 3 (*133*).

**Histology**

Intestinal tissues and Malpighian tubules were dissected in phosphate-buffered saline (PBS) and fixed at room temperature for 45 min in PBS, 4% paraformaldehyde. Staining of muscle was performed using a mushroom phalloidin AlexFluor647 probe (Life Technologies, A22287, diluted 1:200 in PBS + 0.2% Triton X-100) for 45 min. Samples were mounted in Vectashield with DAPI (Vector Labs, H-1200-10).

A Leica STELLARIS-DIVE confocal microscope was used to generate all confocal images, using a water-immersive 40× objective or an oil-immersive 63× objective. The images were acquired using both HyDs as well as PMTs tailored for the fluorophores of each sample.

**Statistics and reproducibility**

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.

829
830
831 **Fig. 3. A genome-wide pooled CRISPR screen for OA susceptibility and**
832 **resistance genes in *D. melanogaster* cultured cells.**
833 **A** Left: images of S2R+ cells cultured in the absence or presence of OA. Scale bar,
834 50 μm. Right: dose-dependent effects of OA on S2R+ cell viability, through
835 measurement of ATP levels with a CellTiter-Glo assay (see Materials and Methods).
836 **B** Schematic of CRISPR screen design.
837 **C** Top 5% "susceptibility" genes (see Materials and Methods and Data S4). Box plots
838 represent the interquartile ranges for all the replicates of the Z-scores obtained through
839 maximum likelihood estimation to test for selection. The inset (bottom-right) displays
840 the gene-ontology analysis of these genes.
841 **D** As in **C**, for the top 5% "resistance" genes.
842 **E** Venn diagram showing the overlap between all the genes identified in the
843 experimental evolution at G25 and G50, and genes associated with susceptibility and
844 resistance in the CRISPR screen.
845 **F** Subportion of the top panel (CMH test) of the Manhattan plot from Fig. 2C for
846 chromosome arms 2L and 3L, with the red arrows highlighting the regions magnified
847 on the right showing the significant variants and linkage disequilibrium blocks
848 (triangles) for *kraken* and *Alkbh7* (CG14130).
849
850
851 **Fig. 4. Sequence and expression analyses of *kraken* and *Alkbh7*.**
852 **A** AlphaFold protein models for *D. melanogaster* Kraken (AF-O18391-F1-model_v4)
853 and Alkbh7 (AF-Q9VTP1-F1-model_v4) (*134, 135*), highlighting a conserved putative
854 catalytic serine in the active site in Kraken, and the mitochondrial targeting sequence
855 in Alkbh7 (predicted by MitoFates (*77*)). See also Fig. S2A,B.
856 **B** Tissue-specific expression of *D. melanogaster kraken* and *Alkbh7* from bulk RNA-
857 sequencing data from the Fly Atlas 2.0 (*136*).
858 **C** tSNE plots illustrating *D. melanogaster kraken* expression in single-cell
859 transcriptomes of the Malpighian tubules and gut from the Fly Cell Atlas (Stringent 10x
860 datasets) (*72*).
861 **D** Left: whole gut of an adult female expressing GFP (green, endogenous fluorescence
862 signal) under the control of a *kraken* promoter Gal4 driver (genotype: *kraken-*
863 *Gal4*/*UAS-FB1.1*, aged 4-5 days); muscles are labelled by phalloidin (red) to reveal
864 tissue morphology. Imaging was performed with a 40× objective; scale bars, 200 μm.
865 Right: higher magnification images of the regions indicated with dashed boxes.
866 Imaging was performed with a 63× objective; scale bars, 100 μm.
867 **E** Expression levels of the *kraken* and *Alkbh7* in the indicate strains of adult female *D.*
868 *sechellia* and *D. simulans* measured by qPCR (Data S6). Expression is represented
869 as calibrated and normalized relative quantities (CNRQ). Significance was assessed
870 using the unpaired t-test correcting for multiple testing and represented using letter
871 codes.
872 **F** Expression levels of *kraken* and *Alkbh7* in the indicated strains and species at adult
873 or larval stages as indicated from published RNA-seq data (*83-85*). Significance was
874 assessed using the unpaired t-test correcting for multiple testing and represented
875 using letter codes.
876 **G** Expression levels of the candidate genes at generation 0, 25 and 50 of the
877 experimentally-evolved *D. simulans* populations measured by qPCR. Significance was

assessed using the paired t-test correcting for multiple testing and represented using 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.

904 **Supplementary Information**
905
906 **Supplementary Tables**
907
908 **Table S1. *Drosophila* strains.**
909

| 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)* | (*93*) |
| *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)* | (*137*) |
| *Dmel\kraken$^{1}$* | This study |
| *Dmel\Alkbh7$^{1}$* | This study |
| *Dsec\kraken$^{RFP}$* | This study |
| *Dsec\Alkbh7$^{RFP}$* | This study |

910
911
912 **Table S2. *D. simulans* lines used to generate the base populations.**
913 *- see separate Excel file.*
914
915

916 **Table S3. Effective population size estimates from the experimentally-evolved**
917 **populations.**
918

| | 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 |

919
920 **Table S4. qPCR primer efficiency.**
921

| **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 |

922
923 **Table S5. geNorm analysis of reference target stability.**
924 *- see separate Excel file.*
925
926 **Table S6. Oligonucleotides.**
927 *- see separate Excel file.*
928
929

## 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 (*138*).
**B** Bar plots showing the expression levels of *kraken* and *Alkbh7* across various larval *Drosophila* tissues and species. Data from (*83*) ("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

1011 **References**
1012


1013 1. M. Erb, P. Reymond, Molecular Interactions Between Plants and Insect
1014 Herbivores. *Annu Rev Plant Biol* **70**, 527-557 (2019).
1015 2. A. C. Jones, G. W. Felton, J. H. Tumlinson, The dual function of elicitors and
1016 effectors from insects: reviewing the 'arms race' against plant defenses. *Plant*
1017 *Mol Biol* **109**, 427-445 (2022).
1018 3. J. J. Smith, V. Herzig, G. F. King, P. F. Alewood, The insecticidal potential of
1019 venom peptides. *Cell Mol Life Sci* **70**, 3665-3693 (2013).
1020 4. N. Umetsu, Y. Shirai, Development of novel pesticides in the 21st century. *J*
1021 *Pestic Sci* **45**, 54-74 (2020).
1022 5. T. Raisch, S. Raunser, The modes of action of ion-channel-targeting
1023 neurotoxic insecticides: lessons from structural biology. *Nat. Struct. Mol. Biol.*
1024 **30**, 1411-1427 (2023).
1025 6. T. C. Sparks *et al.*, Insecticides, biologics and nematicides: Updates to IRAC's
1026 mode of action classification - a tool for resistance management. *Pesticide*
1027 *biochemistry and physiology* **167**, 104587 (2020).
1028 7. J. R. Busvine, Mechanism of resistance to insecticide in houseflies. *Nature*
1029 **168**, 193-195 (1951).
1030 8. K. Dong, Insect sodium channels and insecticide resistance. *Invert Neurosci*
1031 **7**, 17-30 (2007).
1032 9. K. Matsuda, M. Ihara, D. B. Sattelle, Neonicotinoid Insecticides: Molecular
1033 Targets, Resistance, and Toxicity. *Annu Rev Pharmacol Toxicol* **60**, 241-255
1034 (2020).
1035 10. D. M. Soderlund, J. R. Bloomquist, Neurotoxic actions of pyrethroid
1036 insecticides. *Annu. Rev. Entomol.* **34**, 77-96 (1989).
1037 11. G. Di Prisco *et al.*, Neonicotinoid clothianidin adversely affects insect
1038 immunity and promotes replication of a viral pathogen in honey bees.
1039 *Proceedings of the National Academy of Sciences of the United States of*
1040 *America* **110**, 18466-18471 (2013).
1041 12. C. S. Clarkson, H. J. Temple, A. Miles, The genomics of insecticide
1042 resistance: insights from recent studies in African malaria vectors. *Curr Opin*
1043 *Insect Sci* **27**, 111-115 (2018).
1044 13. E. R. Lucas *et al.*, Genome-wide association studies reveal novel loci
1045 associated with pyrethroid and organophosphate resistance in *Anopheles*
1046 *gambiae* and *Anopheles coluzzii*. *Nat Commun* **14**, 4946 (2023).
1047 14. F. Catteruccia, Malaria-carrying mosquitoes get a leg up on insecticides.
1048 *Nature* **577**, 319-320 (2020).
1049 15. V. Balabanidou *et al.*, Mosquitoes cloak their legs to resist insecticides. *Proc*
1050 *Biol Sci* **286**, 20191091 (2019).
1051 16. V. Balabanidou, L. Grigoraki, J. Vontas, Insect cuticle: a critical determinant of
1052 insecticide resistance. *Curr Opin Insect Sci* **27**, 68-74 (2018).
1053 17. D. Weetman, L. S. Djogbenou, E. Lucas, Copy number variation (CNV) and
1054 insecticide resistance in mosquitoes: evolving knowledge or an evolving
1055 problem? *Curr Opin Insect Sci* **27**, 82-88 (2018).
1056 18. C. T. Zimmer *et al.*, Neofunctionalization of Duplicated P450 Genes Drives the
1057 Evolution of Insecticide Resistance in the Brown Planthopper. *Curr Biol* **28**,
1058 268-274 e265 (2018).

19. A. D. Gloss *et al.*, Evolution in an ancient detoxification pathway is coupled with a transition to herbivory in the drosophilidae. *Mol. Biol. Evol.* **31**, 2441-2456 (2014).
20. C. D. Jones, The genetics of adaptation in *Drosophila sechellia*. *Genetica* **123**, 137-145 (2005).
21. T. O. Auer, M. P. Shahandeh, R. Benton, *Drosophila sechellia*: A Genetic Model for Behavioral Evolution and Neuroecology. *Annu Rev Genet* **55**, 527-554 (2021).
22. S. R'Kha, P. Capy, J. R. David, 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 (1991).
23. T. Dekker, I. Ibba, K. P. Siju, M. C. Stensmyr, B. S. Hansson, Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila melanogaster* sibling, *D. sechellia*. *Curr Biol* **16**, 101-109 (2006).
24. R. Alvarez-Ocana *et al.*, Odor-regulated oviposition behavior in an ecological specialist. *Nat Commun* **14**, 3041 (2023).
25. M. P. Shahandeh *et al.*, Circadian plasticity evolves via regulatory changes in a neuropeptide gene. *Nature* **635**, 951-959 (2024).
26. T. O. Auer *et al.*, Olfactory receptor and circuit evolution promote host specialization. *Nature* **579**, 402-408 (2020).
27. C. D. Jones, The genetic basis of *Drosophila sechellia*'s resistance to a host plant toxin. *Genetics* **149**, 1899-1908 (1998).
28. L. Legal, B. Moulin, J. M. Jallon, 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 (1999).
29. T. A. Markow, Host use and host shifts in *Drosophila*. *Curr Opin Insect Sci* **31**, 139-145 (2019).
30. L. Legal, B. Chappe, J. M. Jallon, Molecular basis of *Morinda citrifolia* (L.): Toxicity on *Drosophila*. *J Chem Ecol* **20**, 1931-1943 (1994).
31. A. Kaczmarek, A. Katarzyna Wronska, M. Irena Bogus, 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 (2024).
32. A. Kaczmarek, A. K. Wronska, M. Kazek, M. I. Bogus, 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**, (2022).
33. M. Amlou, E. Pla, B. Moreteau, J. R. David, 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 (1997).
34. M. Amlou, B. Moreteau, J. R. David, Larval tolerance in the *Drosophila melanogaster* species complex toward the two toxic acids of the *D. sechellia* host plant. *Hereditas* **129**, 7-14 (1998).
35. Y. Huang, D. Erezyilmaz, The Genetics of Resistance to *Morinda* Fruit Toxin During the Postembryonic Stages in *Drosophila sechellia*. *G3 (Bethesda)* **5**, 1973-1981 (2015).

36. Z. Sun, S. Inagaki, K. Miyoshi, K. Saito, S. Hayashi, Osiris gene family defines the cuticle nanopatterns of *Drosophila*. *Genetics* **227**, (2024).
37. T. Ando *et al.*, Nanopore Formation in the Cuticle of an Insect Olfactory Sensillum. *Curr Biol* **29**, 1512-1520 e1516 (2019).
38. J. M. Andrade Lopez *et al.*, Genetic basis of octanoic acid resistance in *Drosophila sechellia*: functional analysis of a fine-mapped region. *Mol. Ecol.* **26**, 1148-1160 (2017).
39. S. M. Lanno, S. J. Shimshak, R. D. Peyser, S. C. Linde, J. D. Coolon, Investigating the role of Osiris genes in *Drosophila sechellia* larval resistance to a host plant toxin. *Ecology and evolution* **9**, 1922-1933 (2019).
40. C. Schlotterer, R. Kofler, E. Versace, R. Tobler, S. U. Franssen, Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity (Edinb)* **116**, 248 (2016).
41. R. Viswanatha, Z. Li, Y. Hu, N. Perrimon, Pooled genome-wide CRISPR screening for basal and context-specific fitness gene essentiality in *Drosophila* cells. *Elife* **7**, (2018).
42. N. Barghi *et al.*, Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLOS Biol.* **17**, e3000128 (2019).
43. T. L. Turner, P. M. Miller, Investigating natural variation in *Drosophila* courtship song by the evolve and resequence approach. *Genetics* **191**, 633-642 (2012).
44. J. K. Kelly, K. A. Hughes, Pervasive Linked Selection and Intermediate-Frequency Alleles Are Implicated in an Evolve-and-Resequencing Experiment of *Drosophila simulans*. *Genetics* **211**, 943-961 (2019).
45. P. Shahrestani *et al.*, The molecular architecture of *Drosophila melanogaster* defense against *Beauveria bassiana* explored through evolve and resequence and quantitative trait locus mapping. *G3 (Bethesda)* **11**, (2021).
46. M. G. Zoh *et al.*, Experimental evolution supports the potential of neonicotinoid-pyrethroid combination for managing insecticide resistance in malaria vectors. *Sci Rep* **11**, 19501 (2021).
47. C. G. Sadia *et al.*, 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 (2024).
48. I. Colson, *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 (2004).
49. A. Long, G. Liti, A. Luptak, O. Tenaillon, Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat. Rev. Genet.* **16**, 567-582 (2015).
50. H. Teotonio, I. M. Chelo, M. Bradic, M. R. Rose, A. D. Long, Experimental evolution reveals natural selection on standing genetic variation. *Nat Genet* **41**, 251-257 (2009).
51. P. D. Keightley, R. W. Ness, D. L. Halligan, P. R. Haddrill, Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics* **196**, 313-320 (2014).
52. S. A. Signor, F. N. New, S. Nuzhdin, A Large Panel of *Drosophila simulans* Reveals an Abundance of Common Variants. *Genome Biol. Evol.* **10**, 189-206 (2018).

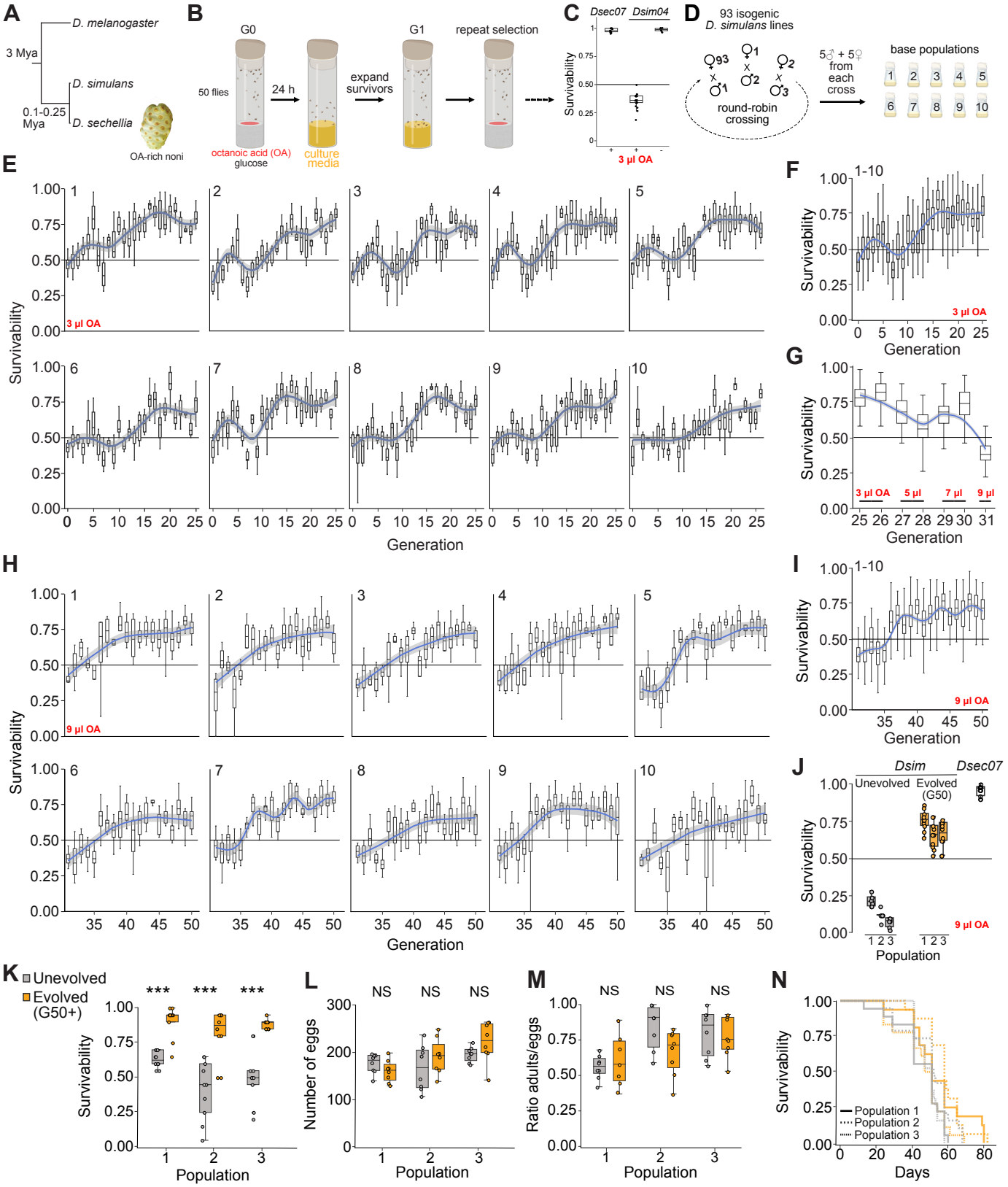
1158 53. S. R'Kha, B. Moreteau, J. A. Coyne, J. R. David, Evolution of a lesser fitness
1159 trait: egg production in the specialist *Drosophila sechellia*. *Genet Res* **69**, 17-
1160 23 (1997).
1161 54. T. A. Markow, S. Beall, L. M. Matzkin, Egg size, embryonic development time
1162 and ovoviviparity in *Drosophila* species. *J Evol Biol* **22**, 430-434 (2009).
1163 55. M. Abe *et al.*, Shortened lifespan induced by a high-glucose diet is associated
1164 with intestinal immune dysfunction in *Drosophila sechellia*. *J Exp Biol* **225**,
1165 (2022).
1166 56. M. Watada *et al.*, Divergence of *Drosophila* species: Longevity and
1167 reproduction under different nutrient balances. *Genes Cells* **25**, 626-636
1168 (2020).
1169 57. C. Vlachos *et al.*, Benchmarking software tools for detecting and quantifying
1170 selection in evolve and resequencing studies. *Genome Biol* **20**, 169 (2019).
1171 58. Y. Kojima, H. Matsumoto, H. Kiryu, Estimation of population genetic
1172 parameters using an EM algorithm and sequence data from experimental
1173 evolution populations. *Bioinformatics* **36**, 221-231 (2020).
1174 59. C. Barata, R. Borges, C. Kosiol, Bait-ER: A Bayesian method to detect targets
1175 of selection in Evolve-and-Resequence experiments. *J Evol Biol* **36**, 29-44
1176 (2023).
1177 60. W. G. Cochran, Some Methods for Strengthening the Common Chi-Squared
1178 Tests. *Biometrics* **10**, 417-451 (1954).
1179 61. N. Mantel, W. Haenszel, Statistical aspects of the analysis of data from
1180 retrospective studies of disease. *J Natl Cancer Inst* **22**, 719-748 (1959).
1181 62. P. Orozco-terWengel *et al.*, Adaptation of *Drosophila* to a novel laboratory
1182 environment reveals temporally heterogeneous trajectories of selected alleles.
1183 *Mol. Ecol.* **21**, 4931-4941 (2012).
1184 63. Y. Hu, A. Comjean, N. Perrimon, S. E. Mohr, The *Drosophila* Gene
1185 Expression Tool (DGET) for expression analyses. *BMC Bioinf.* **18**, 98 (2017).
1186 64. Y. Xu *et al.*, CRISPR screens in *Drosophila* cells identify Vsg as a Tc toxin
1187 receptor. *Nature* **610**, 349-355 (2022).
1188 65. A. Borrull, G. Lopez-Martinez, M. Poblet, R. Cordero-Otero, N. Rozes, New
1189 insights into the toxicity mechanism of octanoic and decanoic acids on
1190 *Saccharomyces cerevisiae*. *Yeast* **32**, 451-460 (2015).
1191 66. A. V. Emelyanov *et al.*, *Drosophila* TAP/p32 is a core histone chaperone that
1192 cooperates with NAP-1, NLP, and nucleophosmin in sperm chromatin
1193 remodeling during fertilization. *Genes Dev* **28**, 2027-2040 (2014).
1194 67. K. A. Galindo, W. J. Lu, J. H. Park, J. M. Abrams, The Bax/Bak ortholog in
1195 *Drosophila*, Debcl, exerts limited control over programmed cell death.
1196 *Development* **136**, 275-283 (2009).
1197 68. H. Y. Edwin Chan, S. J. Harris, C. J. O'Kane, Identification and
1198 characterization of kraken, a gene encoding a putative hydrolytic enzyme in
1199 *Drosophila melanogaster*. *Gene* **222**, 195-201 (1998).
1200 69. L. S. Zhang *et al.*, ALKBH7-mediated demethylation regulates mitochondrial
1201 polycistronic RNA processing. *Nat Cell Biol* **23**, 684-691 (2021).
1202 70. A. Solberg *et al.*, Deletion of mouse Alkbh7 leads to obesity. *J Mol Cell Biol* **5**,
1203 194-203 (2013).
1204 71. K. Kumar, A. Mhetre, G. S. Ratnaparkhi, S. S. Kamat, A Superfamily-wide
1205 Activity Atlas of Serine Hydrolases in *Drosophila melanogaster*. *Biochemistry*
1206 **60**, 1312-1324 (2021).

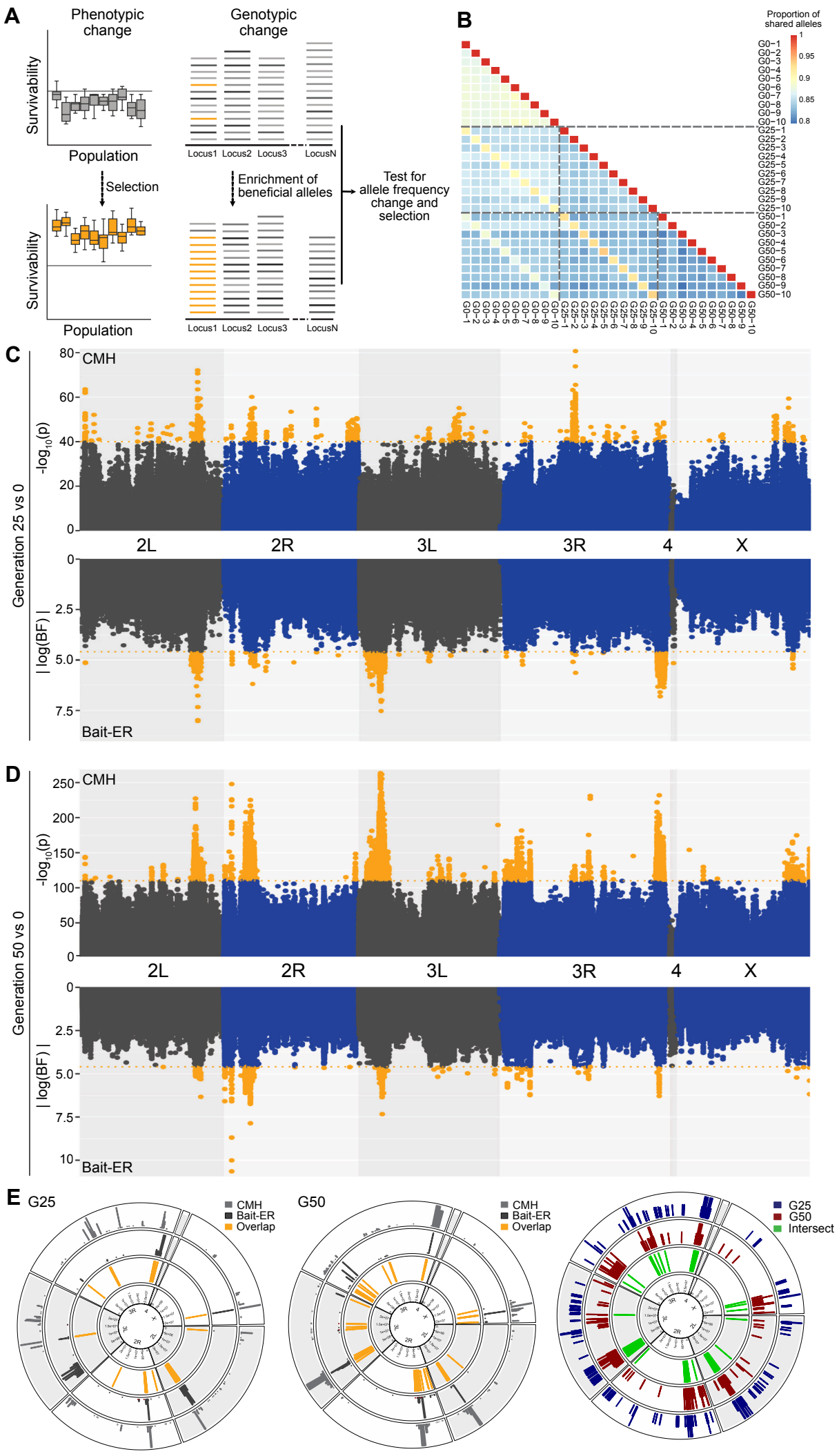
1207 72. H. Li *et al.*, Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult
1208 fruit fly. *Science* **375**, eabk2432 (2022).
1209 73. B. Lemaitre, I. Miguel-Aliaga, The digestive tract of *Drosophila melanogaster*.
1210 *Annu Rev Genet* **47**, 377-404 (2013).
1211 74. F. Teufel *et al.*, SignalP 6.0 predicts all five types of signal peptides using
1212 protein language models. *Nat. Biotechnol.* **40**, 1023-1025 (2022).
1213 75. J. Yang *et al.*, A *Drosophila* systems approach to xenobiotic metabolism.
1214 *Physiol Genomics* **30**, 223-231 (2007).
1215 76. B. Xu, D. Liu, Z. Wang, R. Tian, Y. Zuo, Multi-substrate selectivity based on
1216 key loops and non-homologous domains: new insight into ALKBH family. *Cell*
1217 *Mol Life Sci* **78**, 129-141 (2021).
1218 77. Y. Fukasawa *et al.*, MitoFates: improved prediction of mitochondrial targeting
1219 sequences and their cleavage sites. *Molecular & cellular proteomics : MCP*
1220 **14**, 1113-1126 (2015).
1221 78. C. L. Chen *et al.*, Proteomic mapping in live *Drosophila* tissues using an
1222 engineered ascorbate peroxidase. *Proceedings of the National Academy of*
1223 *Sciences of the United States of America* **112**, 12093-12098 (2015).
1224 79. A. Sen, R. T. Cox, Loss of *Drosophila* Clueless differentially affects the
1225 mitochondrial proteome compared to loss of Sod2 and Pink1. *Frontiers in*
1226 *physiology* **13**, 1004099 (2022).
1227 80. D. R. Schrider, J. Ayroles, D. R. Matute, A. D. Kern, Supervised machine
1228 learning reveals introgressed loci in the genomes of *Drosophila simulans* and
1229 *D. sechellia*. *PLOS Genet.* **14**, e1007341 (2018).
1230 81. J. H. McDonald, M. Kreitman, Adaptive protein evolution at the Adh locus in
1231 *Drosophila*. *Nature* **351**, 652-654 (1991).
1232 82. B. Murrell *et al.*, FUBAR: a fast, unconstrained bayesian approximation for
1233 inferring selection. *Mol. Biol. Evol.* **30**, 1196-1205 (2013).
1234 83. K. Watanabe *et al.*, Interspecies Comparative Analyses Reveal Distinct
1235 Carbohydrate-Responsive Systems among *Drosophila* Species. *Cell Rep* **28**,
1236 2594-2607 e2597 (2019).
1237 84. S. Ma *et al.*, Comparative transcriptomics across 14 *Drosophila* species
1238 reveals signatures of longevity. *Aging Cell* **17**, e12740 (2018).
1239 85. S. Kalra, S. Lanno, G. Sanchez, J. D. Coolon, cis- and trans-regulatory
1240 contributions to a hierarchy of factors influencing gene expression variation.
1241 *Communications biology* **7**, 1563 (2024).
1242 86. M. N. Mota, M. Matos, N. Bahri, I. Sa-Correia, Shared and more specific
1243 genetic determinants and pathways underlying yeast tolerance to acetic,
1244 butyric, and octanoic acids. *Microb Cell Fact* **23**, 71 (2024).
1245 87. P. Liu *et al.*, Membrane stress caused by octanoic acid in *Saccharomyces*
1246 *cerevisiae*. *Appl Microbiol Biotechnol* **97**, 3239-3251 (2013).
1247 88. .
1248 89. J. A. T. Dow, M. Simons, M. F. Romero, *Drosophila melanogaster*: a simple
1249 genetic model of kidney structure, function and disease. *Nat Rev Nephrol* **18**,
1250 417-434 (2022).
1251 90. Y. T. Gutierrez-Guerrero *et al.*, Genomic consequences of dietary
1252 diversification and parallel evolution due to nectarivory in leaf-nosed bats.
1253 *Gigascience* **9**, (2020).
1254 91. L. Gandara *et al.*, Pervasive sublethal effects of agrochemicals on insects at
1255 environmentally relevant concentrations. *Science* **386**, 446-453 (2024).

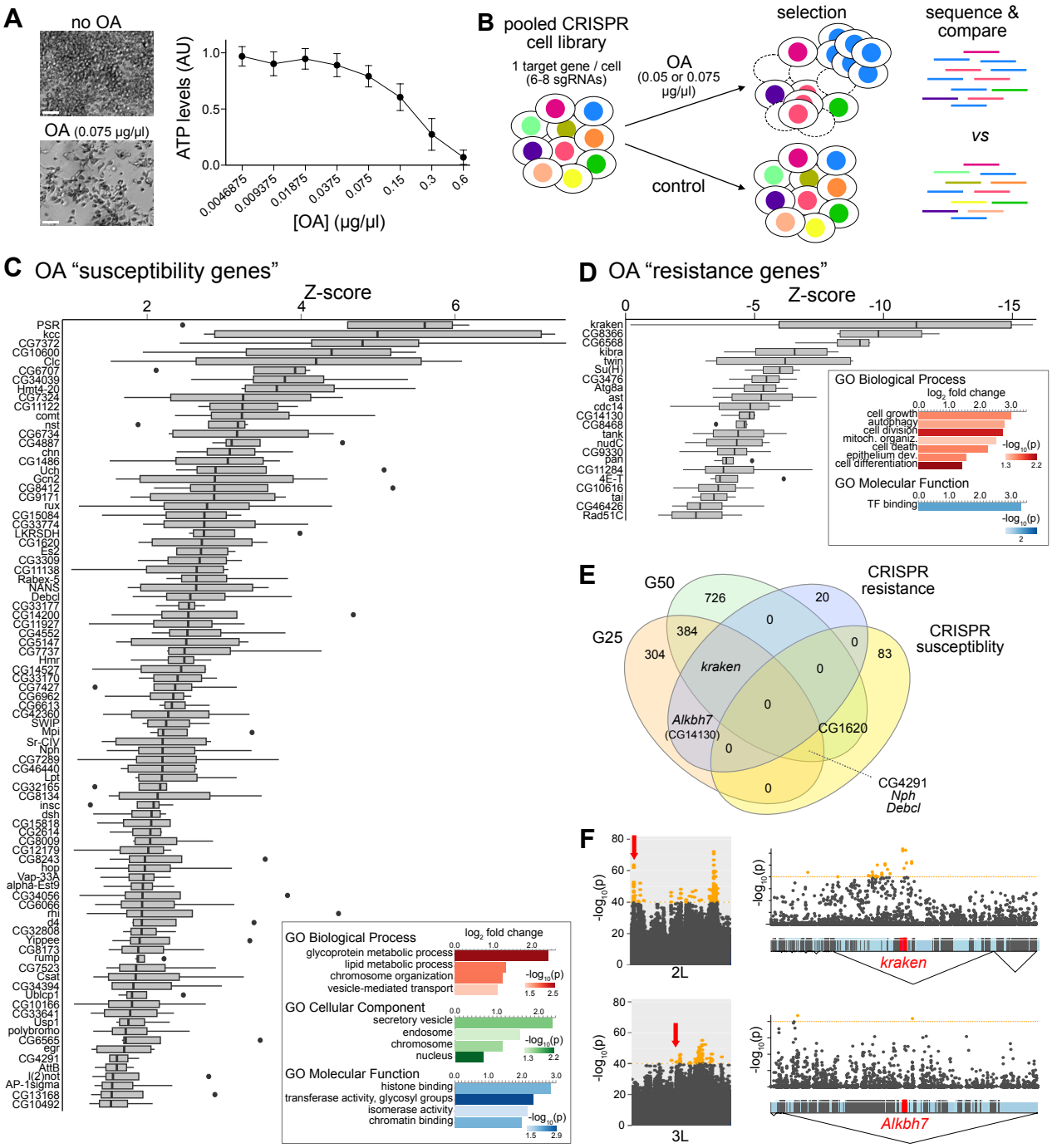
92. F. Port *et al.*, A large-scale resource for tissue-specific CRISPR mutagenesis in *Drosophila*. *Elife* **9**, (2020).
93. T. Lenče, Johannes Gutenberg University Mainz (2021).
94. S. Kondo, R. Ueda, Highly Improved Gene Targeting by Germline-Specific Cas9 Expression in *Drosophila*. *Genetics*, (2013).
95. V. Croset *et al.*, Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLOS Genet.* **6**, e1001064 (2010).
96. J. Bischof, R. K. Maeda, M. Hediger, F. Karch, K. Basler, An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 3312-3317 (2007).
97. T. M. Therneau. (2024).
98. S. N. Wood, Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 3-36 (2011).
99. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
100. H. Li, Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv*, (2013).
101. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
102. R. Kofler, R. V. Pandey, C. Schlotterer, PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435-3436 (2011).
103. P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
104. S. Purcell *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).
105. S. Purcell. (2020).
106. C. Vlachos, R. Kofler, MimicrEE2: Genome-wide forward simulations of Evolve and Resequencing studies. *PLoS Comput Biol* **14**, e1006413 (2018).
107. T. Taus, A. Futschik, C. Schlotterer, Quantifying Selection with Pool-Seq Time Series Data. *Mol. Biol. Evol.* **34**, 3023-3034 (2017).
108. C. Zhang, S. S. Dong, J. Y. Xu, W. M. He, T. L. Yang, PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786-1788 (2019).
109. S. S. Dong *et al.*, LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in bioinformatics* **22**, (2021).
110. J. C. Barrett, B. Fry, J. Maller, M. J. Daly, Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265 (2005).
111. Z. Gu, L. Gu, R. Eils, M. Schlesner, B. Brors, circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).
112. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
113. A. M. Altenhoff *et al.*, OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res.* **49**, D373-D379 (2021).

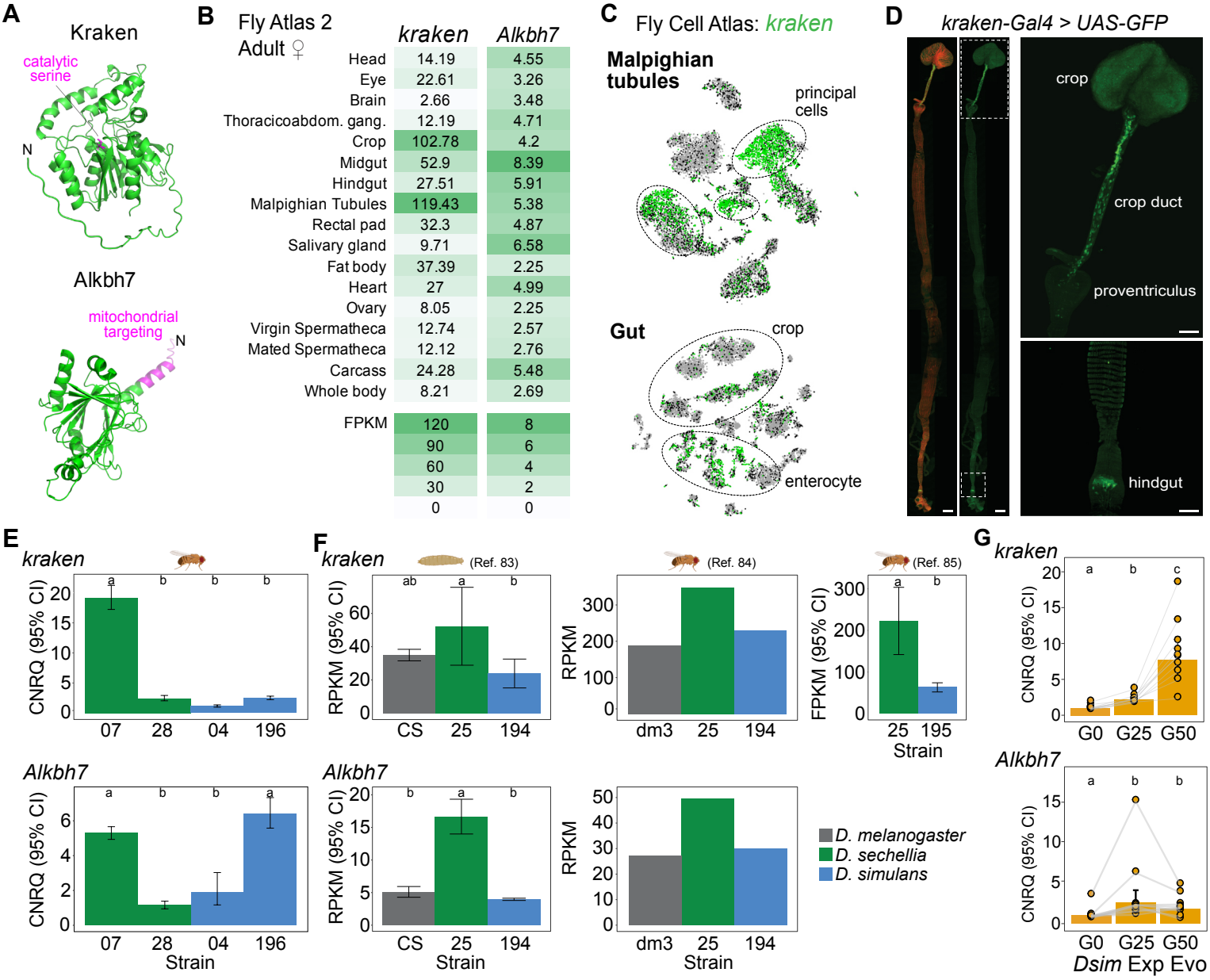
114. B. T. Sherman *et al.*, DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216-W221 (2022).
115. W. Huang da, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
116. C. Camacho *et al.*, BLAST+: architecture and applications. *BMC Bioinf.* **10**, 421 (2009).
117. E. W. Sayers *et al.*, Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res.* **53**, D20-D29 (2025).
118. A. Ozturk-Colak *et al.*, FlyBase: updates to the *Drosophila* genes and genomes database. *Genetics* **227**, (2024).
119. R. Viswanatha *et al.*, Higher resolution pooled genome-wide CRISPR knockout screening in *Drosophila* cells using integration and anti-CRISPR (IntAC). *Nat Commun* **16**, 6498 (2025).
120. R. Viswanatha *et al.*, Pooled CRISPR Screens in *Drosophila* Cells. *Current protocols in molecular biology* **129**, e111 (2019).
121. W. Li *et al.*, MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).
122. Y. Hu *et al.*, PANGEA: a new gene set enrichment tool for *Drosophila* and common research organisms. *Nucleic Acids Res.* **51**, W419-W426 (2023).
123. J. Hellemans, G. Mortier, A. De Paepe, F. Speleman, J. Vandesompele, qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**, R19 (2007).
124. J. Lu, C. Yang, Y. Zhang, H. Pan, 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 (2018).
125. J. Vandesompele *et al.*, Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, RESEARCH0034 (2002).
126. R. L. Rogers *et al.*, Landscape of standing variation for tandem duplications in *Drosophila yakuba* and *Drosophila simulans*. *Mol. Biol. Evol.* **31**, 1750-1766 (2014).
127. R. Leinonen *et al.*, Improvements to services at the European Nucleotide Archive. *Nucleic Acids Res.* **38**, D39-45 (2010).
128. B. J. Walker *et al.*, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014).
129. F. Abascal, R. Zardoya, M. J. Telford, TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7-13 (2010).
130. F. Madeira *et al.*, The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* **52**, W521-W525 (2024).
131. J. Rozas *et al.*, DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol. Biol. Evol.* **34**, 3299-3302 (2017).
132. S. L. Kosakovsky Pond *et al.*, HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol. Biol. Evol.* **37**, 295-299 (2020).

133. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320-324 (2014).
134. M. Varadi *et al.*, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439-D444 (2022).
135. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
136. S. A. Krause, G. Overend, J. A. T. Dow, D. P. Leader, FlyAtlas 2 in 2022: enhancements to the *Drosophila melanogaster* expression atlas. *Nucleic Acids Res.* **50**, D1010-D1015 (2022).
137. D. Hadjieconomou *et al.*, Flybow: genetic multicolor cell labeling for neural circuit analysis in *Drosophila melanogaster*. *Nat Methods* **8**, 260-266 (2011).
138. G. Bontonou *et al.*, Evolution of chemosensory tissues and cells across ecologically diverse Drosophilids. *Nat Commun* **15**, 1047 (2024).

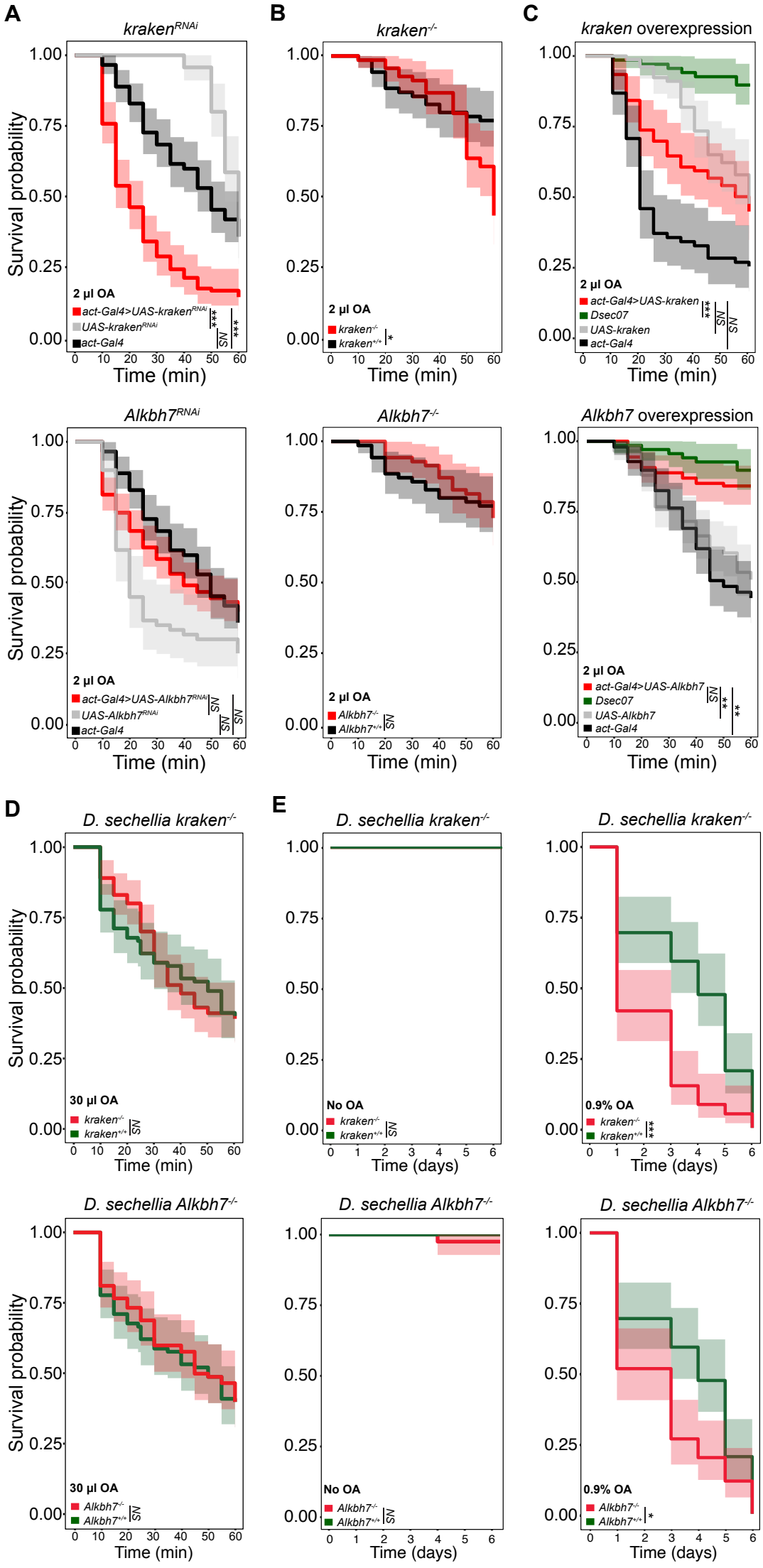

A
D. melanogaster
3 Mya
D. simulans
0.1-0.25 Mya
D. sechellia
OA-rich noni
B
G0
50 flies
24 h
octanoic acid (OA)
glucose
culture media
expand survivors
G1
repeat selection
C
Dsec07
Dsim04
Survivability
3 µl OA
D
93 isogenic D. simulans lines
round-robin crossing
5♂ + 5♀ from each cross
base populations
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
-log10(p)
| log(BF) |
Bait-ER
2L 2R 3L 3R 4 X
E
G25
G50
CMH
Bait-ER
Overlap
G25
G50
Intersect

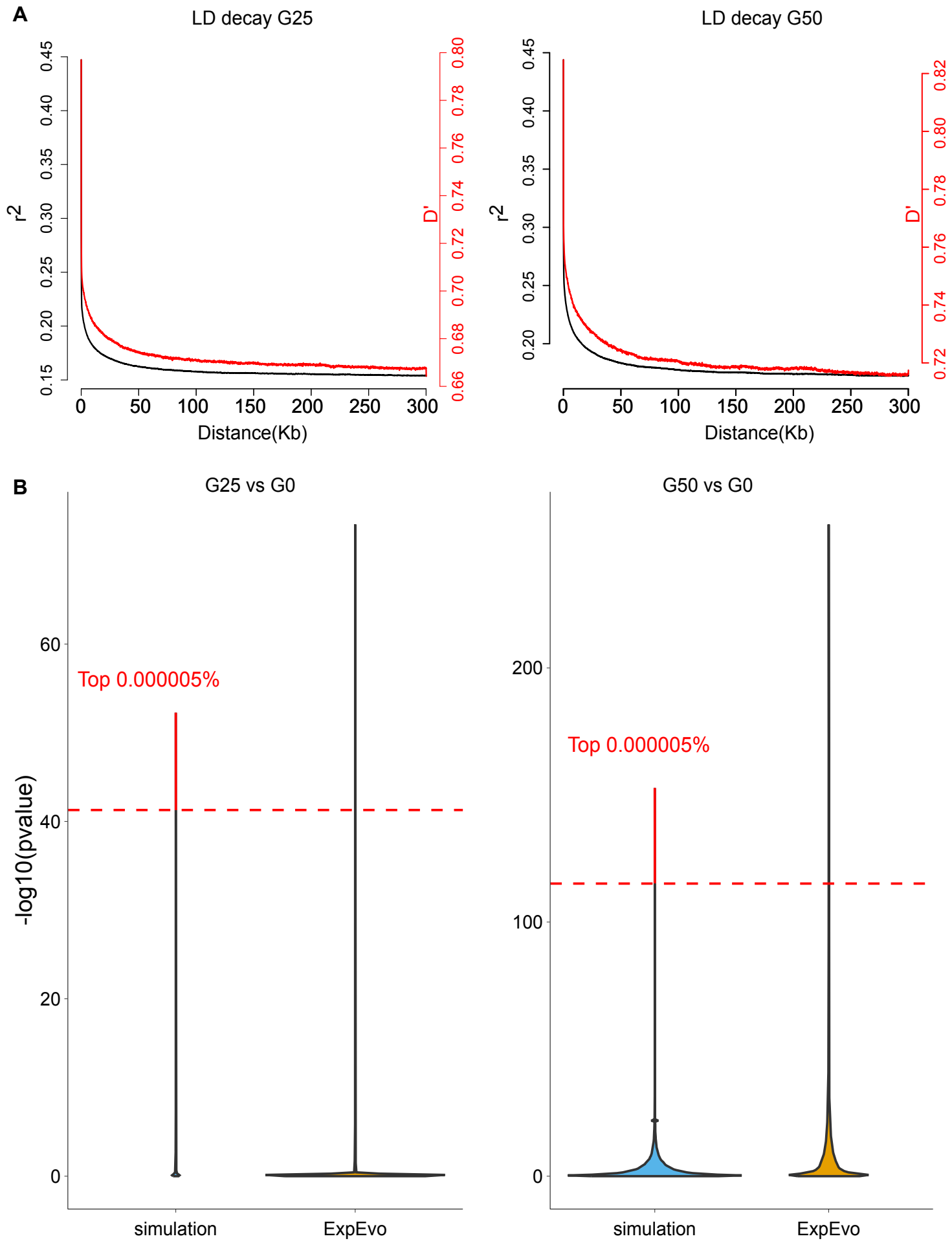

A
no OA
OA (0.075 µg/µl)
ATP levels (AU)
1.0
0.5
0.0
0.0046875
0.009375
0.01875
0.0375
0.075
0.15
0.3
0.6
[OA] (µg/µl)
B
pooled CRISPR cell library
1 target gene / cell (6-8 sgRNAs)
OA (0.05 or 0.075 µg/µl)
control
selection
sequence & compare
vs
C
OA "susceptibility genes"
Z-score
2
4
6
PSR
kcc
CG7372
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
Ubiscp1
CG10166
CG33641
Usp1
polybromo
CG6565
egr
CG4291
AttB
l(2)not
AP-1sigma
CG13168
CG10492
GO Biological Process
log2 fold change
glycoprotein metabolic process
lipid metabolic process
chromosome organization
vesicle-mediated transport
-log10(p)
GO Cellular Component
secretory vesicle
endosome
chromosome
nucleus
GO Molecular Function
histone binding
transferase activity, glycosyl groups
isomerase activity
chromatin binding
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)
GO Molecular Function
TF binding
E
G50
G25
CRISPR resistance
CRISPR susceptiblity
726
20
384
304
83
kraken
Alkbh7 (CG14130)
CG1620
CG4291
Nph
Debcl
F
-log10(p)
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
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. 83)
RPKM (95% CI)
CS 25 194
(Ref. 84)
RPKM
dm3 25 194
(Ref. 85)
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

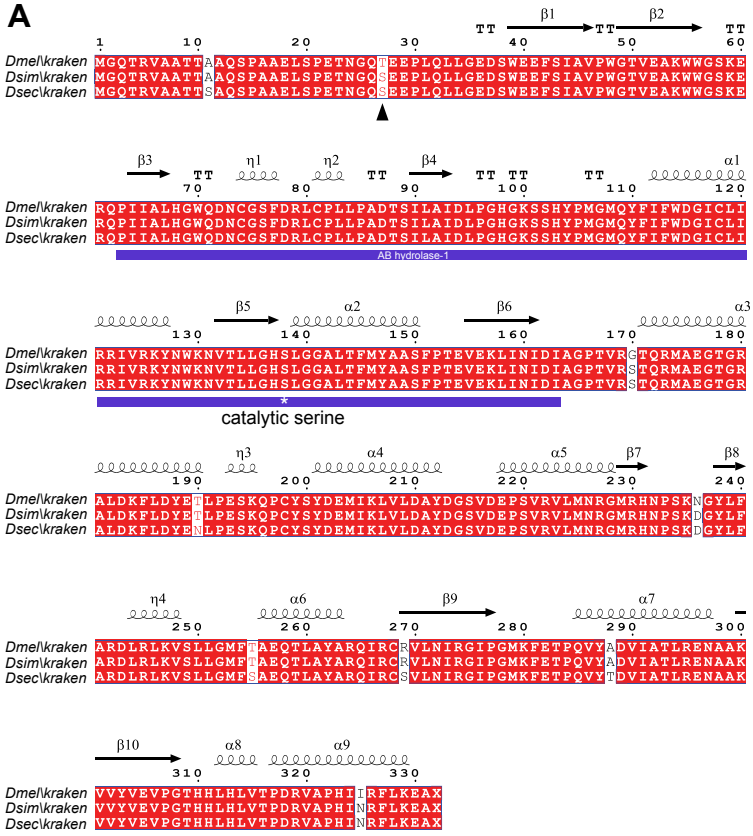
Marconcini, Figure 5



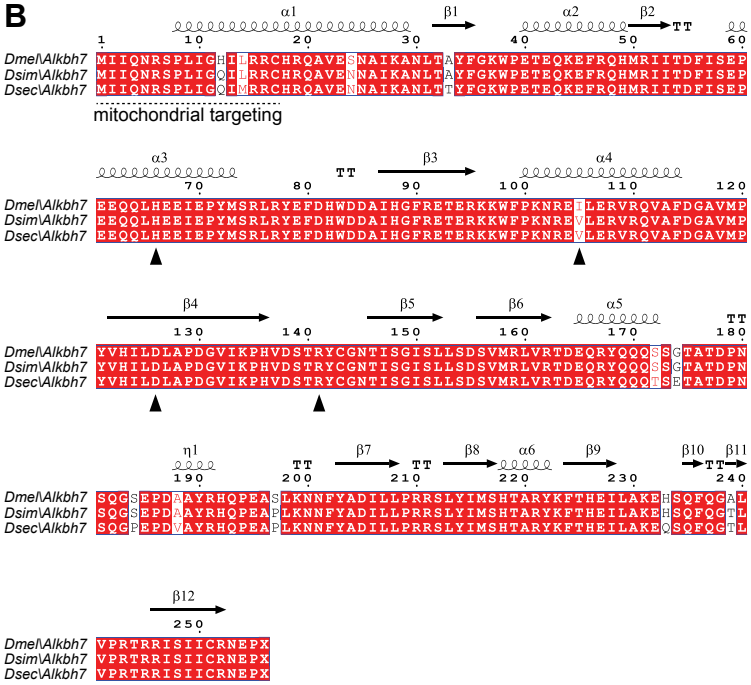
Marconcini, Figure S1



**A**




**B**




**C**

McDonald & Kreitman Test

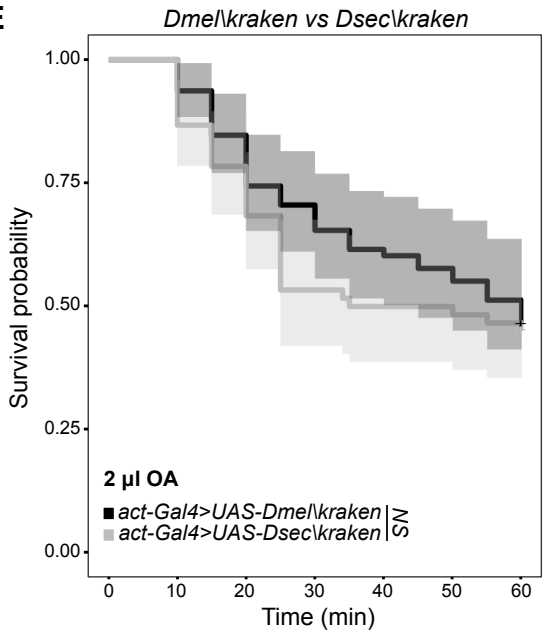
| | NI | alpha | p-value |
|---|---|---|---|
| ***kraken*** | 0.083 | 0.917 | 0.018* |
| ***Alkbh7*** | 0.544 | 0.456 | 0.477 |

**D**

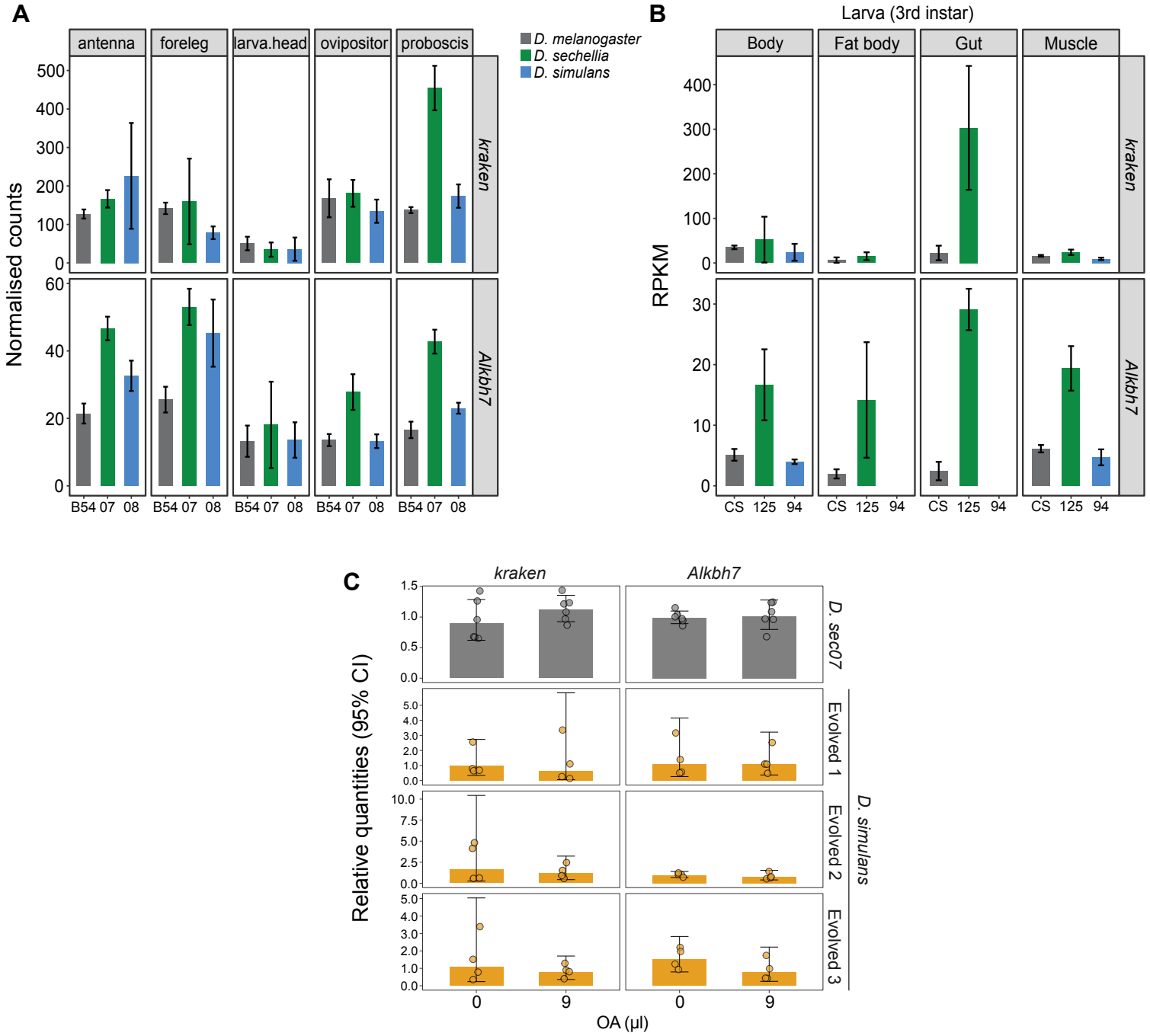
HyPhy FUBAR

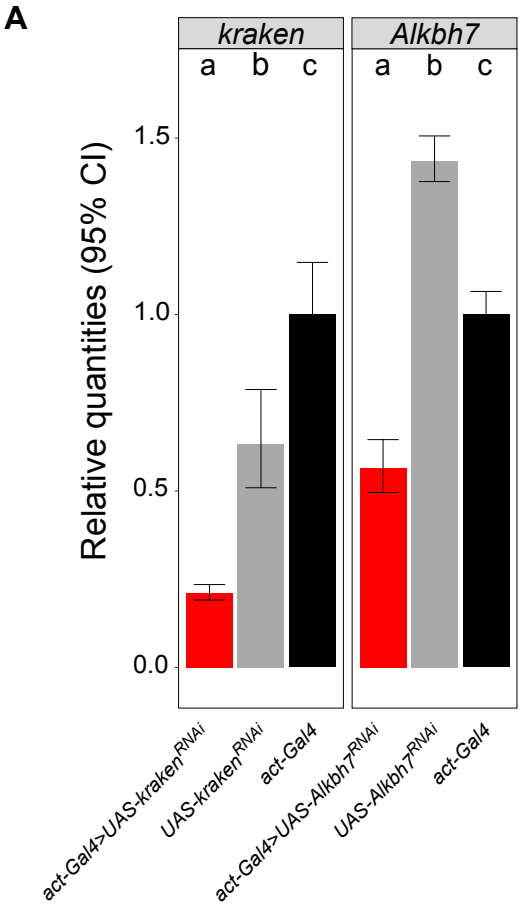
| | position | alpha | beta | post. prob. | codons | Freq *Dsec* | Freq *Dsim* |
|---|---|---|---|---|---|---|---|
| ***kraken*** | 27 | 2.503 | 40.230 | 0.979 | S / A / T | 1 / 0 / 0 | 0.658 / 0.237 / 0.105 |
| ***Alkbh7*** | 66 | 0.768 | 9.440 | 0.977 | H / Y | 1 / 0 | 0.806 / 0.194 |
| | 105 | 0.665 | 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 |

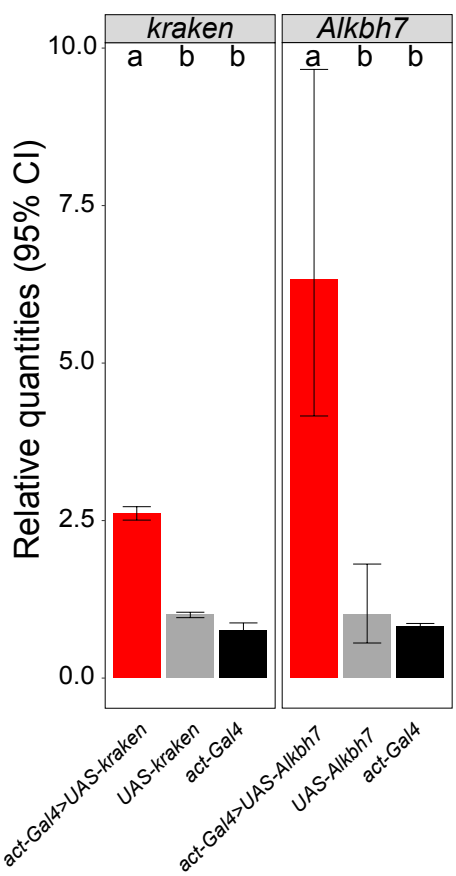
**E**

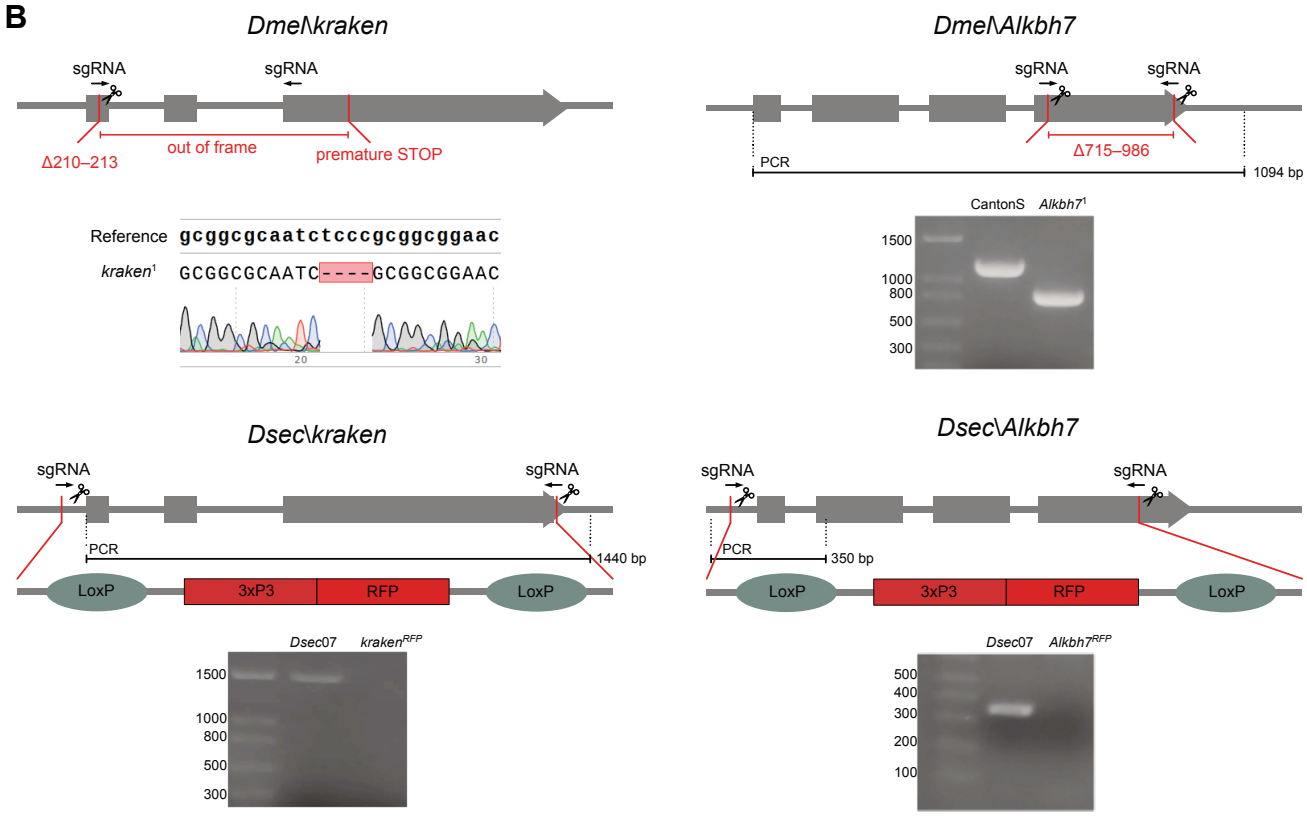


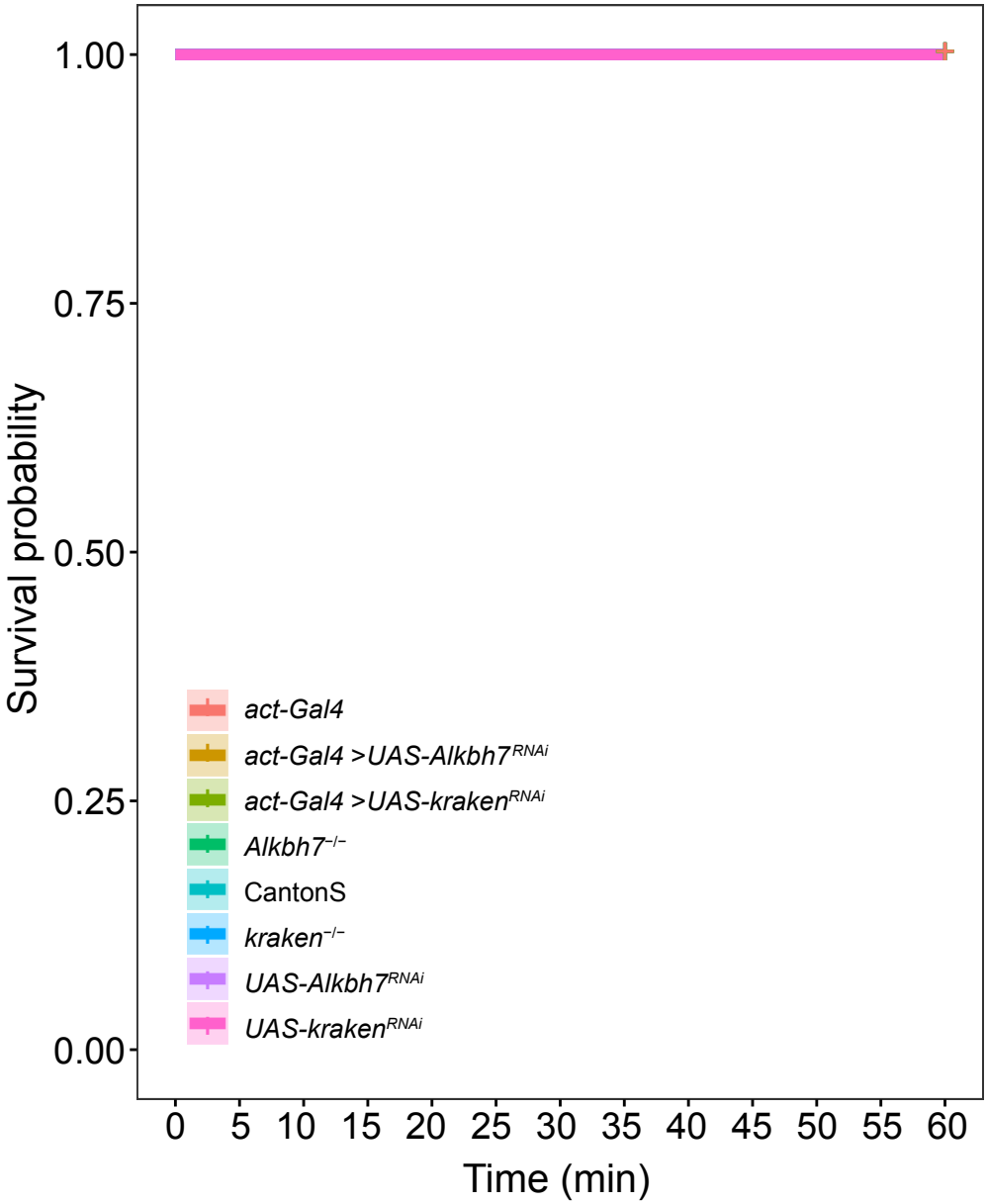
Marconcini, Figure S3




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
20
30
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