RESEARCH ARTICLE

Figures and figure supplements

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

Michele Marconcini *et al*.

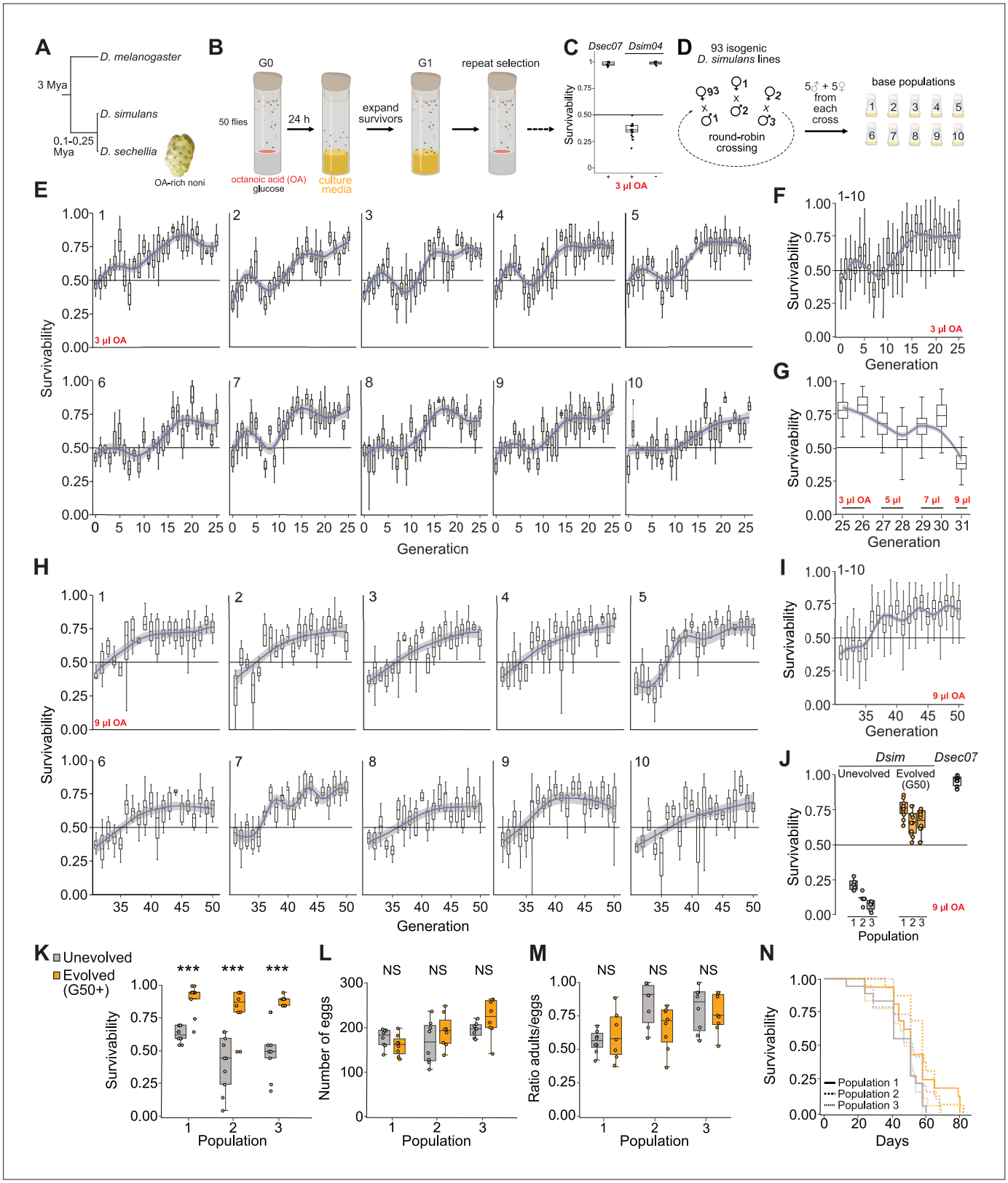

Figure 1. Experimental evolution of *D. simulans* with increased octanoic acid (OA) resistance. (**A**) Phylogeny of the drosophilid trio studied in this work. Mya: million years ago. (**B**) Schematic of the OA resistance tube assay (see Materials and methods) and selection procedure. (**C**) Differences 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 10 *D. simulans* populations over generations 0–25. Box plots represent the

Figure 1 continued on next page

Figure 1 continued

interquartile ranges of the survivability for all the replicates; blue lines show the generalized additive model (GAM) regression; gray 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 generations 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 ***Figure 1—source data 1*** for raw data. Panel B was created with BioRender.com.

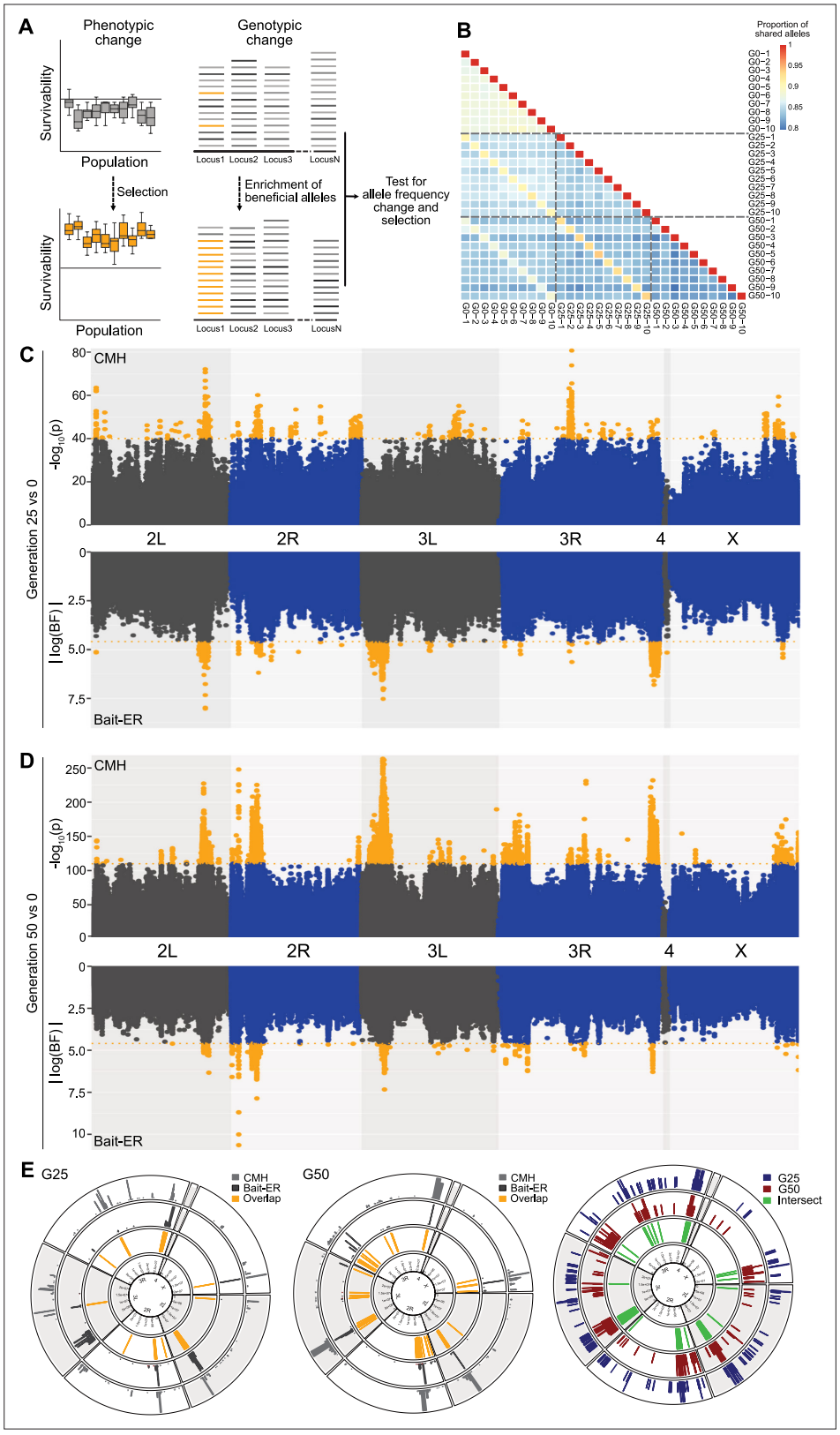

Figure 2. Dynamic directional selection of allele frequencies during experimental evolution. (**A**) Schematic of the analysis. (**B**) Identity-by-state analysis. The heatmap depicts the proportion of shared alleles between populations, with values ranging from 0.8 (blue, lowest shared proportion) to 1 (red, identical alleles). (**C,D**) Manhattan plots showing the probability of allele frequency changes between (**C**) generations 0 and 25, and (**D**) generations 0

Figure 2 continued on next page

Figure 2 continued

and 50. Top panel shows the results of the Cochran-Mantel-Haenszel (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 (***Figure 2—figure supplement 1***). 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). (**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 generations 25 and 50 (right). Raw data in ***Figure 2—source data 1***.

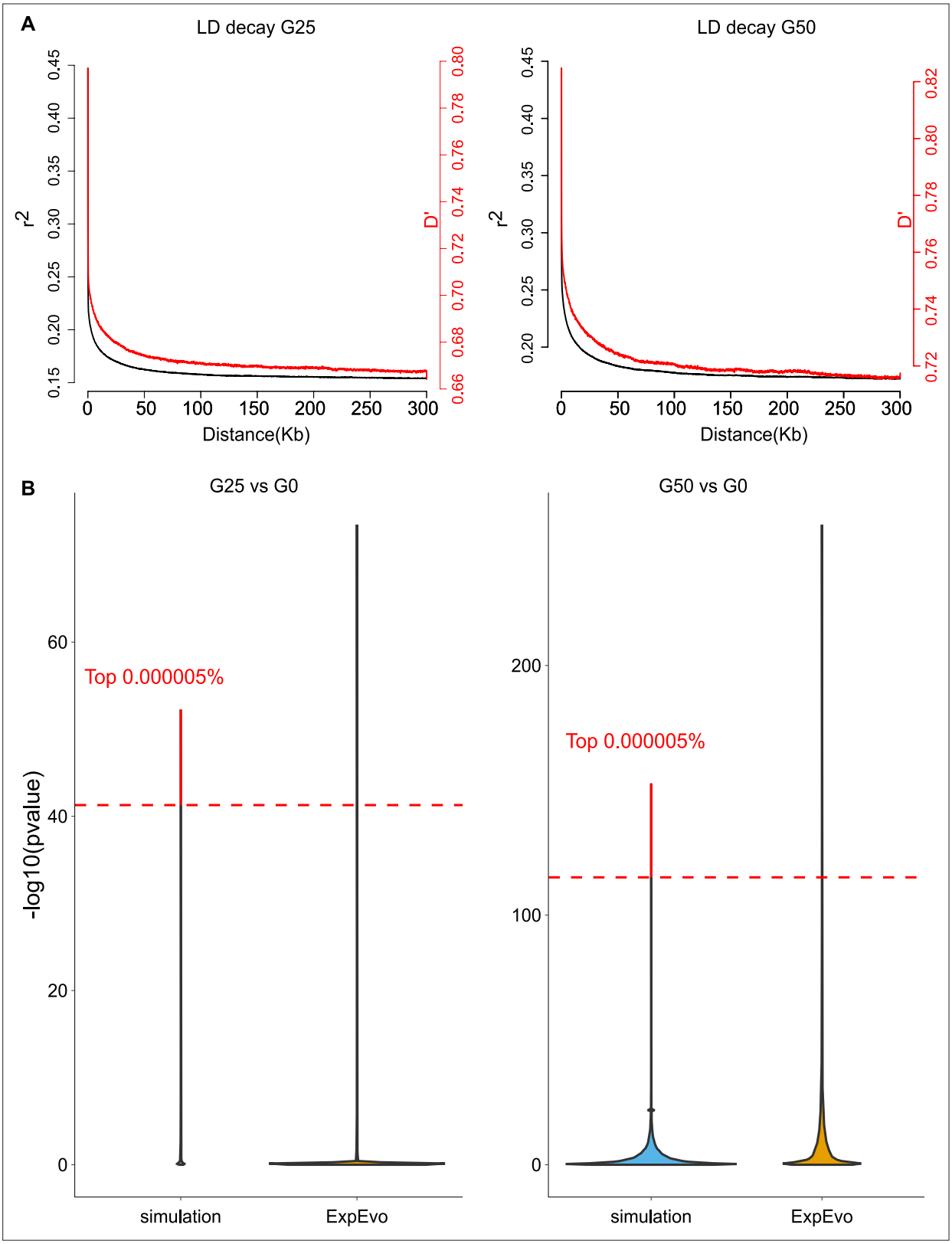

Figure 2—figure supplement 1. Linkage disequilibrium decay, and comparison of simulated and observed p-value distributions in the experimentally evolved populations. (**A**) Linkage disequilibrium decay at generations 25 (left) and 50 (right). Black curves represent the trajectory of r2 values (y-axis 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 Cochran-Mantel-Haenszel (CMH) test, comparing allele frequency changes between generations 25 and 0 (left), and generations 50 and

Figure 2—figure supplement 1 continued on next page

Figure 2—figure supplement 1 continued

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.

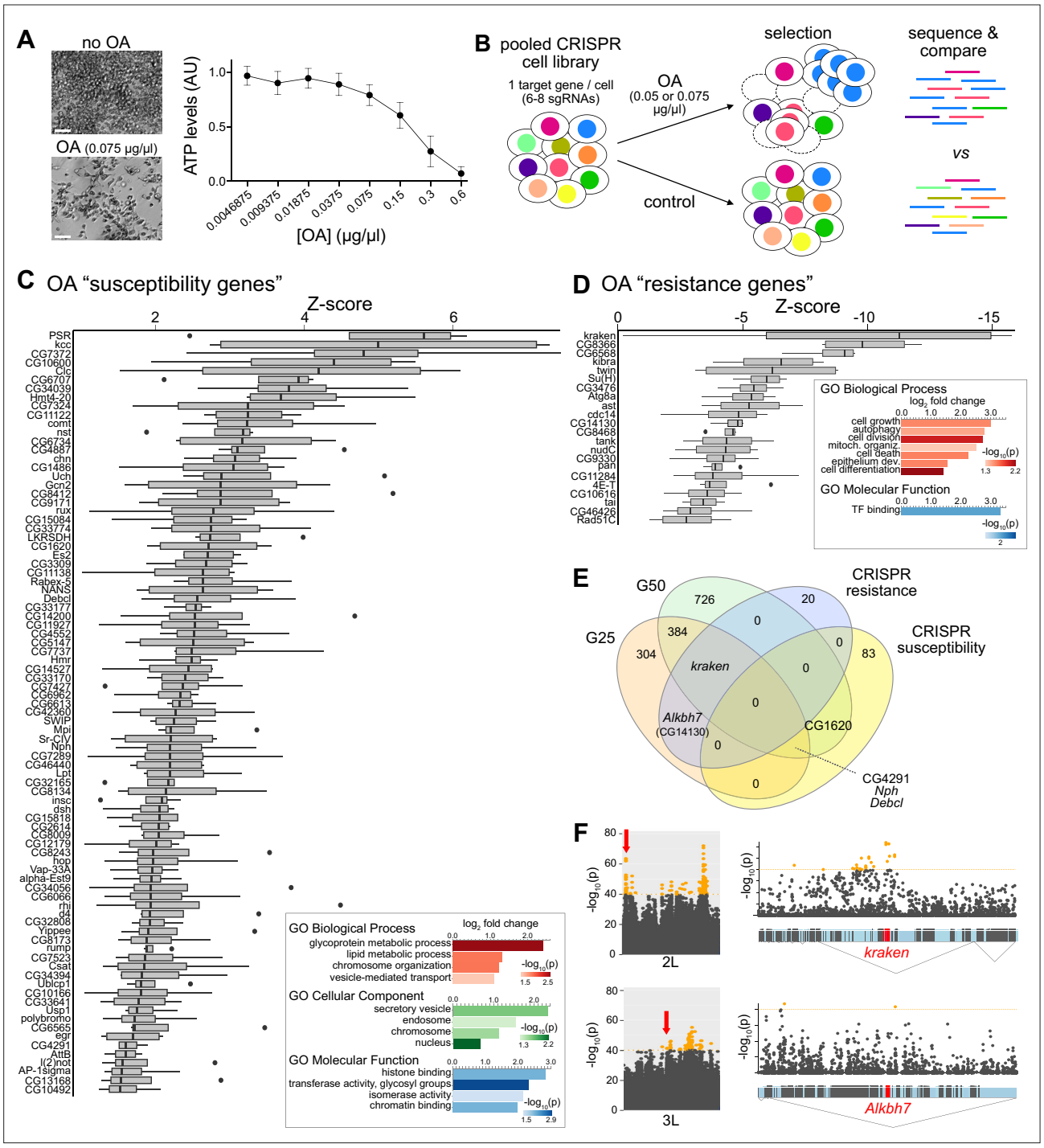

Figure 3. A genome-wide pooled CRISPR screen for octanoic acid (OA) susceptibility and resistance genes in *D. melanogaster* cultured cells. (**A**) Left: images of S2R+ cells cultured in the absence or presence of OA. Scale bar, 50 µm. Right: dose-dependent effects of OA on S2R+ cell viability, through measurement of ATP levels with a CellTiter-Glo assay (see Materials and methods). (**B**) Schematic of CRISPR screen design. (**C**) Top 5% 'susceptibility' genes (see Materials and methods and ***Figure 3—source data 1***). Box plots represent the interquartile ranges for all the replicates of the Z-scores obtained through maximum likelihood estimation to test for selection. The inset (bottom-right) displays the gene ontology analysis of these genes. (**D**) As in **C**, for the top 5% 'resistance' genes. (**E**) Venn diagram showing the overlap between all the genes identified in the experimental evolution at G25 and G50, and genes associated with susceptibility and resistance in the CRISPR screen. (**F**) Subportion of the top panel (Cochran–Mantel–Haenszel [CMH] test) of the Manhattan plot from ***Figure 2C*** for chromosome arms 2L and 3L, with the red arrows highlighting the regions magnified on the right showing the significant variants and linkage disequilibrium blocks (triangles) for *kraken* and *Alkbh7* (CG14130).

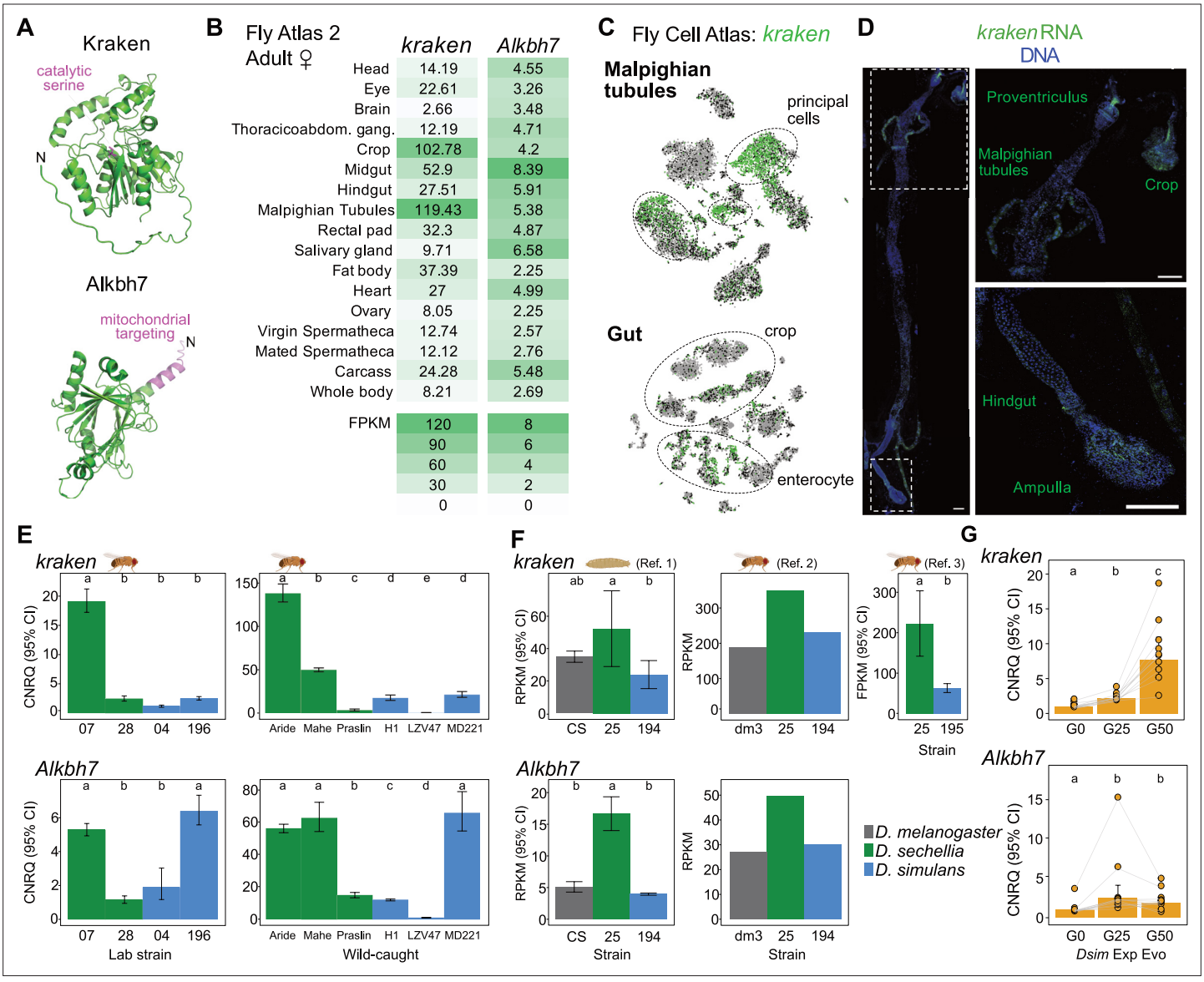

Figure 4. Sequence and expression analyses of *kraken* and *Alkbh7*. (**A**) AlphaFold protein models for *D. melanogaster* Kraken (AF-O18391-F1-model_v4) and Alkbh7 (AF-Q9VTP1-F1-model_v4) (***Jumper et al., 2021***; ***Varadi et al., 2022***), highlighting a conserved putative catalytic serine in the active site in Kraken, and the mitochondrial targeting sequence in Alkbh7 (predicted by MitoFates; ***Fukasawa et al., 2015***). See also ***Figure 4—figure supplement 1***. (**B**) Tissue-specific expression of *D. melanogaster kraken* and *Alkbh7* from bulk RNA-sequencing data from the Fly Atlas 2.0 (***Krause et al., 2022***). (**C**) tSNE plots illustrating *D. melanogaster kraken* expression in single-cell transcriptomes of the Malpighian tubules and gut from the Fly Cell Atlas (Stringent 10× datasets) (***Li et al., 2022***). (**D**) Left: RNAscope detection of *kraken* (green) transcripts in the gut and Malpighian tubules, with nuclei counterstained with DAPI (blue). Right: higher-magnification images showing *kraken* transcript expression in the indicated tissues. Scale bars, 200 µm. This expression pattern was observed in tissues from >20 individuals. (**E**) Expression levels of the *kraken* and *Alkbh7* in the indicated laboratory strains (left panel) and wild-caught strains (right panel) (***Shahandeh et al., 2024***; ***Shahandeh et al., 2026***) of adult female *D. sechellia* and *D. simulans* measured by quantitative PCR (qPCR) (***Figure 4—source data 1***). Expression is represented as calibrated and normalized relative quantities (CNRQ). Significance was assessed using the unpaired t-test correcting for multiple testing and represented using letter codes. (**F**) Expression levels of *kraken* and *Alkbh7* in the indicated strains and species at larval or adult stages as indicated from published RNA-sequencing data (References: 1, ***Watanabe et al., 2019***; 2 ***Ma et al., 2018***; 3 ***Kalra et al., 2024***). Significance was assessed using the unpaired t-test correcting for multiple testing and represented using letter codes when multiple replicates were present in the original data. (**G**) Expression levels of the candidate genes at generations 0, 25, and 50 of the experimentally evolved *D. simulans* populations measured by qPCR (***Figure 4—source data 1***). Significance was assessed using the paired t-test correcting for multiple testing and represented using letter codes. Panels E and F were created with BioRender.com.

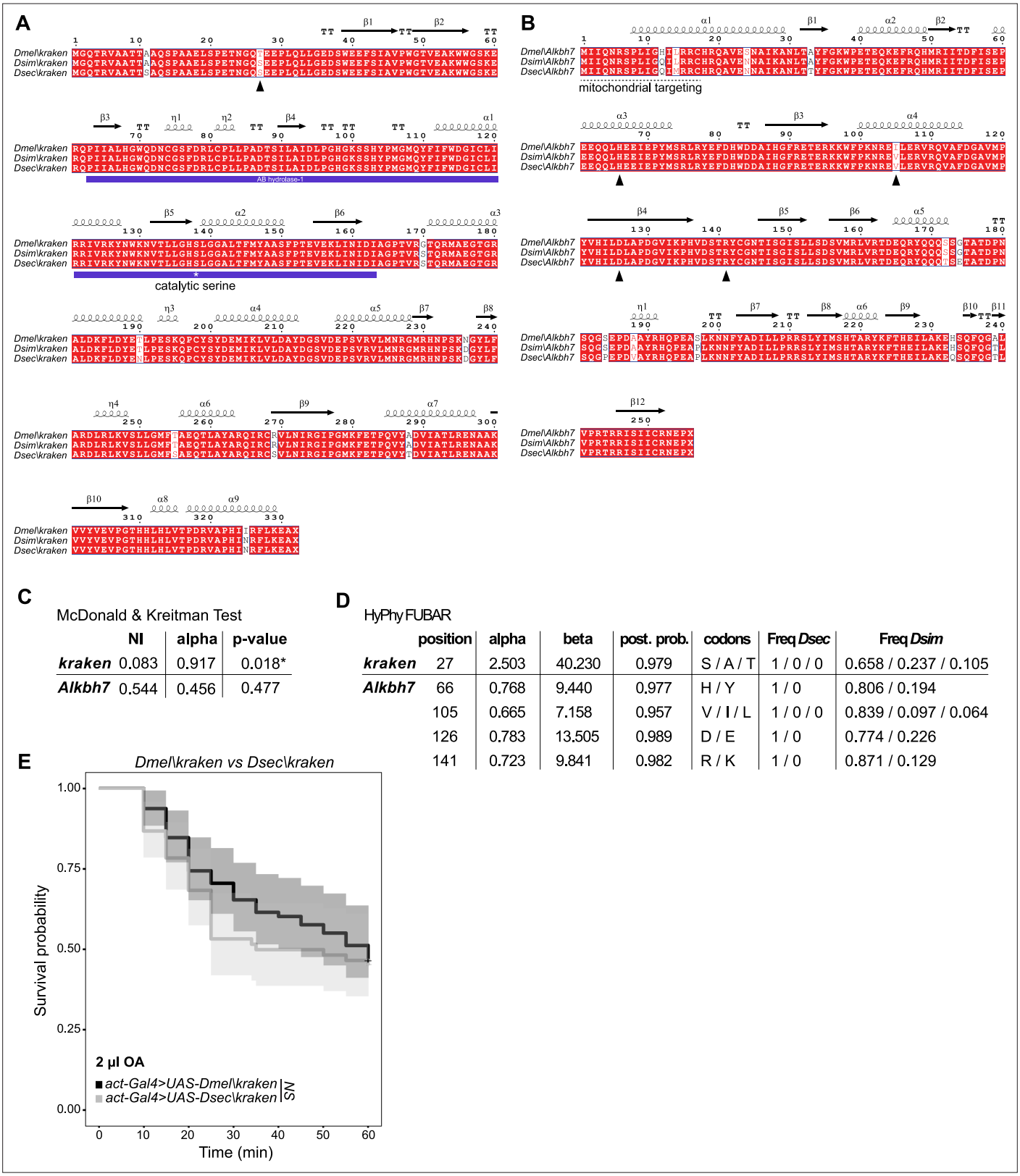

Figure 4—figure supplement 1. Sequence and molecular evolutionary analyses of *kraken* and *Alkbh7*. (**A,B**) Protein sequence alignment and secondary structure of Kraken (**A**) and Alkbh7 (**B**) from the reference sequence of *D. melanogaster* (NP_477265.1 and NP_648511.2) and the consensus sequences 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 (Fast, Unconstrained Bayesian AppRoximation) analysis (see **D**). (**C**) Results from the

Figure 4—figure supplement 1 continued on next page

Figure 4—figure supplement 1 continued

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 ten 2- to 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-effects Cox regression model. NS p>0.05. Raw data in ***Figure 4—figure supplement 1—source data 1***.

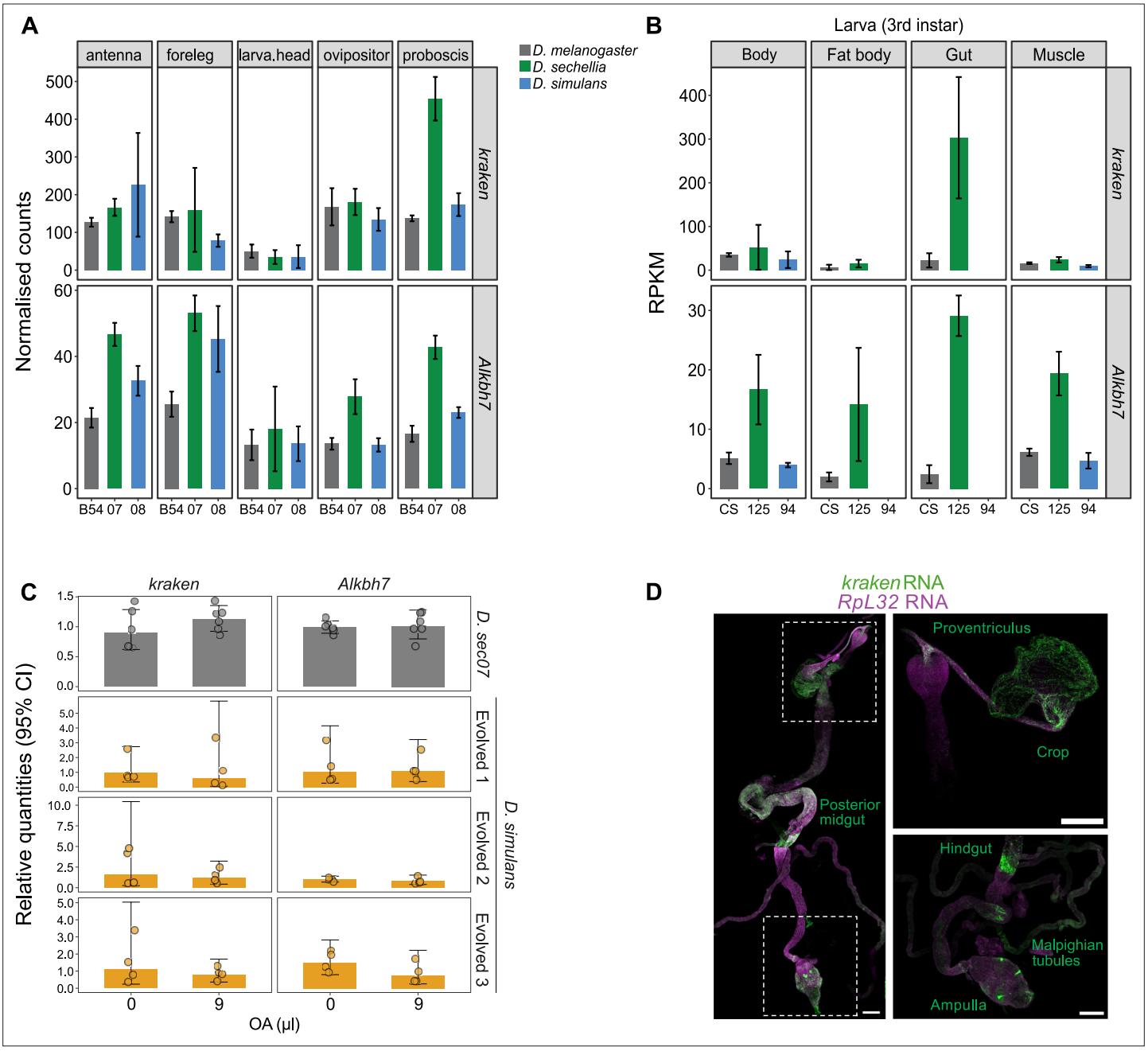

Figure 4—figure supplement 2. Additional expression analyses of *kraken* and *Alkbh7*. (**A**) Bar plots showing the expression levels of *kraken* and *Alkbh7* across various adult *Drosophila* tissues and species. Data from ***Bontonou et al., 2024***. (**B**) Bar plots showing the expression levels of *kraken* and *Alkbh7* across various larval *Drosophila* tissues and species. Data from ***Watanabe et al., 2019*** ('medium diet' condition). (**C**) Bar plots showing the effect of exposure to 9 µl octanoic acid (OA) (in the tube assay) on *kraken* and *Alkbh7* expression in *D. sechellia* and three evolved (**G50**) *D. simulans* populations. (**D**) Left: RNAscope detection of *kraken* (green) and *Rpl32* (magenta) transcripts in the gut and Malpighian tubules. Right: higher-magnification showing *kraken* transcript expression in the indicated tissues. Scale bars, 200 µm. This expression pattern was observed in tissues from >10 individuals.

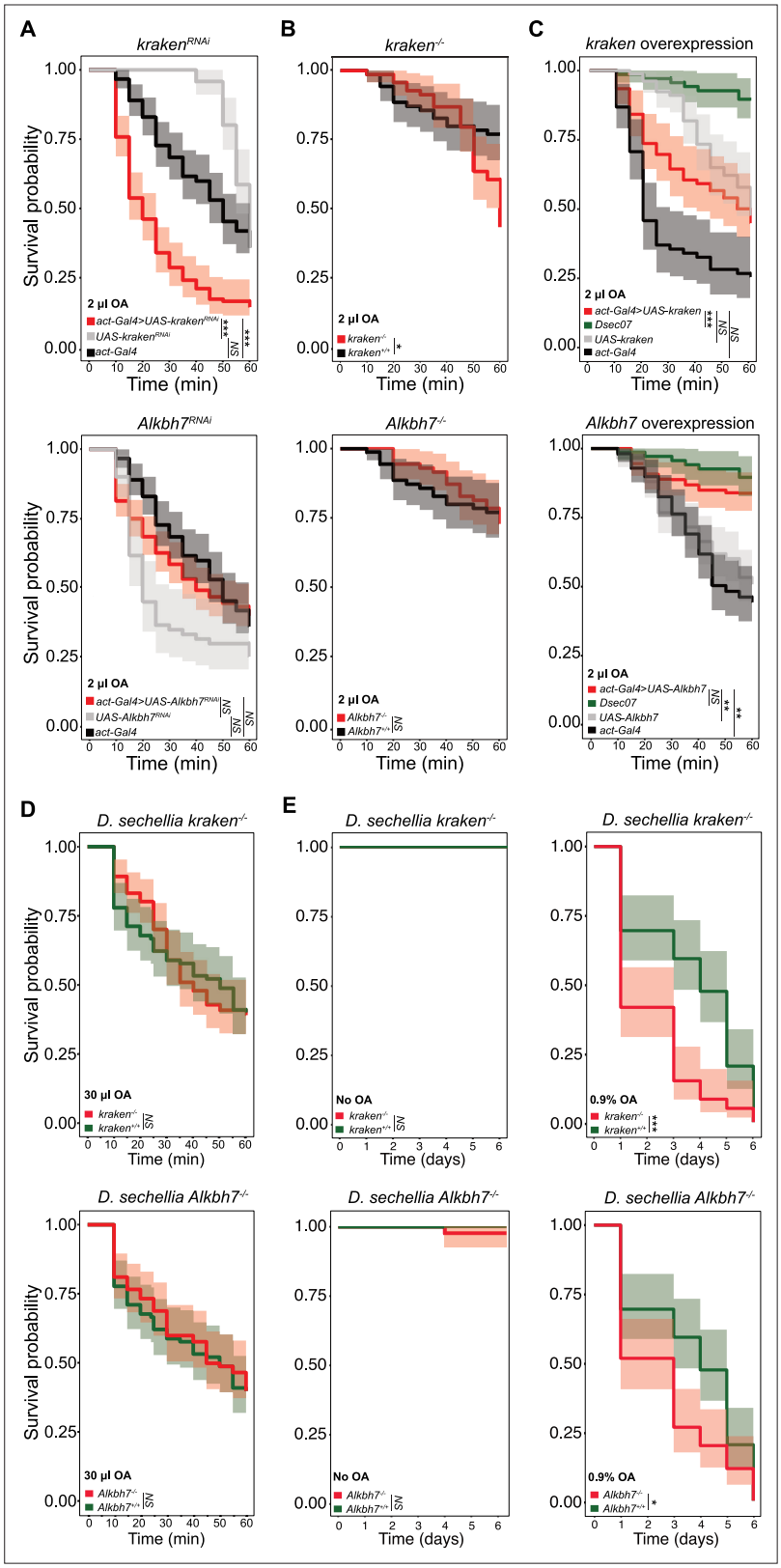

Figure 5. Functional validation of the contribution of *kraken* and *Alkbh7* to octanoic acid (OA) resistance. (**A–C**) Survival curves of adult flies of the indicated genotypes tested in the plate assay with 2 µl OA for *kraken* and *Alkbh7* RNA interference (RNAi) (**A**), mutants (**B**), and overexpression (**C**). Each replicate consisted of ten 2- to 7-day-old female flies. N replicates ≥6, n flies ≥60. Shading indicates 95% confidence intervals. Genotypes: **A** *act-*

Figure 5 continued on next page

Figure 5 continued

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-effects 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 ***Figure 5—source data 1***.

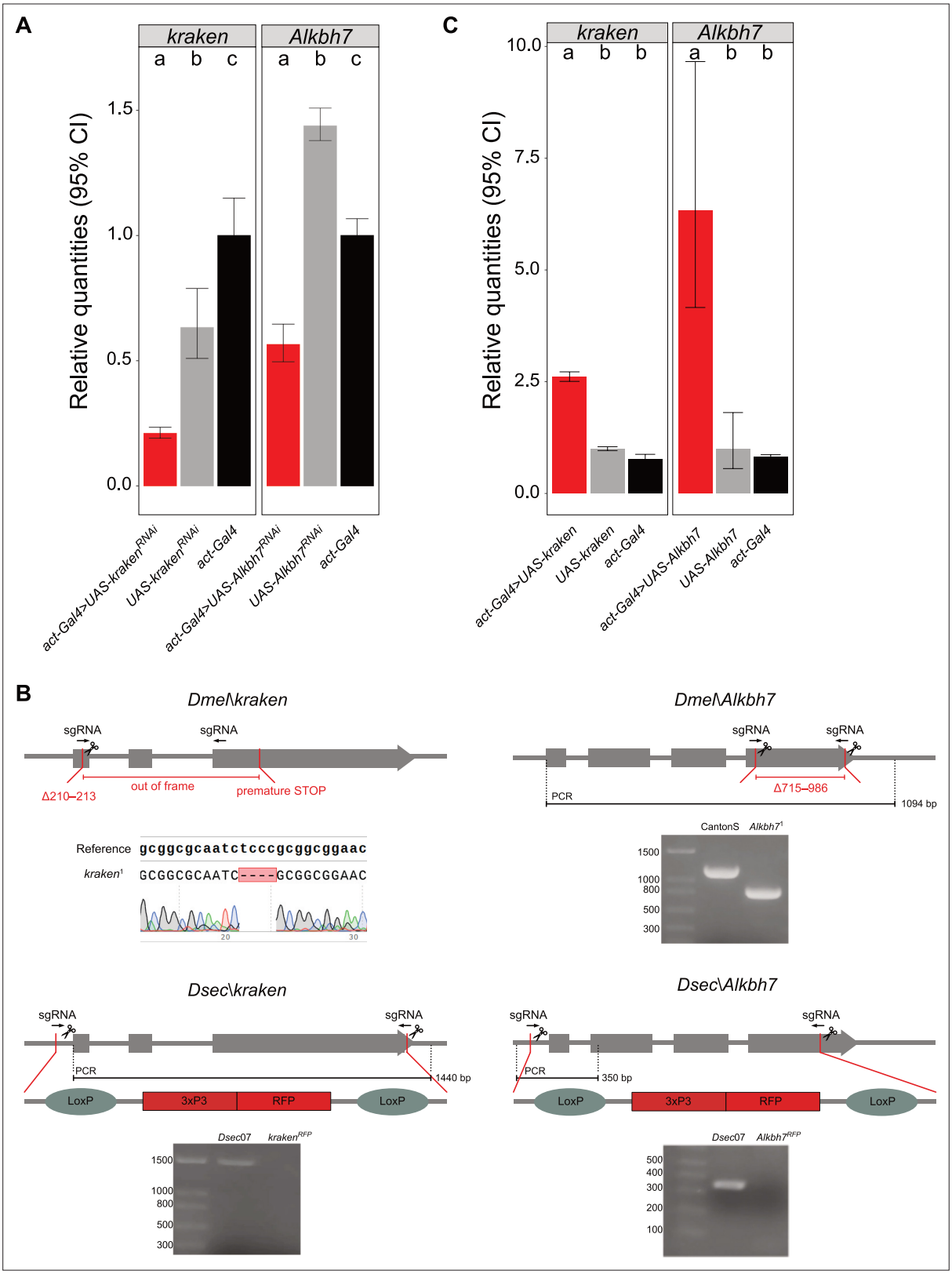

Figure 5—figure supplement 1. Validation of RNA interference (RNAi) and overexpression lines and generation of mutants for *kraken* and *Alkbh7*. (**A**) Quantitative RT-PCR analysis of RNAi-mediated knockdown of *kraken* and *Alkbh7* in *D. melanogaster*. Genotypes as in ***Figure 5A***. (**B**) Schematics of the CRISPR/Cas9 design to generate *D. melanogaster* and *D. sechellia* mutants for *kraken* and *Alkbh7*. Validation was performed by PCR (***Figure 5—figure supplement 1—source data 1***) and Sanger sequencing. The *Dsec\kraken*^{RFP} mutant comprised a 1431 bp deletion (–44 to +1387 nt relative to

Figure 5—figure supplement 1 continued on next page

Figure 5—figure supplement 1 continued

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 ***Figure 5C***.

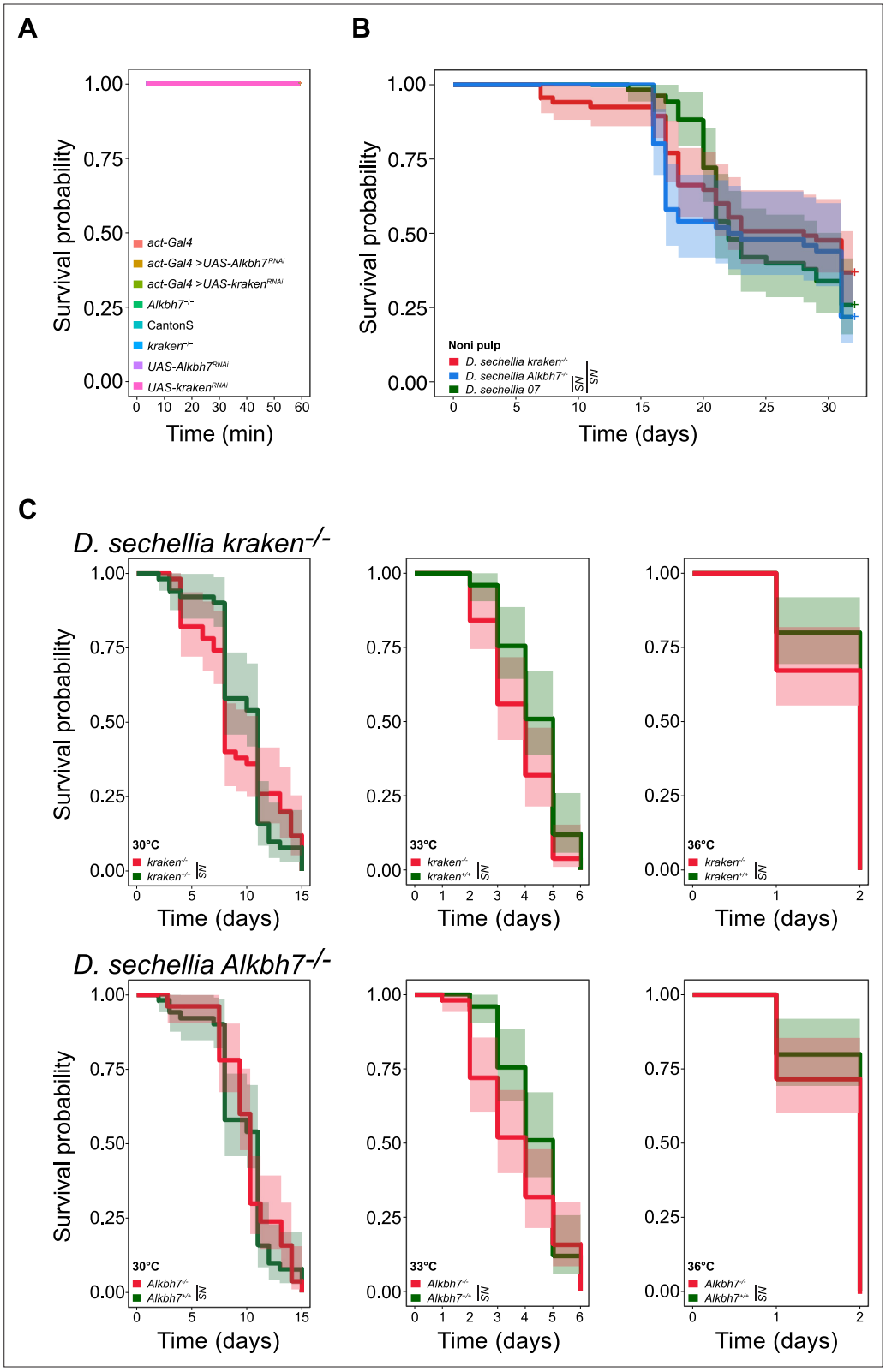

Figure 5—figure supplement 2. Control survival curves in the absence of octanoic acid (OA). (**A**) Survival curves of adult flies of the indicated genotypes tested in the plate assay without OA. Each replicate consisted of ten 2- to 7-day-old female flies. N replicates ≥5, n flies ≥50. All genotypes display full viability, rendering data points invisible beneath the topmost displayed genotype (*UAS-kraken*^{RNAi}). Genotypes as in ***Figure 5A and B***. (**B**) Survival

Figure 5—figure supplement 2 continued on next page

Figure 5—figure supplement 2 continued

curves of wild-type (*Dsec07*), and *kraken* (*Dsec\kraken*^{RFP}) and *Alkbh7* (*Dsec\Alkbh7*^{RFP}) mutant *D. sechellia* on homogenized noni pulp. N=50 per strain. Significance is based on the resulting mixed-effects Cox regression model. NS p>0.05. (**C**) Survival curves of *kraken* (*Dsec\kraken*^{RFP}, top) and *Alkbh7* (*Dsec\Alkbh7*^{RFP}, bottom) mutant *D. sechellia* compared to wild-type (*Dsec07*) flies at 30°C, 33°C, 36°C. N≥50 per strain/condition. Statistics as in **B**.