Diet unmasks genetic variants that regulate lifespan in outbred Drosophila

Authors: Luisa F. Pallares[1,2]†, Amanda J. Lea[1,2]†, Clair Han[3], Elena V. Filippova[1], Peter Andolfatto[4]‡, Julien F. Ayroles[1,2]‡*

Affiliations:

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton NJ, USA.

[2]Ecology and Evolutionary Biology Department, Princeton University, Princeton NJ, USA.

[3]Janelia Research Campus of the Howard Hughes Medical Institute, Ashburn VA, USA

[4]Department of Biological Sciences, Columbia University, New York, NY, USA.

†These authors contributed equally.

‡These authors contributed equally.

*Correspondence to: *jayroles@princeton.edu.*

Abstract:

Several evolutionary forces are thought to maintain genetic variation for fitness-related traits, such as lifespan, but experimental support is limited. Using a powerful experimental design, we identified lifespan-associated variants by exposing outbred *Drosophila melanogaster* to standard and high-sugar diets and tracking genome-wide allele frequency changes as the flies aged. We mapped alleles associated with early vs late life tradeoffs, late-onset effects, and genotype-by-environment (GxE) interactions – all of which are predicted by long-standing theories to maintain genetic variation for lifespan. We also validated an environmentally-dependent role for *nAChRα4* in regulating lifespan; the ortholog of this gene is one of the few lifespan-associated genes in humans (CHRNA3). Our results provide insight into the highly polygenic and context-dependent genetic architecture of lifespan, as well as the evolutionary processes that shape this key trait.

Main text:

Lifespan, a major component of fitness and a key life history trait, has a genetic basis: it is modestly heritable in humans and other organisms (h^2~10%) (*1*) and dozens of lifespan-reducing alleles have now been identified (*2*, *3*). However, the fact that genetic variation for lifespan exists at all presents an evolutionary puzzle, as it is expected that natural selection will purge fitness-reducing alleles from the gene pool. Evolutionary theory provides several potential, non-mutually exclusive explanations for this conundrum. Lifespan-reducing alleles may persist because: (i) they are only deleterious in late-life, when selection is relatively weak (the mutation accumulation theory (*4*)), (ii) they provide benefits early in life that outweigh their late-life costs (the antagonistic pleiotropy theory (*5*)), and (iii) their effects vary across environments (genotype-by-environment, GxE) making them difficult to purge through purifying selection.

Notably, a special class of GxE interactions, driven by evolutionarily recent changes in human diet and lifestyle (*6*–*8*), are thought to be particularly important for human disease. Specifically, it has been proposed that complex diseases are caused by alleles that evolved under stabilizing or positive selection throughout human history, but are now "mismatched" to obesogenic diets and other aspects of modern life (*6*–*9*). While this explanation is compelling, empirical data is limited due to the difficulty of identifying GxE interactions at genome-wide scale with high power (but see (*10*)). As a result, the degree to which exposure to evolutionarily novel environments alters the relationship between genetic variation and fitness-related traits remains unclear.

To address this question, we leveraged the tractable experimental and genomic tools of *Drosophila melanogaster* to map loci that affect lifespan in two environments. Specifically, we exposed an outbred population of flies to two diets: a standard laboratory diet and a high sugar diet containing more sucrose than flies would encounter in nature and that is known to cause obesity, diabetes, and reduced lifespan in this species (*11*, *12*). Drawing inspiration from a recent study of human longevity (*13*), we tracked genome-wide allele frequency changes in age-matched flies across their entire adult life. Using this high-powered experimental approach (Figure 1A-B), we were able to identify thousands of lifespan-reducing alleles that decrease in frequency as individuals grow older, as well as to classify them into: (i) GxE alleles that have stronger effects on lifespan on one diet, potentially due to risk alleles being exposed by the novel high sugar diet, (ii) alleles with tradeoffs between early and late life that first increase and then decrease (antagonistic pleiotropy theory), and (iii) late-onset alleles that only decrease at late ages (mutation accumulation theory). Together, our study provides insight into the genetic architecture and environmental sensitivity of a major life history trait, and experimentally addresses long-standing theories for why fitness-reducing alleles abound in nature.

Sex and genotype have environment-specific effects on lifespan

To identify loci associated with lifespan variation and evaluate their context-dependence, we exposed large, replicate populations of age-matched outbred flies to standard laboratory (from now on "control" or "CTRL") and high sugar ("HS") diets for one generation (n=3 replicates of ~10,000 flies per diet; Figure 1A). To prevent overlapping generations, the food containers where flies also lay eggs were exchanged every three days. We drew a random sample of ~2000 flies at the beginning of the experiment (T_0), and continued to sample ~500 flies from each population at regular intervals. When only the ~500 longest-lived flies were left in a given replicate cage, we collected a final sample (T_N) (Table S1,2). In total, 10,637 flies were genotyped using individually barcoded low-coverage genome sequencing (Fig. S1) to estimate age-specific genome-wide allele frequencies on each diet.

While all replicates for the two diets started from a common pool of standing genetic variation at the beginning of the experiment (Figure 1C; Table S3), we observed a consistent ~1.6 fold reduction in lifespan for flies on the HS diet, as expected (*12*). We also observed substantial and unexpected interactions between diet and sex: while the sex ratio remained roughly 1:1 as flies aged on the HS diet, males far outlived females on the CTRL diet resulting in a sex ratio of ~100:1 by the end of the experiment (Figure 1D; Table S4). We replicated this observation in independent experiments where time-to-death of individual flies was quantified, suggesting that it is a repeatable characteristic of the fly population used here (Cox proportional hazards: p(sex-by-diet) = 0.026) (Fig. S2, Table S5). While others have also observed that sex-specific lifespans in flies are sometimes environmentally-dependent (*14*, *15*), future work is necessary to uncover the proximate mechanisms at play here.

We estimated allele frequencies at 291,319 common SNPs (MAF>0.05) and tested for alleles that exhibited a significantly lower frequency at the end of the experiment (T_N; n=1443 and 1866

sequenced flies for CTRL and HS, respectively) compared to the beginning of the experiment (T_0; n=2104 sequenced flies). Such decreases in frequency indicate that individuals carrying a focal allele die at younger ages relative to individuals carrying the alternative allele (Figure 1B). We identified 4294 genetic variants that fit this pattern distributed among 2425 genes (FDR<1%, beta-binomial model; Figure 2, Table S6, Fig. S3). The average absolute decrease in allele frequency (between T_N and T_0) for these lifespan-associated SNPs was 0.08, with most changes falling between 0.05-0.11 (Fig. S4). The majority of these lifespan-associated variants (2812; 66%) had similar or "shared" effect sizes between environments (defined as FDR<1% in one environment and p<0.05 in the other). However, many lifespan-associated SNPs (1482; 34%) exhibited evidence for GxE interactions, and had stronger effects on lifespan on one diet relative to the other ("GxE" defined as FDR<1% in one environment and p>0.05 in the other environment, which our simulations suggest picks up a conservative set of SNPs with environmentally-dependent effect sizes; Fig. S5). Strikingly, out of the 1482 SNPs with GxE effects on lifespan, 91% had larger effects on the HS diet, indicating that their effects are magnified under dietary stress (Figure 2). These results suggest that a substantial amount of genetic variation that appears to have little effect on phenotypic variation under one set of conditions might indeed play a fundamental role in new or stressful environments.

Given the striking sex-by-diet interaction on overall mortality patterns (Figure 1D), we also evaluated the degree to which the genetic basis of lifespan differs between the sexes using two approaches. First, we tested for allele frequency differences between males and females at T_0. We do not expect any sex differences at T_0 unless there are loci with sex-specific effects on viability during development (e.g., loci that have genetic incompatibilities with the Y chromosome or that interact with X chromosome dosage (*16*). We found 663 SNPs with viability effects (beta-binomial model, FDR<1%; Figure 3A, Table S7). Two notable examples fall near *nAChRα4,* the ortholog of *CHRNA3* which has sex-specific effects on longevity in humans (*3*, *13*). Second, we

tested for alleles with time (T_N vs T_0) by sex interactions which would point to SNPs with sex-specific effects on lifespan. For this analysis we used data from the HS replicates only; the absence of females at T_N in the CTRL replicates prevented their use for this purpose (Figure 1D). Only 29 SNPs showed significant sex-dependent effects on lifespan (beta-binomial model, FDR<1%; Table S8). Because we found only modest evidence that sex modifies the genetic architecture of lifespan in adult flies, we combined both sexes for most subsequent analyses.

Biological and functional insight into the genetic basis of lifespan

To understand the biology of loci that contribute to lifespan similarly on both diets (n=2812 "shared" SNPs) or more so on one diet (n=136 and 1346 SNPs with stronger effects on the CTRL and HS diets, respectively) we first asked whether shared or GxE lifespan-associated SNPs were enriched in particular genomic features. We found that both SNP groups are enriched in protein-coding regions (Fisher's exact test: odds ratio=1.15 and 1.12, $p=2.7x10^{-4}$ and p=0.039, for shared and GxE SNPs, respectively; Table S9). Nevertheless, using RNA-seq data generated from an independent replicate of the main experimental set up (n=162, Figure 4A Table S10, Fig. S6), we found that lifespan-associated SNPs were 1.4x more likely to fall near genes differentially expressed between old (T_N) and young (T_0) flies (relative to SNPs not associated with lifespan; Fisher's exact test, $p=4.4x10^{-11}$). This suggests that while lifespan-altering genetic variation is slightly biased toward protein-coding regions, it likely also plays an important regulatory role for genes relevant to aging.

Second, we asked whether genes harboring shared or GxE lifespan-associated SNPs were enriched for any particular molecular processes. We found that while both gene sets were strongly enriched for genes identified in previous studies of *D. melanogaster* longevity (Figure 4B, Table S12), they were not significantly enriched for any particular molecular pathway nor for "canonical" longevity genes (FDR<1%; Table S12). Restricting our analyses to genes that both contain

144 longevity alleles and are differentially expressed with age did uncover some enrichment,
145 particularly for oogenesis and neuronal related processes (FDR<1%; Table S13). However, these
146 results more generally support a highly polygenic model in which genetic variation segregating in
147 wild-derived populations of *D. melanogaster* does not localize to the canonical biological
148 pathways associated with aging and lifespan (*17*, *18*), as was also observed by (*14*).
149
150 Many of the lifespan-associated genes we identified perform essential functions. For example,
151 *midway*, involved in fat metabolism and oogenesis (*19*), and *lovit*, involved in neurophysiology
152 (*20*), both contain lifespan-associated SNPs. Using loss-of-function mutant lines, we validated
153 their effects on lifespan (*lovit: p-value* = 0.042; *midway:* p-value < 10^{-16}), and show that *midway*
154 acts in an environment-dependent manner (p-value = $7x10^{-4}$, Figure 4C-D, Table S14). We also
155 used loss-of-function mutant lines to investigate *nAChRα4* – the ortholog of a key human longevity
156 gene, *CHRNA3* – for which our main experiment revealed complex sex-by-diet-by-genotype
157 interactions (Fig 3C-D). Our follow up experiments confirmed this three-way interaction: wild type
158 males maintained a large survival advantage over *nAChRα4* mutants on both diets, while in
159 females this effect was modest on a CTRL diet and completely reversed on a HS diet (Fig. 3C-D,
160 Table S14). These results are in agreement with human studies where variants near *CHRNA3*
161 have stronger effects on lifespan in males (*3*, *13*); however, they also suggest that dampened
162 *CHRNA3*-lifespan associations in human females may be driven by environmental interactions
163 that obscure main effects of genotype.
164
165 **Testing evolutionary theories of aging and longevity**
166 Fitness-reducing alleles are thought to largely be governed by mutation-selection balance, in
167 which mutation continuously generates deleterious alleles and purifying selection eliminates them
168 (*21*). In support of this idea (*22*), we find that the minor allele reduces lifespan 98% of the time.
169 That said, these risk alleles are by no means rare in the population (mean frequency +/- SD at T_0

= 0.363 +/- 0.121; Table S6), suggesting that other evolutionary forces maintain them at moderate frequencies. Our results establish GxE interactions as one such key factor. We next asked if two additional forces, antagonistic pleiotropy and mutation accumulation, may also be important contributors to this feature of the data. Specifically, we estimated allele frequencies at several time points between T_0 and T_N to determine the trajectory of lifespan-reducing alleles (Figure 1A, Table S1,2). We then asked whether these alleles exhibited (i) a U-shaped pattern indicative of trade-offs and differential fitness effects at young versus old ages, as predicted by antagonistic pleiotropy theory (*5*); (ii) evidence for fitness-effects only at old ages, as predicted by mutation accumulation theory (*4*); or (iii) an evolutionary "null" model of constant fitness-effects at all ages (Figure 5A).

Of the 4294 SNPs with shared or GxE effects on lifespan, we confidently assigned 61 to one of the three trajectories described above ($_{\Delta}$AIC between the best and second-best trajectory > 99% of permutations). 49% of these 61 SNPs followed an antagonistic pleiotropy pattern, and 46% a pattern consistent with mutation accumulation theory. Importantly, only 3 SNPs exhibited constant changes in frequency with age, a pattern that does not correspond to any of the evolutionary models we have considered (Figure 5, Table S15). In further support of antagonistic pleiotropy theory, we also find that genes near any shared or GxE lifespan-associated SNPs (not just the 61 with assigned trajectories) significantly overlap with a previous study of age-specific fertility in flies (Figure 4B; Table S12; (*23*)). This overlap further indicates that many longevity-reducing alleles are maintained because they provide other benefits, for example to fertility in early adulthood, that outweigh their late life costs.

The evolution of alleles with GxE effects on lifespan

We have shown evidence that lifespan-reducing alleles are maintained at intermediate frequencies in outbred flies through several evolutionary forces, namely GxE interactions,

mutation accumulation, and antagonistic pleiotropy. Strikingly, we find that exposing flies to unusually high sugar concentrations reveals many previously hidden/cryptic fitness-associated variants, as has been predicted repeatedly (*6*) but rarely tested experimentally. This result has implications for human health, as it is thought that rapid shifts in human diet and lifestyle following the Industrial Revolution have caused previously adaptive or neutral alleles to become "mismatched" or "maladaptive", such that they are currently associated with diseases that impact lifespan (Figure 5E; (*6*, *7*, *10*)).

Both the standard lab (CTRL) and the HS diets represent substantial dietary changes for our experimental flies, compared to conditions "in the wild". Thus, it is likely that alleles under selection to maximize longevity in wild conditions are not optimally matched to either of the diets in our experiment. This is especially likely of the HS diet for which we observed a substantially shortened lifespan. If the variants contributing to lifespan differences on the two diets in our experiment were previously neutral, we would expect no bias with regard to whether variants causing lifespan differences are associated with the ancestral or derived allele. Contrary to this expectation, we find that, for both shared and GxE SNPs, lifespan-reducing alleles are more likely to be derived. This pattern is most dramatic for HS-specific risk alleles: 67% of lifespan-reducing alleles are derived compared to background expectations of 52% (estimated using 1000 randomly drawn pools of non-significant SNPs) (Figure 5F). Thus, it is clear that lifespan-reducing alleles are unlikely to be neutral, and their high frequencies in the source population also make it unlikely that they reflect mutation-selection balance. We speculate that directional selection in response to spatial and temporal environmental heterogeneity has resulted in high frequency-derived alleles that contribute to reduced longevity in new environments (*24*).

Implications for understanding the genetic basis of lifespan variation

Long-standing population genetic and evolutionary theories have proposed several forces at play in the maintenance of genetic variation for fitness-related traits (*4*–*6*, *8*). However, the predictions of these theories have rarely been experimentally tested, due to the difficulty of mapping fitness-related genetic effects. For example, a recent human study using a similar approach to ours but a 10-fold larger sample size found only two lifespan-associated loci near the *APOE* and *CHRNA3* genes (*13*). Instead, we identified thousands of lifespan-associated loci, many of which have larger effects on the HS diet, uncovering a highly polygenic and context-dependent architecture. We estimate that in the absence of environmental heterogeneity, both studies have similar statistical power (Fig. S7); the fact that we find orders of magnitude more lifespan-associated SNPS here highlights the utility of well-controlled experimental designs in model organisms for the study of complex traits. Because our high-powered design allowed us to identify many lifespan-reducing alleles, we could evaluate the generality of long-standing theories for why alleles that shorten lifespan persist in nature and how they evolve. Specifically, we identified GxE interactions, as well as mutation accumulation and antagonistic pleiotropy, as factors maintaining genetic variation for lifespan. We also provide experimental insight into how interactions between derived genetic variation and novel environmental conditions may shorten lifespan.

References and Notes:

1. J. G. Ruby, K. M. Wright, K. A. Rand, A. Kermany, K. Noto, D. Curtis, N. Varner, D. Garrigan, D. Slinkov, I. Dorfman, J. M. Granka, J. Byrnes, N. Myres, C. Ball, Estimates of the Heritability of Human Longevity Are Substantially Inflated due to Assortative Mating. *Genetics*. **210**, 1109–1124 (2018).

2. M. Kuningas, S. P. Mooijaart, D. V. Heemst, B. J. Zwaan, P. E. Slagboom, R. G. J. Westendorp, Genes encoding longevity: from model organisms to humans. *Aging Cell*. **7**, 270–280 (2008).

3. P. K. Joshi, K. Fischer, K. E. Schraut, H. Campbell, T. Esko, J. F. Wilson, Variants near CHRNA3/5 and APOE have age- and sex-related effects on human lifespan. *Nature Communications*. **7**, 11174 (2016).

4. P. B. Medawar, *An unsolved problem of biology* (Published for the College by H.K. Lewis, London, 1952).

5. G. C. Williams, Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution*. **11**, 398–411 (1957).

6. G. Gibson, Decanalization and the origin of complex disease. *Nature Reviews Genetics*. **10**, 134–140 (2009).

7. S. Corbett, A. Courtiol, V. Lummaa, J. Moorad, S. Stearns, The transition to modernity and chronic disease: mismatch and natural selection. *Nature Reviews Genetics*. **19**, 419–430 (2018).

8. G. Gibson, I. Dworkin, Uncovering cryptic genetic variation. *Nature Reviews Genetics*. **5**, 681–690 (2004).

9. A. D. Rienzo, R. R. Hudson, An evolutionary framework for common diseases: the ancestral-susceptibility model. *Trends in Genetics*. **21**, 596–601 (2005).

10. L. Quintana-Murci, Understanding rare and common diseases in the context of human evolution. *Genome Biology*. **17**, 225 (2016).

11. L. P. Musselman, J. L. Fink, K. Narzinski, P. V. Ramachandran, S. S. Hathiramani, R. L. Cagan, T. J. Baranski, A high-sugar diet produces obesity and insulin resistance in wild-type Drosophila. *Disease Models & Mechanisms*. **4**, 842–849 (2011).

12. J. Na, L. P. Musselman, J. Pendse, T. J. Baranski, R. Bodmer, K. Ocorr, R. Cagan, A Drosophila Model of High Sugar Diet-Induced Cardiomyopathy. *PLOS Genetics*. **9**, e1003175 (2013).

13. H. Mostafavi, T. Berisa, F. R. Day, J. R. B. Perry, M. Przeworski, J. K. Pickrell, Identifying genetic variants that affect viability in large cohorts. *PLOS Biology*. **15**, e2002458 (2017).

14. W. Huang, T. Campbell, M. A. Carbone, W. E. Jones, D. Unselt, R. R. H. Anholt, T. F. C. Mackay, Context-dependent genetic architecture of Drosophila life span. *PLoS Biol.* **18**, e3000645 (2020).

15. J. M. S. Burger, D. E. L. Promislow, Sex-Specific Effects of Interventions That Extend Fly Life Span. *Sci. Aging Knowl. Environ.* **2004**, pe30 (2004).

16. D. A. Barbash, Genetic Testing of the Hypothesis That Hybrid Male Lethality Results From a Failure in Dosage Compensation. *Genetics*. **184**, 313–316 (2010).

17. M. D. W. Piper, L. Partridge, Drosophila as a model for ageing. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. **1864**, 2707–2717 (2018).

18. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, The Hallmarks of Aging. *Cell*. **153**, 1194–1217 (2013).

19. M. Buszczak, X. Lu, W. A. Segraves, T. Y. Chang, L. Cooley, Mutations in the midway gene disrupt a Drosophila acyl coenzyme A: diacylglycerol acyltransferase. *Genetics*. **160**, 1511–1518 (2002).

20. Y. Xu, T. Wang, LOVIT Is a Putative Vesicular Histamine Transporter Required in Drosophila for Vision. *Cell Reports*. **27**, 1327-1333.e3 (2019).

21. J. F. Crow, M. Kimura, *An Introduction to Population Genetics Theory* (Blackburn Press, 2009).

22. T. Kido, W. Sikora-Wohlfeld, M. Kawashima, S. Kikuchi, N. Kamatani, A. Patwardhan, R. Chen, M. Sirota, K. Kodama, D. Hadley, A. J. Butte, Are minor alleles more likely to be risk alleles? *BMC Medical Genomics*. **11**, 3 (2018).

23. M. F. Durham, M. M. Magwire, E. A. Stone, J. Leips, Genome-wide analysis in Drosophila reveals age-specific effects of SNPs on fitness traits. *Nature Communications*. **5**, 4338 (2014).

24. E. Sezgin, D. D. Duvernell, L. M. Matzkin, Y. Duan, C.-T. Zhu, B. C. Verrelli, W. F. Eanes, Single-Locus Latitudinal Clines and Their Relationship to Temperate Adaptation in Metabolic Genes and Derived Alleles in Drosophila melanogaster. *Genetics*. **168**, 923–931 (2004).

25. M. U. Carnes, T. Campbell, W. Huang, D. G. Butler, M. A. Carbone, L. H. Duncan, S. V. Harbajan, E. M. King, K. R. Peterson, A. Weitzel, S. Zhou, T. F. C. Mackay, The Genomic Basis of Postponed Senescence in Drosophila melanogaster. *PLOS ONE*. **10**, e0138569 (2015).

26. D. K. Fabian, K. Garschall, P. Klepsatel, G. Santos-Matos, É. Sucena, M. Kapun, B. Lemaitre, C. Schlötterer, R. Arking, T. Flatt, Evolution of longevity improves immunity in Drosophila. *Evolution Letters*. **2**, 567–579 (2018).

27. K. M. Hoedjes, J. van den Heuvel, M. Kapun, L. Keller, T. Flatt, B. J. Zwaan, Distinct genomic signals of lifespan and life history evolution in response to postponed reproduction and larval diet in Drosophila. *Evol Lett*. **3**, 598–609 (2019).

28. S. C. Remolina, P. L. Chang, J. Leips, S. V. Nuzhdin, K. A. Hughes, Genomic basis of aging and life-history evolution in Drosophila melanogaster. *Evolution*. **66**, 3390–3403 (2012).

29. R. Tacutu, D. Thornton, E. Johnson, A. Budovsky, D. Barardo, T. Craig, E. Diana, G. Lehmann, D. Toren, J. Wang, V. E. Fraifeld, J. P. de Magalhães, Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res.* **46**, D1083–D1090 (2018).

30. T. M. Therneau, *coxme: Mixed Effects Cox Models* (2020; https://CRAN.R-project.org/package=coxme).

31. T. M. Therneau, P. M. Grambsch, *Modeling Survival Data: Extending the Cox Model* (Springer-Verlag, New York, 2000; https://www.springer.com/gp/book/9780387987842), *Statistics for Biology and Health*.

32. T. M. Therneau, T. L. (original S.->R port and R. maintainer until 2009), A. Elizabeth, C. Cynthia, *survival: Survival Analysis* (2020; https://CRAN.R-project.org/package=survival).

33. S. Picelli, A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

34. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. **17**, 10–12 (2011).

35. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).

36. G. A. Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*, 11–10.

37. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

38. L. F. Pallares, S. Picard, J. F. Ayroles, TM3'seq: a tagmentation-mediated 3'sequencing approach for improving scalability of RNAseq experiments. *G3: Genes, Genomes, Genetics*. **10**, 143–150 (2020).

39. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

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

41. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. **30**, 923–930 (2014).

42. B. S. Weir, C. C. Cockerham, Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. **38**, 1358–1370 (1984).

43. R. M. Durbin, D. Altshuler, R. M. Durbin, G. R. Abecasis, D. R. Bentley, A. Chakravarti, A. G. Clark, F. S. Collins, F. M. De La Vega, P. Donnelly, M. Egholm, P. Flicek, S. B. Gabriel, R. A. Gibbs, B. M. Knoppers, E. S. Lander, H. Lehrach, E. R. Mardis, G. A. McVean, D. A. Nickerson, L. Peltonen, A. J. Schafer, S. T. Sherry, J. Wang, R. K. Wilson, R. A. Gibbs, D. Deiros, M. Metzker, D. Muzny, J. Reid, D. Wheeler, J. Wang, J. Li, M. Jian, G. Li, R. Li, H. Liang, G. Tian, B. Wang, J. Wang, W. Wang, H. Yang, X. Zhang, H. Zheng, E. S. Lander, D. Altshuler, L. Ambrogio, T. Bloom, K. Cibulskis, T. J. Fennell, S. B. Gabriel, D. B. Jaffe, E. Shefler, C. L. Sougnez, D. R. Bentley, N. Gormley, S. Humphray, Z. Kingsbury, P. Kokko-Gonzales, J. Stone, K. J. McKernan, G. L. Costa, J. K. Ichikawa, C. C. Lee, R. Sudbrak, H. Lehrach, T. A. Borodina, A. Dahl, A. N. Davydov, P. Marquardt, F. Mertes, W. Nietfeld, P. Rosenstiel, S. Schreiber, A. V. Soldatov, B. Timmermann, M. Tolzmann, M. Egholm, J. Affourtit, D. Ashworth, S. Attiya, M. Bachorski, E. Buglione, A. Burke, A. Caprio, C. Celone, S. Clark, D. Conners, B. Desany, L. Gu, L. Guccione, K. Kao, A. Kebbel, J. Knowlton, M. Labrecque, L. McDade, C. Mealmaker, M. Minderman, A. Nawrocki, F. Niazi, K. Pareja, R. Ramenani, D. Riches, W. Song, C. Turcotte, S. Wang, E. R. Mardis, R. K. Wilson, D. Dooling, L. Fulton, R. Fulton, G. Weinstock, R. M. Durbin, J. Burton, D. M. Carter, C. Churcher, A. Coffey, A. Cox, A. Palotie, M. Quail, T. Skelly, J. Stalker, H. P. Swerdlow, D. Turner, A. De Witte, S. Giles, R. A. Gibbs, D. Wheeler, M. Bainbridge, D. Challis, A. Sabo, F. Yu, J. Yu, J. Wang, X. Fang, X. Guo, R. Li, Y. Li, R. Luo, S. Tai, H. Wu, H. Zheng, X. Zheng, Y. Zhou, G. Li, J. Wang, H. Yang, G. T. Marth, E. P. Garrison, W. Huang, A. Indap, D. Kural, W.-P. Lee, W. Fung Leong, A. R. Quinlan, C. Stewart, M. P. Stromberg, A. N. Ward, J. Wu, C. Lee, R. E. Mills, X. Shi, M. J. Daly, M. A. DePristo, D. Altshuler, A. D. Ball, E. Banks, T. Bloom, B. L. Browning, K. Cibulskis, T. J. Fennell, K. V. Garimella, S. R. Grossman, R. E. Handsaker, M. Hanna, C. Hartl, D. B. Jaffe, A. M. Kernytsky, J. M. Korn, H. Li, J. R. Maguire, S. A. McCarroll, A. McKenna, J. C. Nemesh, A. A. Philippakis, R. E. Poplin, A. Price, M. A. Rivas, P. C. Sabeti, S. F. Schaffner, E. Shefler, I. A. Shlyakhter, D. N. Cooper, E. V. Ball, M. Mort, A. D. Phillips, P. D. Stenson, J. Sebat, V. Makarov, K. Ye, S. C. Yoon, C. D. Bustamante, L. Clarke, P. Flicek, F. Cunningham, J. Herrero, S. Keenen, E. Kulesha, R. Leinonen, W. M. McLaren, R. Radhakrishnan, R. E. Smith, V. Zalunin, X. Zheng-Bradley, J. O. Korbel, A. M. Stütz, S. Humphray, M. Bauer, R. Keira Cheetham, T. Cox, M. Eberle, T. James, S. Kahn, L. Murray, A. Chakravarti, K. Ye, F. M. De La Vega, Y. Fu, F. C. L. Hyland, J. M. Manning, S. F. McLaughlin, H. E. Peckham, O. Sakarya, Y. A. Sun, E. F. Tsung, M. A. Batzer, M. K. Konkel, J. A. Walker, R. Sudbrak, M. W. Albrecht, V. S. Amstislavskiy, R. Herwig, D. V. Parkhomchuk, S. T. Sherry, R. Agarwala, H. M. Khouri, A. O. Morgulis, J. E. Paschall, L. D. Phan, K. E. Rotmistrovsky, R. D. Sanders, M. F. Shumway, C. Xiao, G. A. McVean, A. Auton, Z. Iqbal, G. Lunter, J. L. Marchini, L.

Moutsianas, S. Myers, A. Tumian, B. Desany, J. Knight, R. Winer, D. W. Craig, S. M. Beckstrom-Sternberg, A. Christoforides, The 1000 Genomes Project Consortium, Corresponding author, Steering committee, Production group: Baylor College of Medicine, BGI-Shenzhen, Broad Institute of MIT and Harvard, Illumina, Life Technologies, Max Planck Institute for Molecular Genetics, Roche Applied Science, Washington University in St Louis, Wellcome Trust Sanger Institute, Analysis group: Agilent Technologies, Baylor College of Medicine, Boston College, Brigham and Women's Hospital, T. H. G. M. D. Cardiff University, Cold Spring Harbor Laboratory, Cornell and Stanford Universities, European Bioinformatics Institute, European Molecular Biology Laboratory, Johns Hopkins University, Leiden University Medical Center, Louisiana State University, US National Institutes of Health, Oxford University, The Translational Genomics Research Institute, A map of human genome variation from population-scale sequencing. *Nature*. **467**, 1061–1073 (2010).

44. G. Bhatia, N. Patterson, S. Sankararaman, A. L. Price, Estimating and interpreting FST: The impact of rare variants. *Genome Res*. **23**, 1514–1521 (2013).

45. J. K. Pickrell, J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt, E. Nkadori, J.-B. Veyrieras, M. Stephens, Y. Gilad, J. K. Pritchard, Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*. **464**, 768–772 (2010).

46. A. Lea, J. Tung, X. Zhou, A flexible, efficient binomial mixed model for identifying differential DNA methylation in bisulfite sequencing data. *PLoS Genet.* **11**, e1005650 (2015).

47. J. Tung, X. Zhou, S. C. Alberts, M. Stephens, Y. Gilad, The genetic architecture of gene expression levels in wild baboons. *Elife*. **4**, 1–22 (2015).

48. C. A. Buerkle, Z. Gompert, Population genomics based on low coverage sequencing: how low should we go? *Molecular Ecology*. **22**, 3028–3035 (2013).

49. W. N. Venables, B. D. Ripley, Modern Applied Statistics with S. *Statistics and Computing* (2002)

50. M. Lesnoff, R. Lancelot, aod: Analysis of Overdispersed Data. R package version 1.3. 2012 (2016).

51. R. Kofler, R. V. Pandey, C. Schlötterer, PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*. **27**, 3435–3436 (2011).

52. G. dos Santos, A. J. Schroeder, J. L. Goodman, V. B. Strelets, M. A. Crosby, J. Thurmond, D. B. Emmert, W. M. Gelbart, FlyBase Consortium, FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.* **43**, D690-7 (2015).

53. M. Wang, Y. Zhao, B. Zhang, Efficient Test and Visualization of Multi-Set Intersections. *Scientific Reports*. **5**, 16923 (2015).

54. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

55. A. Zhu, J. G. Ibrahim, M. I. Love, Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*. **35**, 2084–2092 (2019).

56. R. Kofler, C. Schlötterer, Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*. **28**, 2084–2085 (2012).

57. G. F. Berriz, J. E. Beaver, C. Cenik, M. Tasan, F. P. Roth, Next generation software for functional trend analysis. *Bioinformatics*. **25**, 3043–3044 (2009).

58. H. Fang, B. Knezevic, K. L. Burnham, J. C. Knight, XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med*. **8**, 129 (2016).

59. J. Thurmond, J. L. Goodman, V. B. Strelets, H. Attrill, L. S. Gramates, S. J. Marygold, B. B. Matthews, G. Millburn, G. Antonazzo, V. Trovisco, T. C. Kaufman, B. R. Calvi, FlyBase Consortium, FlyBase 2.0: the next generation. *Nucleic Acids Res.* **47**, D759–D765 (2019).

60. A. M. Taverner, L. J. Blaine, P. Andolfatto, Epistasis and physico-chemical constraints contribute to spatial clustering of amino acid substitutions in protein evolution. *bioRxiv*, 2020.08.05.237594

61. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. **6**, 31 (2005).

62. J. B. Lack, J. D. Lange, A. D. Tang, R. B. Corbett-Detig, J. E. Pool, A Thousand Fly Genomes: An Expanded Drosophila Genome Nexus. *Mol. Biol. Evol.* **33**, 3308–3313 (2016).

63. A. Löytynoja, N. Goldman, A model of evolution and structure for multiple sequence alignment. *Philos Trans R Soc Lond B Biol Sci*. **363**, 3913–3919 (2008).

64. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *Journal of Molecular Biology*. **215**, 403–410 (1990).

65. R. C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. **5**, 113 (2004).

66. A. J. Lea, J. Altmann, S. C. Alberts, J. Tung, Developmental Constraints in a Wild Primate. *Am Nat*. **185**, 809–821 (2015).

67. A. J. Lea, D. Martins, J. Kamau, M. Gurven, J. F. Ayroles, Urbanization and market-integration have strong, non-linear effects on cardio-metabolic health in the Turkana. *bioRxiv*, 756866 (2019).

Acknowledgments:

We thank Andrew M. Taverner for providing the multi-species alignments, Erika Gadja for help collecting flies from the wild, and Michael Fernandez for help maintaining the experimental cages. We also thank members of the Ayroles and Andolfatto labs for their general support and helpful scientific discussions. **Funding:** LFP was funded by a Long-Term Postdoctoral Fellowship from the Human Frontiers Science Program. AJL was funded by a postdoctoral fellowship from the Helen Hay Whitney Foundation. JFA is funded by NIH-NIGMS R35GM124881-04 and NIH-NIEHS R01ES029929. **Author contributions:** LFP: conceptualization, formal analysis, investigation, writing-original draft, writing-review and editing; AJL: conceptualization, formal analysis, investigation, writing-review and editing, visualization; CH: conceptualization, formal analysis, investigation, writing-review and editing; EVF: investigation; PA: conceptualization, formal analysis, writing-review and editing, funding acquisition; JFA: conceptualization, formal analysis, writing-review and editing, funding acquisition **Competing interests:** Authors declare no competing interests. **Data and materials availability:** the data and code for analysis will be made available after publication.

Supplementary Materials:

Materials and Methods. References cited only in this section (*30–67*)

Figure S1. Mean read coverage for DNA libraries.

Figure S2. Sex and diet interact to affect survival.

Figure S3. Comparison of the magnitude of the genetic effect on longevity across conditions.

Figure S4. Allele frequency changes between T0 and TN.

Figure S5. Simulations to understand our definitions of "shared" and "GxE" SNPs.

Figure S6. Differential gene expression results between old and young flies.

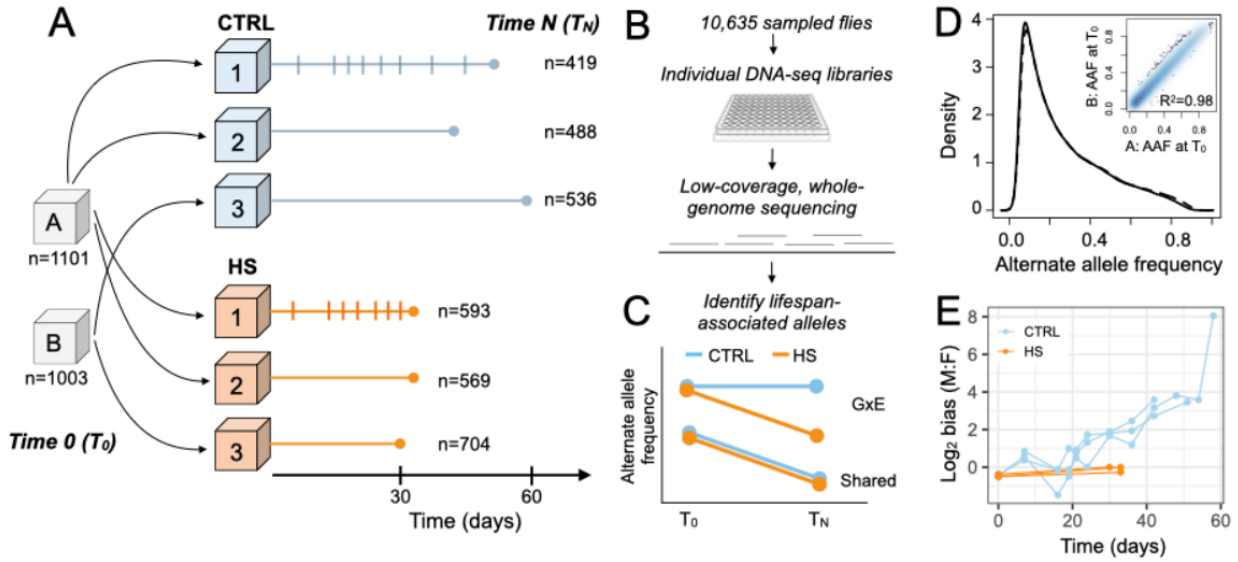
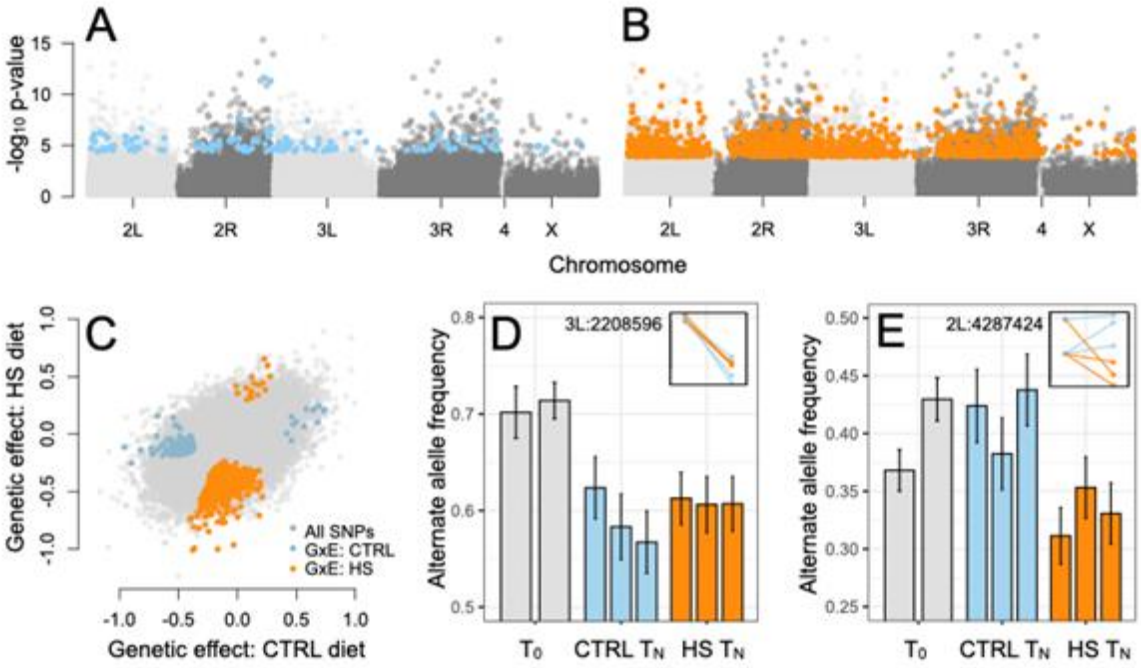


Fig. 1 Experimental design to detect GxE interactions modulating lifespan.

(**A**) *D. melanogaster* flies caught in Princeton, NJ were used to generate a synthetic outbred population that was kept under laboratory conditions for over a year, and split in two replicate cages prior to the beginning of the experiment (A and B). ~1000 flies were collected from A and B at the start of the experiment (T_0, 2±1 days old) and the rest were distributed into 6 replicate cages of ~10,000 flies each (3 cages = standard lab diet (CTRL), blue; 3 cages = high sugar diet, orange). ~500 flies were sampled every 3-7 days from a given cage and a last sample was taken when only ~500 flies were left (T_N) (Table S1-2); sampling schedule are noted by vertical dashed lines for the CTRL1 and HS1 cages. Identical schedules were followed for all other cages within a treatment group. To prevent pupae from the new generation from eclosing inside the experimental cages, food containers were replaced every three days. (**B**) Individually barcoded DNA-seq libraries were prepared from 10,635 individual flies sampled from T_0, T_N, and the intermediate time points. Each library was sequenced at ~1x depth to estimate allele frequencies and test for frequency changes across time (Fig. S1). **(C)** Expected patterns of frequency change are shown for alleles that reduce lifespan in both diets (shared) or more so on the HS diet (GxE). (**D**) The allelic composition of cages A (solid line) and B (dashed line) is very similar at T_0 (n=291,319 SNPs). Inset shows the per-site correlation between the alternate allele frequency (AAF) estimated for cage A versus B at T_0. (**E**) Log_2 ratio of males to females at different timepoints during the experiment. The number of flies sexed at each time point is provided in Table S4.



534

535

536 **Fig. 2 GxE interactions determine lifespan.**

537 (**A**, **B**) Manhattan plots highlighting significant lifespan-associated SNPs with GxE effects. Plots
538 show the -$\log_{10}$ p-value for tests for allele frequency differences between T_N and T_0 on a (A) CTRL
539 and (B) HS diet; colored points passed our significance filters for GxE effects (see methods). (**C**)
540 Comparison of model-estimated effect sizes for a genetic effect on lifespan on CTRL versus HS
541 diets (positive values indicate the alternate allele increases in frequency at T_N versus T_0). Only
542 SNPs with significant evidence for GxE effects are colored. (**D, E**) Allele frequency changes
543 across replicates for (D) an example SNP (3L:2208596) associated with lifespan in both dietary
544 conditions and (E) an example SNP (2L:4287424) with larger effects on lifespan on the HS diet.
545 The estimated alternate allele frequency is shown for each replicate cage, with bars representing
546 the standard error. The two T_0 bars correspond to cage A and B. The inset shows the mean
547 alternate allele frequencies at T_N and T_0, for each replicate CTRL (blue) and HS (orange) cage,
548 using the same x and y axes as in Figure 1B.

549

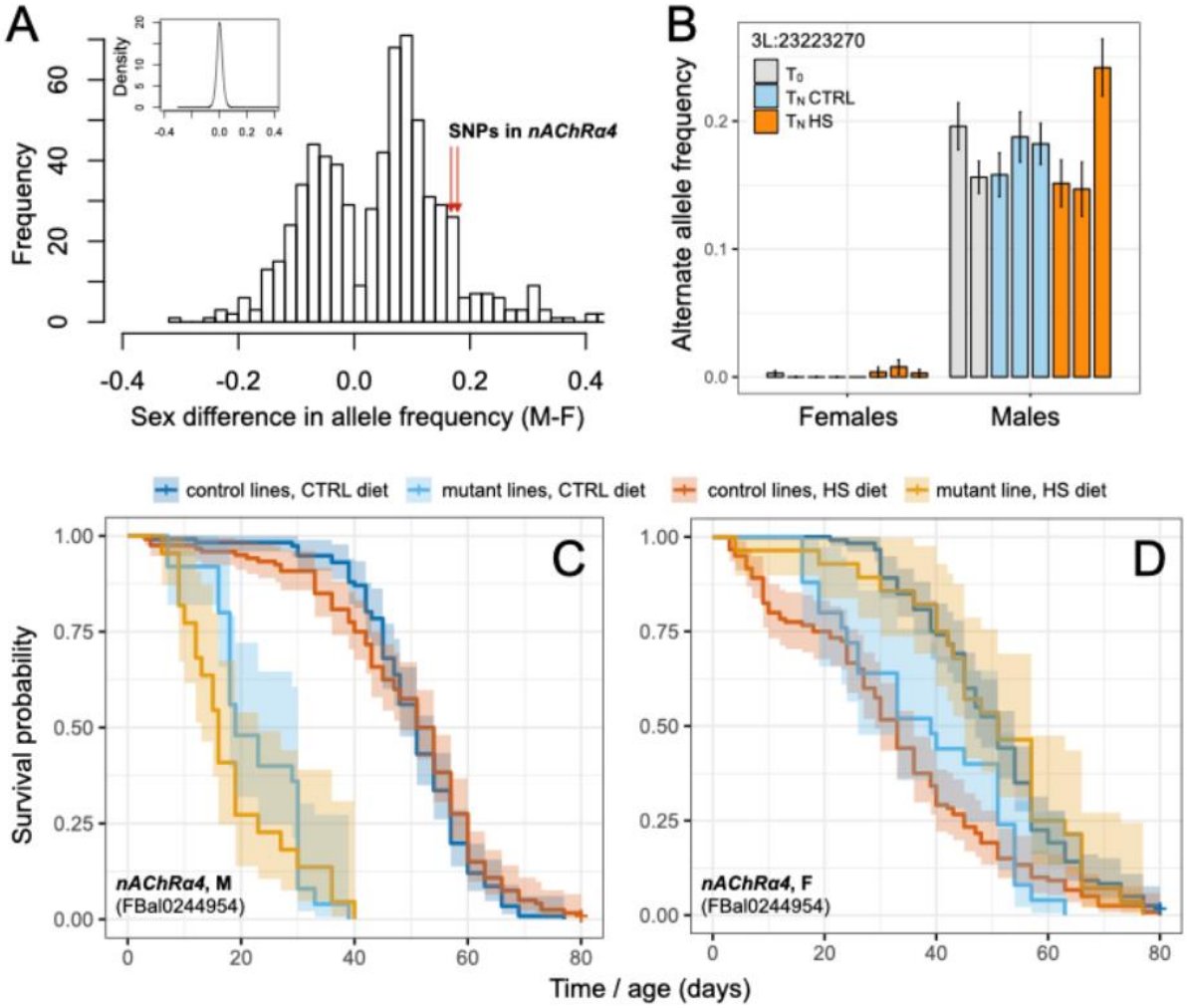
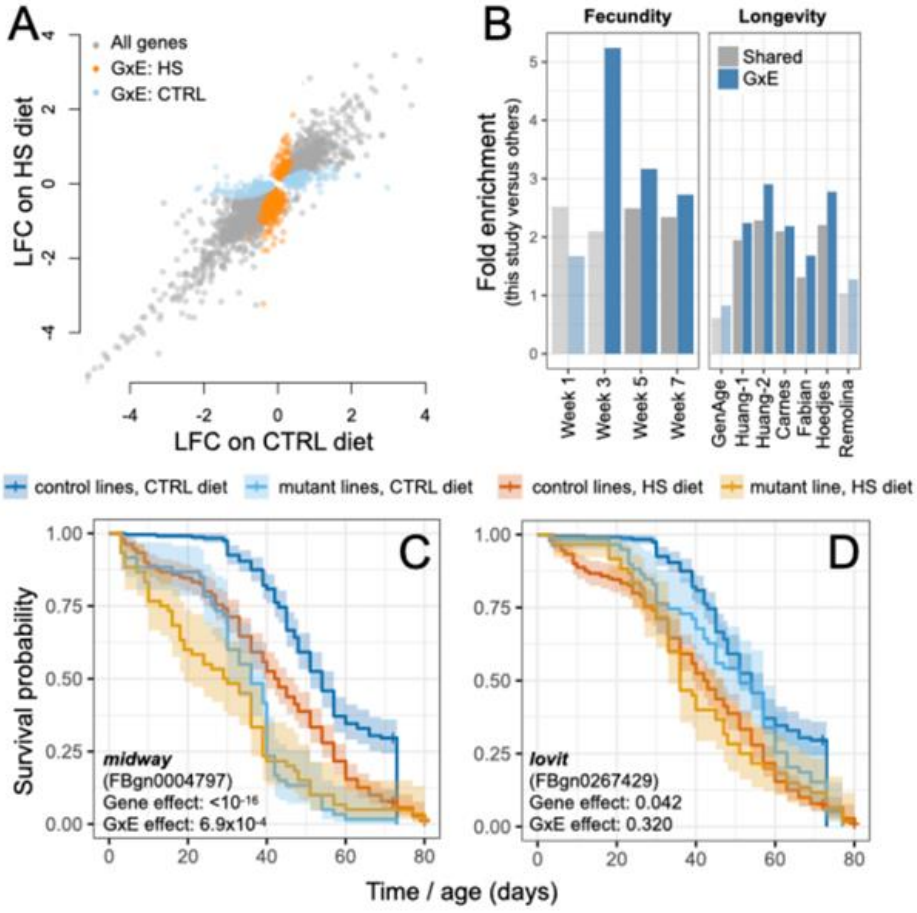


Fig. 3 Sex differences in viability associated with the *nAChRα4* gene.

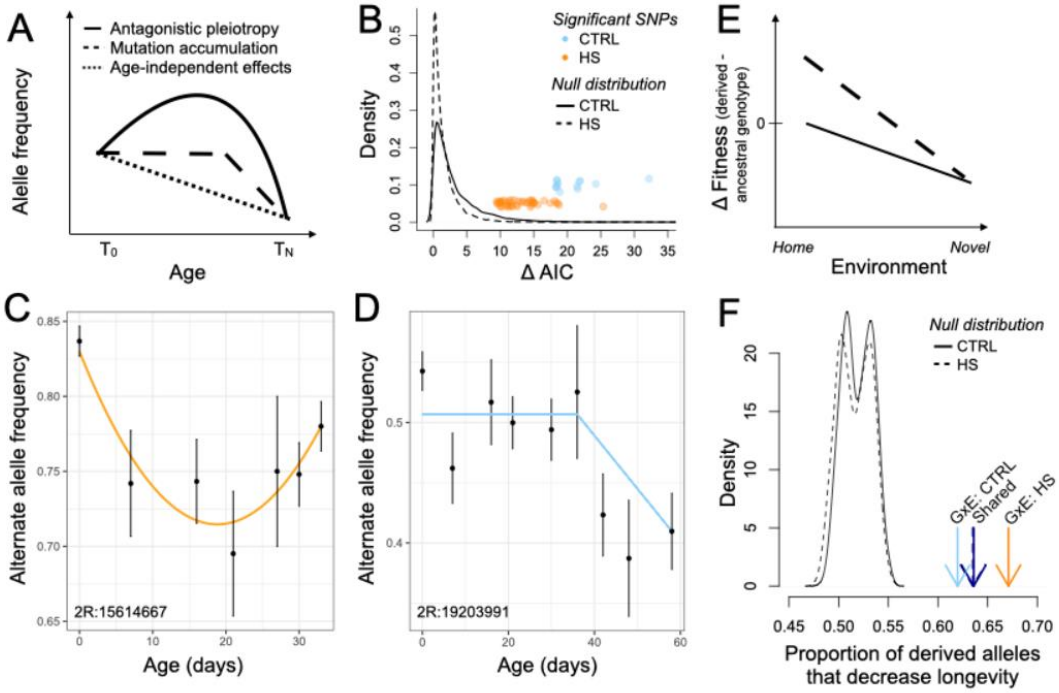
(A) Histogram of the difference in alternate allele frequency between males and females for SNPs with significant sex differences at T_0. Arrows indicate the effect size for SNPs in *nAChRα4*; the human ortholog -*CHRNA3*- has also been associated with lifespan/viability. Inset shows that the genome-wide distribution of alternate allele frequency differences between males and females at T_0 is centered at 0. (**B**) Allele frequency changes across replicates and time points for one of the *nAChRα4* SNPs (3L:23223270). The estimated alternate allele frequency is shown, with bars representing the standard error. (**C, D**) Survival curves for control lines versus loss-of-function *nAChRα4* mutants estimated for (**C**) males and (**D**) females, respectively. Kaplan-Meier survival curves for the control lines include data from four WT lines (DGRP 439, DGRP 181, Canton-S, and yw) (Table S14).



564

565 **Fig. 4 Properties of lifespan-associated genes.**

566 (**A**) Genes differentially expressed more so on a CTRL (blue) or HS (orange) diet. Axes show the
567 estimated log-fold change (LFC) in expression between young and old flies on each diet. Positive
568 values indicate higher expression in older flies. (**B**) Genes in or near (<1kb) SNPs with shared
569 (grey) or GxE (blue) effects on lifespan in this experiment overlap with lifespan and fecundity
570 genes identified in previous studies (identified by first author's last name). The degree of overlap
571 is represented as fold enrichment from a Fisher's exact test, and light bars indicate non-significant
572 overlap. Studies represent several types: GWAS for fecundity measured during weeks 1-7 in
573 inbred lines (*23*); selection for extended lifespan in outbred flies (*25–28*); analyses of standing
574 variation associated with lifespan in inbred lines (*14*); and "canonical" longevity genes from the
575 GenAge database (*29*). Light bars indicate non-significant overlap. (**C, D**) Kaplan-Meier survival
576 curves for two candidate genes, with p-values from a Cox proportional-hazards model testing for
577 an effect of the gene on survival as well as a GxE effect. Survival curves for the control lines
578 include data from four WT lines (DGRP 439, DGRP 181, Canton-S, and yw).



579

580

581 **Fig. 5 Evolutionary insight from longevity-associated SNPs.**

582 **(A)** Allele frequency trajectories across time according to the antagonistic pleiotropy and
583 mutation accumulation theories, and a constant trajectory not expected under evolutionary
584 models. We asked whether each lifespan-associated SNP could be confidently assigned to one
585 of these trajectories. (**B**) ΔAIC between the best and second-best model for each tested SNP. HS
586 and CTRL cages were analyzed separately due to the different age distributions within each
587 treatment (See Figure 1A). SNPs with ΔAIC values >99% of the null distribution are confidently
588 assigned to a given trajectory and their ΔAIC values are plotted as individual points. (**C-D**)
589 Examples of a (C) quadratic trajectory SNP (2R:15614667) in HS suggesting antagonistic
590 pleiotropy and (D) a breakpoint trajectory SNP (2R:19203991) in the standard lab environment
591 (CTRL) suggesting mutation accumulation dynamics. Points represent the mean alternate allele
592 frequency for a given age estimated across all cages, while bars represent the standard error of
593 the estimate. (**E**) Potential predictions from mismatch theory: alleles that evolved more recently
594 in the focal population (derived alleles) are neutral (black line) or advantageous (dashed line) in
595 the "home" environment they evolved in; however, they become deleterious in a "novel"
596 environment. (**F**) The derived allele is more likely than the ancestral allele to reduce lifespan
597 relative to chance expectations. For lifespan-associated SNPs shared between environments or
598 with stronger effects in HS or CTRL, the proportion of SNPs for which the derived allele is the
599 lifespan-reducing allele is noted with an arrow. The two arrows for shared SNPs represent the
600 proportion estimated using effect sizes from the CTRL (dark blue solid arrow) or HS (dark blue
601 dashed arrow) conditions, respectively. Null expectations were derived by performing the same
602 calculations on effect sizes from individual CTRL or HS cages across 1000 randomly drawn pools
603 of non-significant SNPs.