Laboratory evolution of flies to morphogen dosage via rapid maternal changes reveals predictable outcomes

Authors: Xueying C. Li[1,*], Lautaro Gandara[1], Måns Ekelöf[1], Kerstin Richter[1], Theodore Alexandrov[1,2,3], and Justin Crocker[1,*]

[1] European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

[2] Molecular Medicine Partnership Unit between EMBL and Heidelberg University, Heidelberg, Germany

[3] BioInnovation Institute, Copenhagen, Denmark

[*] Corresponding authors: X.C.L., xueying.li@embl.de, J.C., justin.crocker@embl.de

"The world was to me a secret which I desired to devine."

— Mary Shelley, Frankenstein

Abstract:

It remains unknown how developmental systems evolve in response to variable genetic and environmental conditions. Here, we have examined the evolvability of the classic *bicoid* network in *Drosophila*, which is essential for anterior-posterior patterning in the early embryo. This network can be synthetically perturbed by increasing the dosage of *bicoid*, which causes a posterior shift of the network's regulatory outputs and a decrease in fitness. To directly monitor network evolution across populations with extra copies of *bicoid*, we performed genome-wide EMS mutagenesis, followed by experimental evolution. After only 8-15 generations, evolved populations have normalized patterns of gene expression and increased survival. Using a phenomics approach, we find that populations normalized through rapid increases in embryo size driven by maternal changes in metabolism and ovariole development. We extend our results to wild populations of flies, demonstrating strong predictability. Together, our results necessitate a broader view of regulatory network evolution at the systems level. This study highlights the power of synthetic evolution using animal systems, a generalizable platform for the dissection of gene regulation and complex genomes.

Introduction

Changes in gene regulation underlie much of phenotypic evolution (Wittkopp and Kalay, 2012). However, our understanding of regulatory evolution is likely biased (Fuqua et al., 2020), as most evidence is derived from observations of sparse natural variation or limited experimental perturbations (Davies, 2017), especially in a developmental context. Furthermore, developmental networks orchestrate multiple processes that span a range of organizational scales—from single cells to tissues and organs and to even entire organisms (Davidson, 2010). These complex regulatory programs also integrate metabolic states (Miyazawa and Aulehla, 2018) and environmental cues in response to complex ecologies (Bergelson et al., 2021; Gilbert et al., 2015). However, developmental networks are often explored using a reductionist approach, focusing on particular time windows or pathways of development (Weber, 2022). While such approaches have been foundational to our understanding of development, this narrow focus may have limited our understanding of other 'possible' paths of regulatory evolution that are not taken in nature (Jacob, 1982). A more unbiased view might teach us about the constraints that govern evolutionary trajectories.

Quantitative genomics further challenges our models of how regulatory networks function—for complex traits, most of the heritability is likely due to a large number of variants, each with a small effect size (Boyle et al., 2017). Thousands of individual genes may contribute to phenotypes through expression in relevant cells (Boyle et al., 2017), and the contributions of each genetic variant to developmental fates are often small and challenging to measure (Frankel et al., 2011; Wood et al., 2014; Yengo et al., 2018). Therefore, it is essential to consider regulatory evolution and development both at the systems level and across populations (Gandara et al., 2022; Houle et al., 2010; Perkins et al., 2022). Clearly, approaches to elicit the relationships between different phenotypic layers and how these changes manifest across populations are needed to understand the evolution of developmental regulatory networks.

Laboratory evolution is a powerful method to gain insights into the mechanisms of molecular evolution and phenotypic adaptation. Such approaches have explored changes that accumulate in microbial populations during long-term selection experiments in response to specific growth conditions (Barrick et al., 2009), leading to insights into genomic, metabolic, and enzyme evolution (Bershtein and Tawfik, 2008; Dragosits and Mattanovich, 2013; Lenski et al., 2003; Levy et al., 2015; Sandberg et al., 2019). Combined with recent progress in genomics and

phenotyping techniques in multicellular organisms, it has become a powerful approach to measuring the robustness and evolvability of developmental networks (Miles et al., 2011) and patterns of genomic changes underlying adaptation (Mallard et al., 2018; Rudman et al., 2022). Importantly, laboratory "evo-devo" could combine controlled experimental conditions and precise genetic perturbations in the context of a complete developmental system to understand possible evolutionary trajectories without requiring *a priori* knowledge.

In this study, we leverage the power of laboratory evolution in flies (Engstrom, 1971; Payne, 1920) to explore the well-characterized early embryonic segmentation network in *Drosophila* (Nüsslein-Volhard and Wieschaus, 1980) in response to extra copies of *bicoid*, a key morphogen in *Drosophila* embryonic development. We were able to directly monitor developmental changes that rescue or mitigate the phenotypic defects caused by altered gene expression and, in some cases, to even generate novel phenotypes. Examining parallel-evolving populations with high temporal resolution allowed us to address the predictability of developmental evolution, whereas multi-modal phenotypic analyses provided mechanistic detail of the phenotypic changes. We found that compensatory changes for developmental perturbation can evolve rapidly in the lab, with extensive phenotypic changes in gene expression, metabolism and maternal anatomical features. Finally, we suggest that patterns observed in laboratory evolution can recapitulate phenotypic diversity in nature.

Rapid evolutionary responses to extra copies of *bicoid*

The *bicoid* network in *Drosophila melanogaster* is one of the best-understood developmental networks (Briscoe and Small, 2015). Bicoid is a transcription factor, the mRNA of which is maternally deposited at the anterior of the egg and forms a concentration gradient along the anterior-posterior (A-P) axis in the early embryo (**Fig. 1A-B, Supplemental Data File 1**). The gradient provides positional information for downstream target genes, such as *hunchback* (*hb*), *giant* (*gt*), *Kruppel* (*Kr*), and *even-skipped* (*eve*). These genes and others together constitute a complex network that determines segmentation (Driever and Nüsslein-Volhard, 1988) and scaling (Gregor et al., 2005; Houchmandzadeh et al., 2002) along the A-P axis of the embryo. The network directly responds to an increase in the gene dosage of *bicoid* (wild-type to 4x, **Fig. 1A-C**), whereby the increased λ of a steady-state exponential gradient (**Fig. 1D**) results in a shift of the cephalic furrow toward the posterior (Driever and Nüsslein-Volhard, 1988), indicated by a

posterior shift in the expression of *eve*, an essential segmentation gene expressed in a striped pattern (**Fig. 1E-F**). Despite the positional defects, embryos carrying two extra copies of the *bicoid* locus (4x*bcd*) can develop into normal adults—albeit with an increased frequency of cuticle defects (**Fig. 1G-I**) and reduced viability to adulthood (Berleth et al., 1988; Namba et al., 1997) (68.5%, **Fig. 1J**, **Table S1**).

The reduced viability of 4x*bcd* flies is a fitness disadvantage that can be a selection pressure in experimental evolution (**Fig. 1K**). To explore the system's capacity to respond to a perturbation of Bicoid levels, we established 15 parallel laboratory populations from 7 pools of chemically mutagenized 4x*bcd* flies (including replicates, see **Fig. S1**), along with three non-mutagenized populations which represent the standing variation in the lab stock. Based on whole-genome sequencing data, we estimated that the chemical mutagenesis with Ethyl methanesulfonate (EMS) introduced on average 2.7 point mutations per Mb. Thus, we estimated that the founding populations contained 1.7 million novel mutations (see **Methods, Fig. S1C**), providing genetic diversity for selection. We evolved the populations under standard laboratory conditions from the 4th generation after mutagenesis, after the generally deleterious mutations were purged in the first three generations, to select for compensatory mutations that can rescue or mitigate the fitness defect. We primarily used *eve* stripe positions as an indicator for compensatory changes: the compensated embryos should show *eve* stripes positions shifted to the anterior of the ancestral 4x*bcd* line [37.2 ± 0.4% egg length (EL) for the first *eve* stripe, 95% confidence interval, Generation 4] and closer to the wild-type positions (28.3 ± 0.6% EL for the first stripe, VK33)**.**

We found that compensation for the higher *bicoid* dosage occurred rapidly in our experimental populations. From the 4th to the 8th generation, the first *eve* stripe shifted to the anterior (toward the wild-type position) on average by 1.1% EL, from 37.2 ± 0.4% EL to 36.1 ± 0.2% EL ($p < 0.01$, Wilcoxon test) (all populations aggregated, **Fig. 1L, bottom panel**). Other stripes also showed different magnitudes of anterior shifts compared to Generation 4, ranging from 0.4% EL (stripe 7, $p = 0.04$, Wilcoxon test) to 1.0% EL (stripe 3 and 4, $p < 0.01$, Wilcoxon test) (**Fig. 1L, bottom panel**). Among these populations, there were heterogeneous responses in *eve* positions (**Fig. 1L, top panel**), with populations 1-1-3A and 2-6-1A showing compensatory shifts in more than one stripe in Generation 8 (**Fig. 1L**, **Fig. S2A**). We did not find a higher

similarity between replicate populations from the same mutant pool than those from different pools. Interestingly, the compensatory shifts in population 1-1-3A occurred through a shortened anterior region, whereas population 2-6-1A compensated via an expansion in the posterior region, suggesting multiple possible mechanisms for compensation (**Fig. S2B-E**). These shifts could not be explained by loss of *bicoid* expression because the Bicoid gradient in the evolved population remained the same as the 4x*bcd* ancestor line (**Fig. S2F**). Although these shifts are subtle compared to the drastic difference between 2x*bcd* and 4x*bcd*, a shift of 1% EL was the highest level of natural variation ever reported in *D. melanogaster* (Lott et al., 2007), suggesting that the early embryonic segmentation network can evolve rapidly in the lab under directed selection. In addition, the evolved populations showed increased viability, as measured by survival rates to eclosure after 16 generations (74.2 ± 2.5%, averaged across all populations) compared to the ancestral line (66.3 ± 3.4%), consistent with adaptation (**Fig. S1D**).

Unexpectedly, we found that compensation in the *bicoid* network coincided with an increase in egg length across the populations. From the 4th to the 8th generation, median embryo length increased from 550 um to 567 um (all populations aggregated, **Fig. 1M**, histogram, $p = 1.81e-09$, Wilcoxon test). Strikingly, despite variable embryo sizes, nine out of 12 populations showed an increase in median embryo length (1-1-1A, 2-2-1A, 2-2-2A, 2-3-1A, 2-3-2A, 2-4-1A, 2-5-1A, 2-5-2A and 2-6-1A; **Fig. 1M**, colored lines; **Fig. S3**) and three of them (2-2-2A, 2-5-1A, 2-6-1A) were statistically significant ($p < 0.05$, Wilcoxon test). This recurrent pattern suggests that an increase in embryo length might provide a quickly accessible mechanism to buffer the developmental stress caused by overexpression of *bicoid* and thus could drive the rapid compensatory changes we observed.

In parallel to phenotypic changes, we also found recurrent directional changes at the genomic level consistent with selection (**Fig. S4**). We performed low-coverage whole-genome sequencing for all 18 populations at the 3rd and 7th generation, and focused on changes in allele frequency in common variants shared across populations (i.e. standing variation) to understand the population dynamics at a broad scale. We found 16,394 biallelic variants showing consistent increases or decreases in allele frequency in two or more populations (Fisher's exact test, FDR-adjusted $p < 0.05$, **Supplemental Data File 2**). Based on a sign test, 181 of them were biased toward being maintained or purged in six or more populations (**Fig. S4C**). Recurrent gain or loss

of these alleles across multiple populations could suggest selection. For example, a non-synonymous mutation in Melted (F21V) was purged in six populations at the 7th generation (**Fig. S4D**), which could be beneficial because *melted* was linked to growth and metabolic pathways, and its mutant showed nutrient deprivation (Teleman et al., 2005). Other variants potentially under directed selection include those related to metabolism (e.g. *Apoltp*, **Supplemental Data File 2**) and ovariole development (*e.g. mtgo, bru3*, **Fig. S4D, Supplemental Data File 2**) (Lobell et al., 2017). These changes in allele frequency are consistent with rapid adaptation in the laboratory populations, with possible links to maternal and metabolic-related genes.

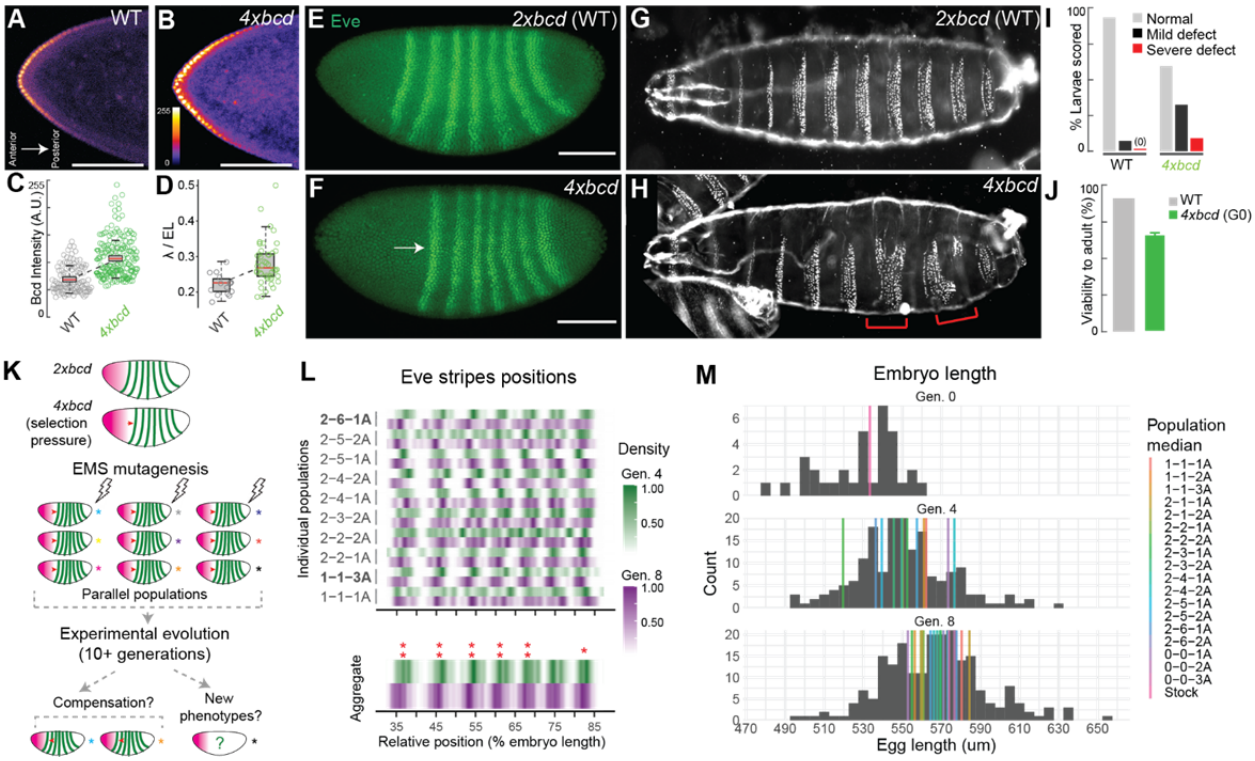


Fig. 1. Rapid laboratory evolution after perturbation of the *bicoid* network.
(**A**, **B**) Bicoid gradient along the anterior-posterior axis in embryos with two (wild-type) or four copies of *bicoid* (anti-Bicoid immunostaining, stage 4 embryos. Scale bar = 100 um.). (**C**) Bicoid levels in the ten most anterior nuclei, quantified across 11 and 12 embryos for wild-type and 4x*bcd*, respectively. (**D**) Bicoid gradient slopes, represented by decay constant λ scaled to egg length (EL). (**E, F**) Expression of *even-skipped* (*eve*) (anti-Eve immunostaining, stage 5 embryos). Scale bar = 100 um. (**G, H, I**) Cuticle phenotypes, with red brackets in (**H**) highlighting severe defects. (**J**) Viability to adulthood, with error bars representing the standard error of three measurements. (**K**) Scheme of experimental evolution. (**L**) Distribution of *eve* stripes positions in mid-stage 5 embryos, detected by *in situ* hybridization. Top, individual populations. Bottom, all populations aggregated (N=60 for Generation 4, N=217 for Generation 8). Intensity represents the scaled density of the designated population. Asterisks indicate shifts in the scaled position between generations. **, $p < 0.01$; *, $p < 0.05$ (Wilcoxon test, FDR-adjusted). (**M**) Distribution of embryo length across generations (grey histogram, all populations aggregated; N=34 for Generation 0, 176 for Generation 4, 217 for Generation 8). Color bars represent the median of each population. Population 0-0-1A, 0-0-2A, and 0-0-3A are non-mutagenized populations representing standing variation in the lab stock.

Compensation of *bicoid* overexpression through an increase in embryo length

To further address the possible link between embryo size and the *bicoid* network, we focused on population 2-6-1A to dissect the developmental changes before and after laboratory evolution. In this line, *eve* stripes consistently shifted to the anterior in the 8th and the 10th generation compared to the 4th generation (**Fig. 2A-C**; **Fig. S5A**), with the shift of the last stripe being the most prominent (**Fig. 2C-D**). We found that the shifts occurred simultaneously with an expansion of the posterior region: the egg length was consistently longer in both generations (540.5 ± 6.5 um at Generation 4, 573.5 ± 13.6 um at Generation 8, and 560.4 ± 7.1 um at Generation 10; **Fig. 2E, Fig. S5B**). The expression of *tailless*, a gap gene that specifies the posterior identity, was also wider in the 8th generation than the 4th generation (**Fig. 2F-I**). While the total number of nuclei along the A-P axis has not significantly changed (**Fig. S5C**), consistent with early embryos' limited capacity to regulate cell number (Busturia and Lawrence, 1994), there was a slight increase in the number of nuclei in the posterior region, from *eve* stripe 7 to the posterior pole at Generation 8 (12.3 ± 0.9 vs. 14.1 ± 1.1, $p = 0.048$, Wilcoxon test, **Fig. 2J**), as well as an overall increase in the distance between nuclei (6.39 ± 0.23 um vs. 6.82 ± 0.13 um, $p = 0.004$, Wilcoxon test, **Fig. S5D-E**). Consistent with compensatory evolution, the evolved line has stabilized phenotypes across phenotypic scales, including cuticle phenotypes (**Fig. 2K**) and viability to adulthood after 15-16 generations (**Fig. 2L**).

The compensation via embryo size appeared to be relatively short-term, because the embryo length of population 2-6-1A peaked at Generation 8 and 10, but gradually reduced after Generation 15 and resumed wild-type level at Generation 49 (**Fig. S5A-B**). This could be due to the fact that overly large embryos might have deleterious effects and cannot persist as a long-term solution in the standard environmental conditions employed in this work. Such a turnover in adaptive strategies is not uncommon in laboratory evolution (Good et al., 2017; Lenski, 2017; Levy et al., 2015). Future research along these lines could reveal alternative strategies to compensate for high *bicoid* dosage that is independent of embryo size, such as the response of Population 1-1-3A, which showed a shortened anterior region (**Fig. 1J**, **Fig. S2**).

Together, these data lead us to hypothesize that the compression of the trunk and tail caused by extra Bicoid might be mitigated in larger embryos due to more space in the posterior region. These results are consistent with previous findings on the interaction between egg size

and the *bicoid* network (Huang et al., 2020; Lott et al., 2007; Miles et al., 2011). Furthermore, because egg size is a highly polygenic and evolvable trait (Azevedo et al., 1996; Church et al., 2019; Jha et al., 2015), it might have provided a large target for selection and thus accessible as a means to respond rapidly.

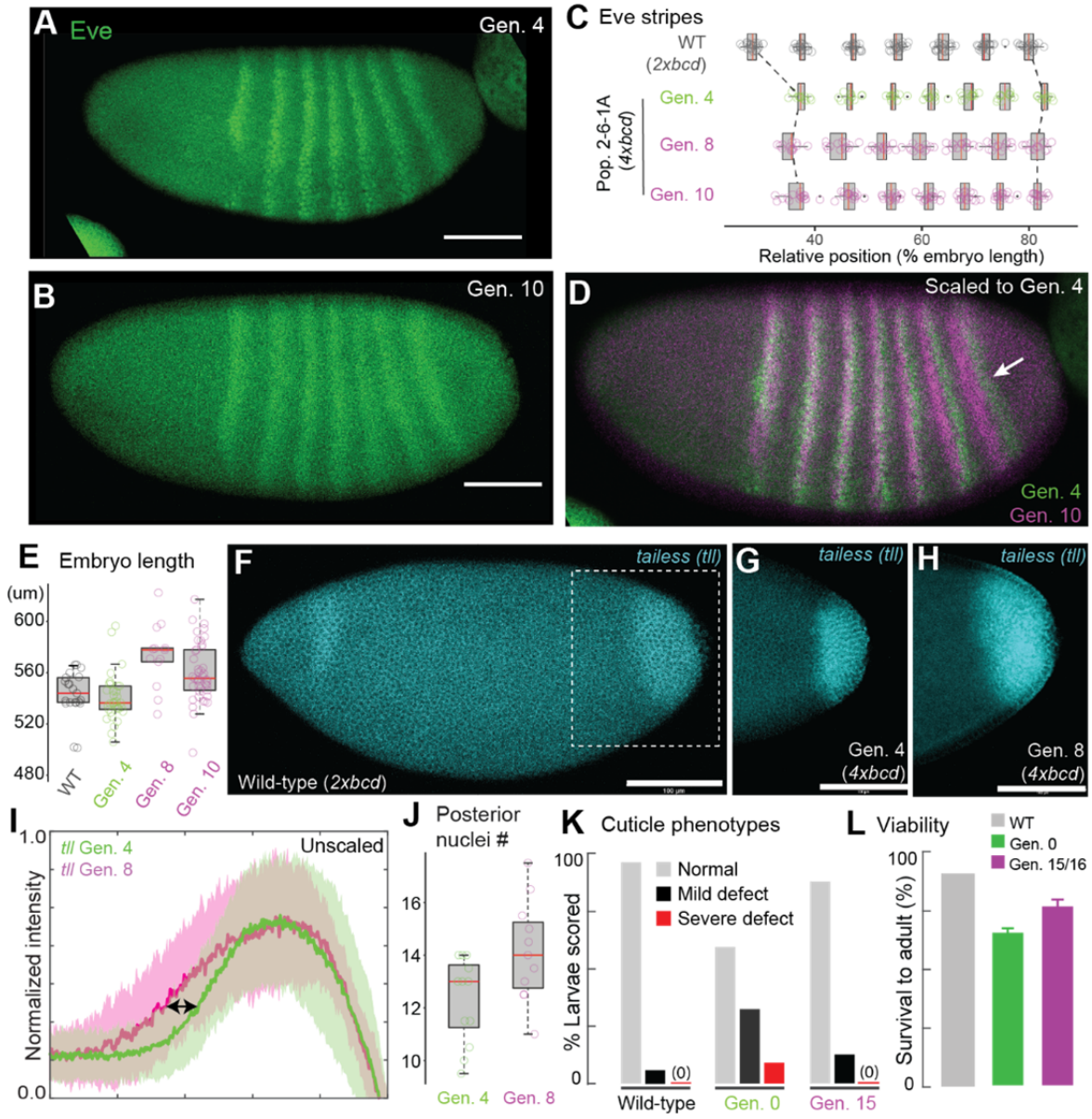


Fig. 2. Compensatory changes in gene expression, embryo length, cuticle, and viability.
(**A-D**) Eve stripes in Population 2-6-1A (anti-Eve staining), with the arrow in (**D**) showing a prominent anterior shift in the 7th stripe**.** The shifts were quantified in (**C**) from *in situ* data (*eve* co-stained with *sna*). (**E**) Increases in embryo length at the 8th and 10th generation. (**F – i**) *tailless* (*tll*) expression, detected by *in situ* hybridization. (**I**) shows the normalized intensity profiles aligned at the posterior end. Solid lines are average *tll* intensity and the shaded panels denote the standard deviation. N = 22 and 14 for the 4th and the 8th generation, respectively. (**J**) The number of nuclei from the posterior boundary of *eve* stripe 7 to the posterior pole. (**K**) Rescue of cuticle defects. (**L**)

Viability to adult, with error bars representing the standard error of three measurements (also see **Fig. S1**). Scale bar = 100 um.

Multi-modal analysis reveals changes in metabolism and ovariole development

To identify possible molecular bases that can support the rapid phenotypic stabilization through changes in egg length, we performed single-nuclei transcriptomics with early embryos in the evolved line (2-6-1A, Generation 20) (**Fig. S6, Table S2**). The evolved line had a striking increase in the proportion of yolk nuclei compared to wild-type or the 4x*bcd* lab stock (6% vs. 1%, **Fig. 3A**), consistent with the increased nutritional need of larger embryos. Among marker genes of the yolk cluster, there were 230 genes differentially expressed in the evolved line, including those related to metabolism (*bmm*, *trbl*, *Lime, Srr*) and cell growth (*crp*, *Traf4*) (**Fig. 3B, C**, and **Supplemental Data File 3**). Previous research suggests that the *Drosophila* body/organ size can be directly controlled by signaling pathways involved in metabolic regulation and cell growth, such as the insulin signaling pathway(Böhni et al., 1999; Oldham et al., 2002). We found a number of metabolic genes differentially expressed in the evolved line across multiple cell types, including epidermal ('*ovo*'), trunk ('*opa*'), anterior ('*oc*') and posterior ('*byn*') clusters in the ectoderm, as well as in mesoderm, endoderm, yolk and pole cells (**Fig. 3D, Supplemental Data File 4**).

The changes in yolk content and gene expression might imply a broader change in maternal metabolism to direct more nutrients into the eggs, and thus enable larger embryo sizes. Indeed, we found that the evolved embryos contained more triglycerides (TG) than two wild-type lines (**Fig. 3E**). Triglycerides are essential components of yolk-related lipid droplets (Welte, 2015) that can act as metabolic fuel for *Drosophila* embryogenesis (Tennessen et al., 2014), and high triglyceride levels have been linked to bigger embryo size in multiple animals (Mensch et al., 2021; Němec, 2002). To further quantify these changes, we performed MALDI-imaging mass spectrometry (Caprioli et al., 1997) on cryo-sectioned slices of ovaries to reconstruct entire mass spectra for single oocytes. We found differences in the lipid signature of oocytes between the evolved 4x*bcd* line (2-6-1A, Generation 42) and wild-type (w1118) (**Fig. S7A**), including elevated levels of triglycerides and decreased levels of glycerophosphocholines in the evolved line (**Fig. 3F-G**, **Fig. S7B-C**). Additionally, there were global differences in the fatty acid (FA) distribution in the evolved line, showing a higher abundance of FAs with 13, 14, and 15 carbons, and reduced levels of FAs with 18 carbons on their chain (**Fig. 3G**). This observation was

230 confirmed by tandem mass spectrometry coupled with MALDI-imaging (**Fig. S7D-F**). Overall,
231 these results show that the evolved line has altered its lipid metabolism in a way that is consistent
232 with bigger embryo sizes and higher energy requirements.

233 The changes in gene expression and lipid composition suggest rapid physiological
234 changes at the maternal level. We examined the ovaries of the evolved populations and found
235 that they tended to have fewer ovarioles (12.4 ± 0.3 vs. 14.8 ± 0.7, all populations aggregated vs.
236 wild-type aggregated, same below) and longer oocytes (498.4 ± 2.9 um vs. 458.0 ± 8.2 um) than
237 wild-type lines (**Fig. 3H-J**). Therefore, the compensation could occur through a trade-off
238 between the two traits (Church et al., 2021), possibly through growth-related mechanisms such
239 as the insulin pathway (Green, 2014; Jha et al., 2015).

240 To further explore this hypothesis, we next tested if the process could be recapitulated
241 genetically. Consistent with this hypothesis and previous reports (Green, 2014; Tu and Tatar,
242 2003), we found that overexpression of the gene *chico*— a key component of the insulin
243 signaling pathway—using the *nos:GAL4* driver and the Trip-OE system (Zirin et al., 2020), led
244 to a reduction in oocyte length of 475.2 ± 8.0 um to 459.8 ± 4.8 um ($p = 0.003$, Wilcoxon test)
245 and reduced levels of triglycerides of 0.346 ± 0.027 nmole/embryo to 0.125 ± 0.061 ($p = 0.00967$,
246 two-sided t-test, $n = 3$), demonstrating that oocyte size could evolve through such genes in a
247 short evolutionary timescale. Furthermore, we found that the change in size was specific to
248 oogenesis and likely to have metabolic rather than behavioral underpinnings because we did not
249 observe significant differences in larval length or larval feeding behavior (**Fig. S8**) (Church et al.,
250 2019).

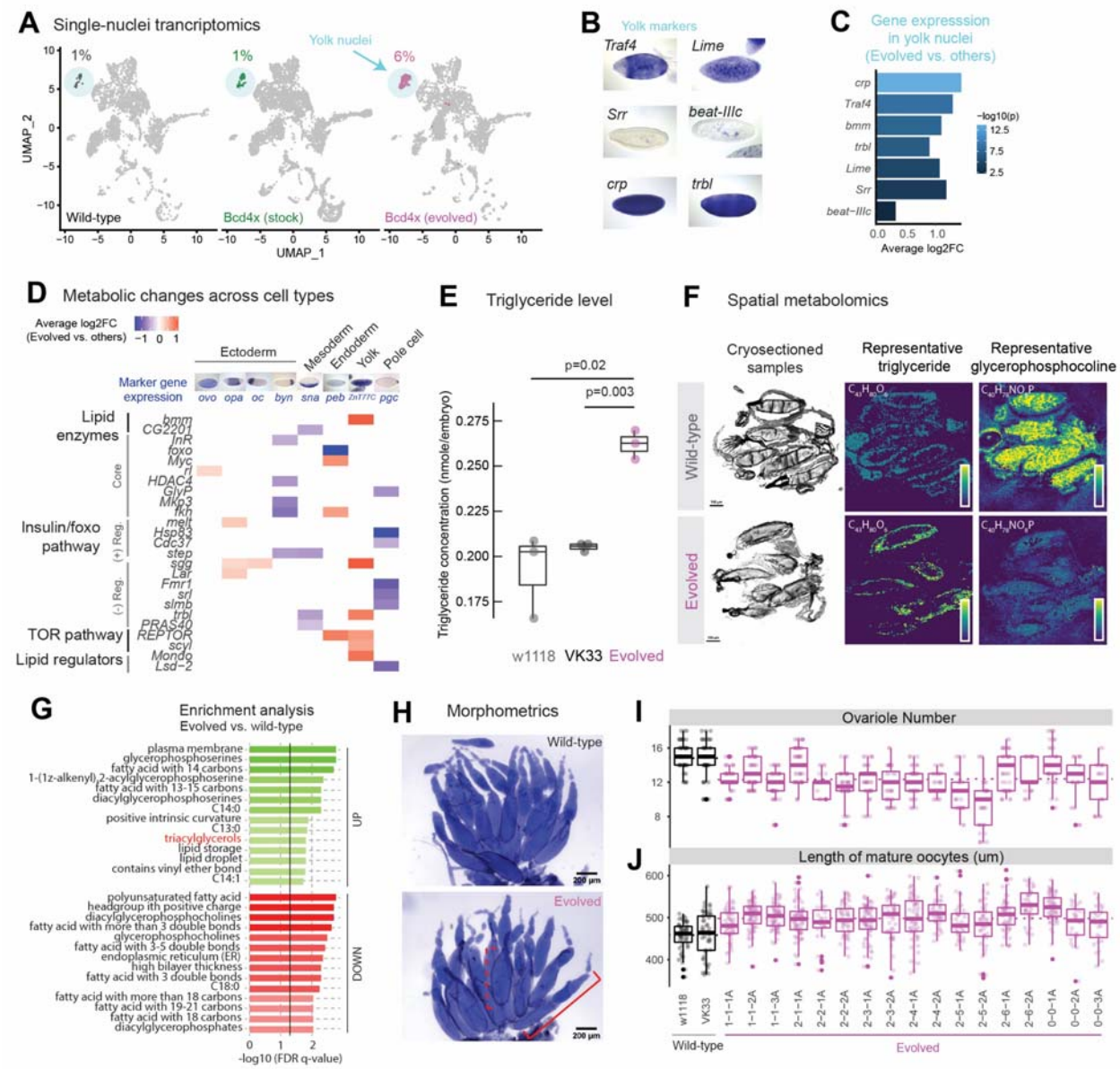


Fig. 3. Phenotypic changes in gene expression, metabolism and ovariole development in the evolved lines. (**A**) UMAP of single-nuclei transcriptomes of stage 5 embryos (see **Fig. S6** for details). The colored clusters show yolk nuclei. Wild-type is VK33. The evolved line is population 2-6-1A at Generation 20. (**B**) Representative marker genes of yolk nuclei. (**C**) Representative marker genes of yolk nuclei that were differentially expressed in the evolved line. (**D**) Changes in expression of metabolic genes across cell types between the evolved line and the other two samples. Only significant changes (adjusted p-value < 0.05) are shown. (+) Reg., positive regulators; (-) Reg., negative regulators. FC, fold change. Images of marker gene expression in (**B**) and (**D**) are from BDGP *in situ* database (Hammonds et al., 2013). (**E**) Enzymatic determination of triglyceride levels in stage 5 embryos (Generation 50 for population 2-6-1A). Points represent values from independent homogenates made from 50 embryos each. P values are from Student's t-test. (**F**) MALDI-imaging of ovaries. Left, middle sections from ovaries employed in MALDI-imaging. Scale bar = 100 µm. Middle, spatial distribution of a representative triglyceride, TG(40:1) at m/z=715.5846 normalized by another triglyceride which showed constant levels across all experiments (TG(44:3) at m/z=767.6159). Right, spatial distribution of a representative glycerophosphocholine, PC(32:1) at m/z=732.5537, in the sectioned ovaries. The evolved line is 2-6-1A from Generation 42. (**G**) Enrichment analysis comparing oocytes from the 2-6-1A and w1118 lines, based on the abundance values for 122 lipids detected through MALDI-imaging in both populations. The vertical solid line indicates a cutoff at FDR q-value of 0.05.

Triacylglycerols (highlighted in red) were enriched in the evolved line, consistent with results in (**E**). (**H**) Ovaries of wild-type (w1118) and evolved (2-6-1A, Generation 39) lines, stained with DAPI. The solid red bracket indicates an ovariole, and the dashed red bracket indicates a mature oocyte. Scale bar = 200 um. (**I**) Ovariole number and (**J**) length of mature oocytes of wild-type and the evolved lines (Generation 39). The horizontal dashed lines represent the mean of all wild-type/evolved lines aggregated (p = 9.783e-09 for ovariole number and p < 2.2e-16 for oocyte length, Wilcoxon test).

Laboratory evolution predicts phenotypes of wild populations

Embryo size is known to vary widely within and between *Drosophila* species (Lott et al., 2007) and across environments (Azevedo et al., 1996). As such, changes in embryo size could provide a way to rapidly mitigate the effects of Bicoid dose. To test if our observations could be extended to wild populations, we examined two near-isogenic lines isolated from the wild, Ind and Canton-S, with the former having larger embryos than the latter (Lott et al., 2007) (**Fig. 4A**). The anterior Bicoid concentration was also higher in the larger Ind embryos (**Fig 4B, C**), consistent with the relationship between Bicoid and embryo size in our laboratory-evolved lines, as well as previous results (Cheung et al., 2014). Strikingly, these two natural isolates also show differences in ovariole number and oocyte length (**Fig 4D**), as well as the level of triglycerides (**Fig. 4E**). Collectively, these observations suggest that the coupling among the *bicoid* network, egg size, maternal physiology, and metabolism could also exist in nature.

Next, to test if the bigger embryo size of the Ind genetic background could relieve the stress on the developmental network elicited by Bicoid overexpression, we crossed the *bicoid* transgenes into these inbred lines. In the crosses, the F1 offspring have 50% of genetic information from the wild isolates and have two extra copies of *bicoid* inserted on the second and the third chromosomes, respectively (4x*bcd* in total, see **Fig. S9A** for the crossing scheme). We also crossed them to a wild-type lab strain (VK33) to control for background effects. We found that embryos from F1 individuals in Ind/lab background were larger than those in Canton-S/lab background (**Fig. S9B**), suggesting that the Ind background had a dominant effect on embryo size. The *eve* stripes in Ind/lab background were located further to the anterior than the Canton-S/lab background in the control crosses (2x*bcd*) (**Fig. S9C**), suggesting natural variation in the capacity for scaling of the network. Such variation might be in favor of buffering stresses such as overexpression of *bicoid* - the difference was also present in embryos with 4x*bcd*, with the *eve* stripes of Ind embryos being anterior to those of Canton-S embryos, i.e. closer to the wild-type positions (**Fig. 4F-I**). Interestingly, the positions of *eve* stripes (**Fig. 4I**) and cuticle phenotypes (**Fig. 4J-L**) of 4x*bcd*-Ind embryos resembled those of the evolved population 2-6-1A in

laboratory evolution. 4x*bcd* embryos in the Ind background also had higher viability to adulthood compared with those in Canton-S or lab background (**Fig. S9D**), consistent with a higher tolerance of *bicoid* overexpression in larger embryos. Together, the evolved line is similar to Ind across a number of key phenotypes, supporting the hypothesis that changes in maternal contributions to embryo sizes could be used to buffer the dosage of *bicoid*.

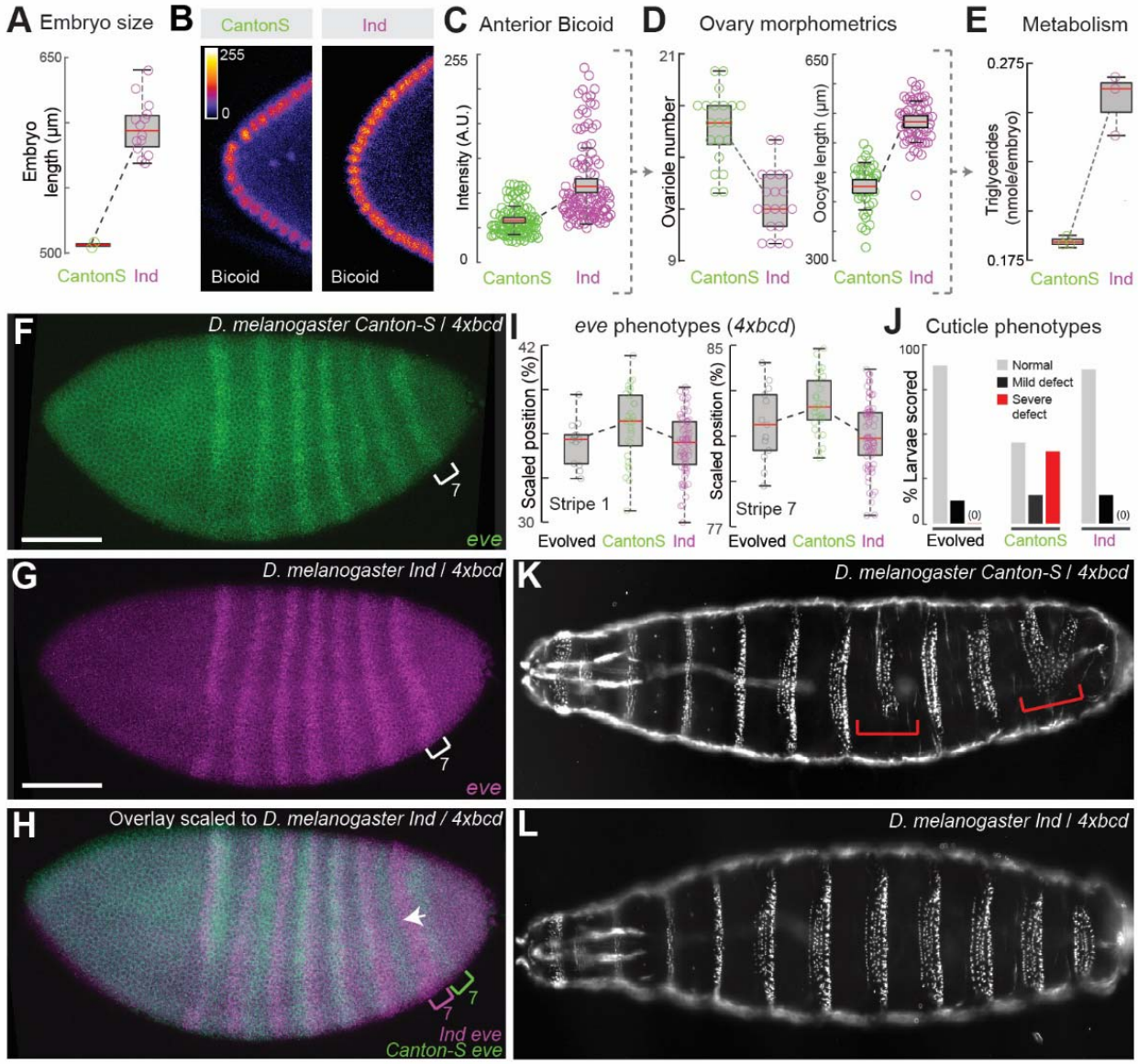


Fig. 4. Wild populations' responses to extra copies of *bicoid* and model for adaptation.
(**A**) Embryo size and (**B-C**) anterior Bicoid concentration (anti-Bicoid staining) of Ind and Canton-S. Each point represents one nucleus in (**C**), quantified across 18 and 10 embryos for Ind and Canton-S, respectively. (**D**) Ovariole number, oocyte length, and (**E**) level of triglycerides per embryo of Ind and Canton-S. (**F-I**) *eve* stripe positions and (**J-L**) cuticle phenotypes of Ind and Canton-S when carrying 4x*bcd*. Scale bar = 100 um. The red brackets in (**K**) highlight severe defects. Data for the evolved line in (**I**) were from Generation 8. See **Fig. S9** for full data.

The trends we found from laboratory evolution are consistent with our findings from the larger *D. melanogaster Ind* line, in line with evidence that *Drosophila* can adapt rapidly to laboratory culture on ecological timescales (Rudman et al., 2022). To explore the broader context of these results, we looked across a number of closely related *Drosophila* species (**Fig. 5A, Fig. S10**), testing the relationship between ovariole number and oocyte lengths (**Fig. 5B**). Strikingly, we see a strong correlation across the *Sophohora* subgenus indicating that such a trait may be consistent across a broader evolutionary context.

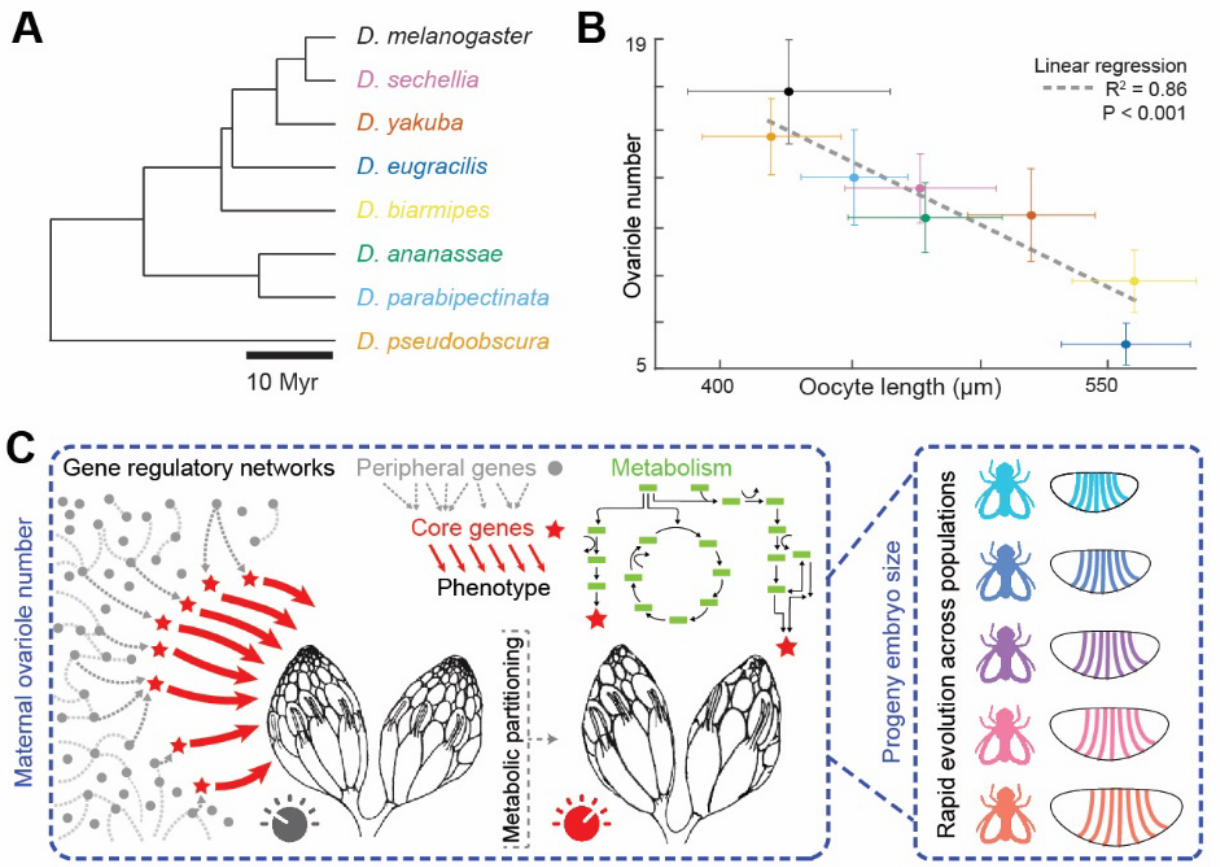


Fig. 5. Laboratory evolution predicts phenotypes of wild species.
(**A**) Phylogeny of species tested (Suvorov et al., 2022). (**B**) The relationship between ovariole number and oocyte length, error bars denote the standard deviation; colors are indicated in panel **A**. *D. melanogaster* was represented by Canton-S. (**C**) Model for maternal adaptation in laboratory evolution (adapted from (Liu et al., 2019)).

Discussion

Little is known about how organisms respond to developmental perturbations in a short evolutionary timescale. The early segmentation network downstream of Bicoid has been characterized as a highly dynamic (Bothma et al., 2014) yet robust network to ensure precise

302 scaling of gap gene boundaries (Cheung and Ma, 2015; Cheung et al., 2014; Gregor et al., 2005;
303 Lott et al., 2007). Perturbations to the network, such as a change in *bicoid* dosage, can lead to
304 substantial patterning defects and fitness disadvantages (Namba et al., 1997) (**Fig. 1**). Leveraging
305 the fitness disadvantage as a selection pressure provided us an opportunity to examine the
306 robustness and evolvability of developmental systems. Strikingly, we found compensatory
307 changes within 8-15 generations, reflected in gene expression, larval morphologies, and survival
308 to adulthood (**Fig. 1-2**), as well as recurrent changes in allele frequency among parallel-evolving
309 populations (**Fig. S3**), suggesting rapid adaptation in response to the developmental perturbation.
310 These results are consistent with the recent findings that adaptation in *Drosophila* was evident
311 over only one to four generations in response to environmental changes, including changes in
312 egg-size (Rudman et al., 2022). Such rapid phenotypic adaptation and large allele-frequency
313 shifts over many independent loci in response to developmental changes may be a common
314 mechanism for gene-regulatory network evolution (Rudman et al., 2022).
315 Our results support previous observations that embryonic geometry can affect the scaling
316 of gap gene boundaries under perturbations (Huang et al., 2020; Miles et al., 2011),
317 demonstrating an inherent link between the embryonic size-control network and the early
318 segmentation network. In particular, we found that the increase in egg length has the most
319 prominent effect on the posterior region (**Fig. 2**), consistent with a recent study showing that
320 posterior boundaries in *Drosophila* embryos are highly dynamic and sensitive to gene dosage
321 (Clark et al., 2022). The rapid phenotypic compensation driven by embryo size is likely related
322 to its genetic architecture. Egg size is a trait known to be both highly polygenic (Jha et al., 2015)
323 and evolvable in both common garden experiments (Miles et al., 2011; Rudman et al., 2022) as
324 well as across natural populations (Azevedo et al., 1996; Church et al., 2019; Lott et al., 2007).
325 As such, the egg-size network might provide a much larger set of targets for selection than
326 targets directly downstream of Bicoid, and hence the change in egg length appeared as the first
327 response in a short evolutionary timescale. These results are consistent with models that posit
328 that phenotypic evolution may be driven by many loci of small effect (Rockman, 2012; Zhang et
329 al., 2021). Furthermore, the rapid changes were associated with changes in ovariole number,
330 which is also known to be controlled by many genes (Lobell et al., 2017), resulting in changes in
331 metabolism and embryo size. Therefore, there could be numerous genes at different phenotypic
332 levels that provide evolutionary accessibility to compensation. It is possible that the

segmentation network, which can readily scale within and between species (Gregor et al., 2005), is the result of selection for a highly evolvable system that provides developmental plasticity for early embryos across variable ecologies (Moczek et al., 2011) (**Fig. 5C**).

The phenotypic differences of the evolved 4x*bcd* line were not limited to early embryonic development but included changes in lipid metabolism (increased yolk content and triglyceride levels), cell-type-specific gene expression (rewiring of metabolic gene network), and maternal anatomy (reduced ovariole numbers) (**Fig. 3**). These results show that perturbation of one node of the developmental network, the *bicoid* dosage, can lead to profound organism-wide responses across multiple phenotypic scales. Metabolic genes such as *melted*, *bmm*, and others identified in the genomic and transcriptomic analyses (**Fig. 3 and Fig. S3**) might provide a molecular basis for the observed evolutionary changes. Importantly, these observations highlight the deep connections between multiple phenotypic layers of multicellular systems and argue for a broader 'phenomics' perspective (Gandara et al., 2022), instead of a strictly gene-centric view. In the future, exploring the interplay of metabolic and developmental networks could transform our understanding of evolution and development across variable ecologies (Miyazawa and Aulehla, 2018; Perkins et al., 2022), as such processes are fundamentally linked (White et al., 2022).

Finally, our experiments show that it is not only possible to observe the evolution of developmental systems in the laboratory in real-time, but also to extend the results to trends found in nature. For example, one laboratory-evolved population (2-6-1A) showed similar rescue effects to a natural line with larger embryos (Ind), and, overall, similar phenotypes—including lipid profiles, ovariole number, and protein concentration—suggesting that the buffering mechanism observed in this study could exist in nature (**Fig. 4-5**). Reducing the inherent complexities among natural populations, such as variable environments, population dynamics, and standing variation, while also leveraging high-dimensional phenomics techniques, could allow us to investigate the evolution of complex developmental systems. Together, our results necessitate a broader view of developmental evolution at the systems level and suggest that such knowledge learned from experimental evolution can help predict evolutionary trends in nature.

Methods

Fly genetics

The eGFP-Bicoid fusion construct was designed according to (Gregor et al., 2007) (see **Supplemental data File 5** for the construct map). The construct was synthesized and cloned into placZattB by Genscript, and was transformed into *D. melanogaster* at the VK18 or VK33 landing site following standard PhiC31 integrase protocol, with the help of injection service provided by Alessandra Reversi at EMBL. The transformants at the VK33 site were homozygosed by sibling crosses to construct a stable 4x*bcd* line and subsequently used in mutagenesis and experimental evolution.

We also established balancer stocks from the transformants at VK18 (second chromosome) and VK33 (third chromosome) sites, and used them to generate a 6x*bcd* line, with an extra copy of *bicoid* on each of the second and the third chromosomes.

Overexpression of *chico* was done by Trip-OE system (Zirin et al., 2020). Virgin flies of NGT40>dCas9-VPR (Bloomington stock #67052; w[*]; P{w[+mC]=GAL4-nos.NGT}40; P{UAS-3XFLAG-dCas9-VPR}attp2) were crossed to males of gRNA lines, targeting sequences near transcription start sites of candidate genes. *chico*: #76114, y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=TOE.GS00909}attP40/CyO. Non-targeting control (QUAS): #67539, y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=GS00089}attP40). Non-CyO F1 females were dissected for ovariole analysis.

To examine the response to extra copies of *bicoid* in wild populations, virgins of Ind ("Mysore" strain, old stock #3114.4 from National Drosophila Species Stock Center, US) and Canton S (Bloomington stock #64349) were crossed to 6x*bcd* males. The F1 flies are heterozygous for the alleles from the wild populations and carry two extra copies of *bicoid*. They were used to set up egg-collection chambers and the F2 embryos were examined for *eve* expression, cuticle phenotypes, and fitness (**Fig. S9A**). To control for background effects, the natural isolates were crossed to the VK33 stock, which has the same background as the 6x*bcd* line.

At Generation 40, we outcrossed 2-6-1A males to wild-type w1118 or VK33 for four generations. In each generation, males with orange eyes (heterozygous for the *egfp-bicoid* transgene) were crossed to virgins of w1118 or VK33. After four generations, males and virgins with orange eyes were mated, and their progeny were selected for homozygotes (red eyes) to

create ‘new’ 4x*bcd* lines. In this way, we expect to remove or ‘dilute’ 2-6-1A-associated mutations and study the effects of 4x*bcd* without any compensatory evolution.

The non-*melanogaster* species were a generous gift from Nicolas Gompel, with the exceptions of *Drosophila parabipectinata* which was kindly provided by Artyom Kopp, and *Drosophila virilis*, which was kindly provided by Eileen Furlong. Strain background: *D. ananassae* (TSC 14024-0371.13), *D. biarmipes* (TSC 14023-0361.01), *D. eugracilis* (from the US National *Drosophila* Species Stock Center), *D. parabipectinata* (inbred derivative of strain TSC 14024-0401.02), *D. pseudoobscura* (TSC 14011-0121-94 USA), *D. sechellia* (TSC 14021-0248-25), *D. yakuba* (TSC 14021-0261.01) and *D. virilis* (w^-).

Mutagenesis and experimental evolution

EMS-mutagenesis was performed according to (Bökel, 2008). Briefly, around 1,000 4x*bcd* male flies (G0) were fed with 1% sucrose solution containing 25mM EMS, and were then mated to 4x*bcd* virgins. Around 3,500 F1 flies were used to establish 7 independent mutant pools, with 400-600 flies per pool. Specifically, the mutagenesis was done in two batches: flies from the first batch were used to establish one mutant pool, labeled 1-1, and flies from the second batch were used to establish six mutant pools, labeled 2-1 to 2-6. Mutation rate did not obviously differ between the two batches based on subsequent genomic analysis (see below).

Each mutant pool was used to seed 2-3 bottles of progenies consecutively (‘set A’) and these bottles were replicated at the 3rd generation (‘set B’), to provide 4-6 replicate populations in total for each mutant pool (**Fig. S1A**). For example, Pool 1-1 was used as parents to produce Populations 1-1-1A, 1-1-2A and 1-1-3A, by transferring the parents to a new bottle every 4-5 days. F3 flies from these populations were used as parents to produce Populations 1-1-1B, 1-1-2B, and 1-1-3B, respectively. Populations in set B were primarily for backups in this study.

The flies were maintained at 25°C under standard fly-rearing condition under non-overlapping generations, to select for rescuing mutations. The population size was approximately 200-500 for each generation. Three populations of non-mutagenized 4x*bcd* flies were maintained under the same condition for comparison (labeled 0-0-1A, 0-0-2A and 0-0-3A). During the first 15 generations, the populations were sampled every 2-5 generations for embryo collection, and the adult flies were frozen for genomic DNA (**Fig. S1B**).

Embryo fixation, antibody staining and fluorescent *in situ* hybridization

Drosophila embryos were fixed and stained following standard protocols (Galupa et al., 2022). In particular, stage 5 embryos were acquired from a 5 hr egg-laying window at room temperature. A fixation time of 18 min was used for these embryos, to adapt to the sensitivity of Eve antibody. The Eve antibody (mouse, Developmental Studies Hybridoma Bank, 2B8-concentrate) was used at 1:20 dilution. Bicoid antibody (rabbit) was a gift from Pinar Onal and Stephen Small, and was used at 1:250. DIG-, FITC- or biotin-labeled, antisense RNA-probes were used to detect gene expression of *eve*, *sna*, or *tll*, respectively. All embryos were mounted in ProLong Gold with DAPI, and imaged on a Zeiss LSM 880 confocal microscope, under 20x (air, 0.8 NA) or 25x (oil, 0.8 NA) objective.

Image analysis

All images were rotated to orient along the A-P axis before analysis.

Position of eve stripes. Images from *in situ* hybridization of *eve*, *snail* (*sna*) and *tailless* (*tll*) were used to quantify *eve* position precisely. We extracted the positions of the intersection of *sna* expression and the anterior boundary of each *eve* stripe in mid-stage 5 embryos (see **Fig. S2C** for an example), staged based on the degree of membrane invagination. The use of *sna* to mark a particular dorsal-ventral position on the *eve* stripes enabled precise quantification of the *eve* positions, which could also explain the differences between our results on Ind and Canton-S and a previous publication (Lott et al., 2007).

Embryo length. Embryo length was extracted from z-stacked confocal images, from anterior to posterior, excluding the pole cells.

Bicoid concentration. Bicoid intensities were acquired from anti-Bicoid staining by extracting the average nuclear intensity for ten nuclei at the anterior pole for each embryo, as per (Dubuis et al., 2013).

Slope of Bicoid gradient. Bicoid intensity along the A-P axis was measured at the depth of mid-embryo, by sliding a rectangle (smaller than a nucleus) along the edge of the embryo, from anterior to posterior (Houchmandzadeh et al., 2002). The shape of Bicoid gradient is described by the length constant λ (Cheung and Ma, 2015). The log-transformed, unscaled intensities between 10% to 50% egg length were fitted to a linear model, and the slope (k) from the linear model was used to calculate λ: $\lambda = -\frac{1}{k}$.

Tll profiles. The intensity profiles were extracted from a rectangular region of 3-4 cells' height along the A-P axis from max-projected confocal images (Crocker et al., 2016), normalized to peak intensities. The dorsal-ventral position was determined using the border of *sna* expression.

Nuclei counts. The number of nuclei along the A-P axis was counted along the *sna* border independently by two experimenters (X.C.L and L.G.). Numbers from the two experimenters were averaged for each embryo (see **Fig. S5** for more details).

Nuclei distance. While counting the nuclei, we marked the center of each nuclei and extracted their coordinates to calculate the distance between neighboring nuclei along the A-P axis.

Cuticle preparation

Overnight embryos were collected, bleached, rinsed and transferred into clean water in a petri dish, where they were allowed to develop for 24h at room temperature. After 24h, the larvae were transferred onto a glass slide and mounted in Hoyer's medium. The slides were baked in an oven at 55°C for 48h and were then imaged with dark field microscopy.

The cuticle images were scored based on the criteria from (Namba et al., 1997): severe defect – fusion or missing segments; mild defect – missing or misaligned denticles in any segment; normal – no visible defects. w1118 was used as wild-type.

Survival assay

Around 100 embryos from an overnight plate were manually transferred onto an apple juice plate with yeast in the center, and left at room temperature for 24h. On the second day, the number of unhatched embryos were counted for each plate, and the entire agar (with larvae and unhatched embryos) was transferred to a food vial. The eclosed adults were counted from day 12 until no adults came out. All the survival assays were performed at room temperature.

Whole-genome sequencing

Genomic DNA extraction and library preparation

We sequenced 20 F1 flies individually to estimate the level of genetic variation in the founding populations (1-4 flies from each mutant pool). To prepare genomic DNA from F1 individuals, each fly was squished and incubated at 37 °C for 30 min in Squish Buffer (10 mM Tris pH 8.0, 1 mM EDTA, 25 mM NaCl, 0.15 mg/ml Proteinase K), followed by a clean-up with

a Genomic DNA Clean & Concentrator kit (Zymo Research). The DNA was tagmented with a customized Tn5 protocol and sequenced in 75 bp (maximum 92 bp) paired-end on an Illumina NextSeq 500 at EMBL GeneCore.

Genomic DNA from the evolved populations was prepared using a Qiagen DNeasy Tissue Kit protocol (from Alexey Veraksa), with around 100 frozen flies (about 400 ul packed flies) per population. There are 38 samples: 18 populations × 2 generations (F3, F7) and 1 focal population (2-6-1A) × 2 additional generations (F9, F15). They were tagmented as described above and sequenced in 50 bp (maximum 88 bp) single-end on an Illumina NextSeq 2000, with a pooling strategy intentionally biased toward higher coverage of 2-6-1A samples.

Read mapping and variant calling

The reads were aligned to the dm6 genome with `Bowtie2` (Langmead and Salzberg, 2012), and duplicated reads were removed with `Picard` tools. To rule out Wolbachia infection, we aligned the reads to a Wolbachia reference genome (wMelPop, GCF_00475015.1), and found 0.0 % of reads aligned in all samples. After pre-processing, we acquired a total of 89.5 million reads for the 20 F1 individuals. As a preliminary analysis, we called variants in F1 individuals with `FreeBayes` (Garrison and Marth, 2012), with a threshold of 30 for mapping quality and 20 for base quality, on sites with a minimum coverage of 4. We found 375,779 variable positions among F1 individuals (variant quality score >10 and allele frequency < 1), suggesting a substantial amount of variation in the starting populations.

For pooled-sequencing (Pool-seq) of evolved populations, we obtained an average of 5 million reads for each non-focal sample, and an average of 16 million reads for 2-6-1A samples after pre-processing. Data from F1 individuals were computationally pooled. Together our reads cover 36.6% of the genome. Despite the shallow coverage, we regard each read to be randomly sampled from the population and the allele frequency may be roughly represented by the ratio of allele depth (AD). To extract this information, we used a pipeline adapted for Pool-seq data (Jha et al., 2015; Schlötterer et al., 2014): first, we realigned the reads around indels and performed base recalibration with `GATK4`, using the list of known variants in F1. Variable sites were then identified with `bcftools mpileup` and `bcftools call`, with allele depth (AD) extracted for each sample. 936,533 positions are found variable among the samples (variant quality score >10 and allele frequency < 1). The variants were then annotated with `ANNOVAR` (Wang et al., 2010).

Unfortunately, the shallow coverage did not allow us to confidently detect EMS-induced mutations in the population data. For the non-focal populations, there were 18-56 variants private to each mutant pool (at sites with sufficient coverage), and there were 1,663 private variants for pool 2-6, which is likely associated with the high coverage on population 2-6-1A. Therefore, we focused on common variants among the populations in the genomic analysis.

The NGS reads are deposited at ArrayExpress (EMBL-EBI) under experiment no. E-MTAB-11768.

Estimation of EMS mutation rate

We used the freebayes calls from the twenty F1 individuals to estimate the mutation rate induced by EMS treatment. To estimate the mutation rate, we needed to apply more stringent filters to remove background mutations. We first removed indels and sites with missing data in more than two individuals. Furthermore, we only kept sites with a mean depth between 4 and 50, and all genotypes with a depth outside this range were considered missing data. We then used `bcftools +prune` to remove small linkage blocks (sites with r^2 higher than 0.6 within a 1kb window), which were likely to be background variation. After these filters, there were 13,292 SNPs in the dataset. We then identified SNPs that were only present in one individual (minor allele count = 1), with a requirement of at least 3 reads supporting the observed allele (AO or RO >2). In this way, we identified 1,036 mutations across 19 mutagenized individuals (on average 55 mutations per individual) and 7 private SNPs in one non-mutagenized individual. Normalized to the number of bases covered in each individual (with the same quality and depth filter as when applying freebayes), the estimated mutation rate was on average 2.7 mutations per Mb, ranging from 0.9 to 5.4 mutations per Mb among individuals (**Fig. S1C**). The mutation rate was not obviously different between the two mutagenesis batches. Based on these data, we estimated the total number of novel mutations introduced to our experimental populations to be 2.7 × 180Mb × 3500 individuals = 1,701,000 mutations.

Changes in allele frequency of common variants

For each population, we used `bcftools +ad-bias` to apply fisher's exact test to compare allele ratio between F3 and F7, with requirements on the minimum alternative allele depth (2) and minimum depth (10). Out of the 450,739 biallelic sites tested, 54,045 (12%) sites show significant changes in allele frequency between generations in at least one population (FDR-adjusted $p < 0.05$). The changes in allele frequency span a wide range, with most changes being

transitions between homozygous and heterozygous states (**Fig. S3A**), which is probably associated with the detection limit imposed by sequencing depth (mean depth is 29 and median depth is 21 for the sites surveyed, **Fig. S3B**).

Since fisher's exact test might be an overly relaxed test on allele frequency and could lead to false positives (Jha et al., 2015; Turner et al., 2011), we applied a sign test (Orr, 1998) to narrow down the list of variants to those showing recurrent changes in multiple populations. Each variant is given a score: $S = N_{\text{REF increase}} - N_{\text{REF decrease}}$, where $N_{\text{REF increase}}$ is the number of populations showing a significant increase in reference allele frequency and $N_{\text{REF decrease}}$ is the number of populations showing a significant decrease in reference allele frequency. Therefore, the S score represents the tendency for the alternative allele to be purged (if $S > 0$) or fixed (if $S < 0$) during evolution. Out of the 450,739 biallelic sites tested, 16,394 sites (4%) showed consistent increases or decreases in allele frequency in more than one population. The mean of S among these sites is 0.56, suggesting a slight systematic bias for detecting decreases in alternative allele frequency, but the majority of the changes among populations are in random directions (mean S is close to 0). By using a cutoff of $S > 5$ or $S < -5$, we report on the top 1% sites (181 among 16,394) that show consistent directional changes across the parallel-evolving populations.

Genotype-phenotype association

Due to the low coverage and small sample size, we used genotype calls instead of allele frequency to perform genotype-phenotype association. We restricted this analysis to sites with a minimum mean depth of 10, leaving 261,167 sites in the dataset. We used the mean length of F4, F8, F10, and F17 embryos as the phenotype, to associate with the 'population genotypes' of their parent generation (F3, F7, F9, and F15). Note that we used the length of F17 embryos as the phenotype of F15 population, due to missing data in F16. For each variant, a linear model is used to estimate the effect size and significance of the genotype. For variants with three genotypes ("0/0", "0/1" and "1/1"), the smaller p-value is used. Due to the small sample size (30 samples at most), we don't think that the association analysis has enough statistical power to support any variant to be an interesting candidate, but the results could be used as a reference to prioritize variants detected by the sign test (e.g. the intronic G>T mutation in *CG1136* in **Fig. S3E**). The p-values are included in **Supplemental Data File 2**.

Single-nuclei transcriptomics

2.5h-to-3.5h-old embryos (developed at room temperature) were dechorionated and flash-frozen in liquid nitrogen for nuclei preparation. The evolved embryos are from population 2-6-1A, at the 20th generation. They were manually examined, and smaller embryos were removed upon collection, to reduce noise and focus on relatively large embryos. A wild-type line (VK33) and the 4x*bcd* lab stock were treated in parallel.

Nuclei isolation was performed following a standard protocol (10x Genomics® Single Cell Protocols, with adaptations from Francisca Hervas-Sotomayor at Heidelberg University). The frozen embryos were squished with a pestle for 20 times in cold homogenisation buffer (HB) [250 mM sucrose, 25 mM KCl, 5 mM $MgCl_2$, 10 mM Tris-HCl (pH 8), 0.1% Nonidet P40/IGEPAL, 1 uM DTT, 0.4 U/ul RNAse Inhibitor (New England Biolabs), 0.2 U/ul SUPERase•In™ RNase Inhibitor (Invitrogen)]. The samples were then centrifuged at 100 g for 1 min to remove unlysed tissue, and the supernatant was centrifuged at 1,000 g for 5 min to pellet the nuclei. The pellet was washed once in HB, filtered twice with Flowmi® Cell Strainers (Sigma), and resuspended in PBS. A subsample of the nuclei prep was DAPI-stained and examined under the microscope, to determine the density of nuclei. For each sample, 7,500 nuclei were used as the input for 10x library construction. RNA-seq was performed on an Illumina NextSeq 500 at EMBL Genomic Core Facilities (GeneCore) in two runs.

The reads were mapped to the *Drosophila* reference genome (dm6) plus the eGFP-Bicoid plasmid sequence and counted with Cell Ranger (6.0.1), with intronic reads included. The count data were analyzed with Seurat (3.9.9.9010) (Stuart et al., 2019) in R, with the three samples merged into one data frame. They were first filtered to remove 1) nuclei with extremely low (< 200) or high (> 4,000) number of expressed genes and 2) nuclei with a high percentage of mitochondrial reads (> 5%). The resulting data were normalized and scored for cell cycle status. The data were then scaled, with the percentage of mitochondrial reads, percentage of ribosomal reads, and cell cycle status regressed out. The scaled data were used for PCA, and `Harmony` (Korsunsky et al., 2019) was used to correct for batch effect, with 30 PCs. A preliminary clustering was done on the corrected data with 30 PCs and three clusters with predominantly cytosolic RNA (high percentage of ribosomal and mitochondrial RNA, low count in the number of genes and number of molecules) were removed.

After the removal, there are 3k to 6k nuclei for each sample. The data were normalized, scaled, 'harmonized' and clustered again as described above, with 30 PCs. There are 21 clusters,

611 with no obvious cluster of doublets based on scores generated by `scrublet` (Wolock et al.,
612 2019). Cell types were inferred based on marker genes (Karaiskos et al., 2017), and 11 clusters
613 were identified as early embryonic cell types based on marker gene expression at stage 4-6 [*in*
614 *situ* database of Berkeley Drosophila Genome Project (Hammonds et al., 2013)] (**Table S2**).
615 Differentially expressed genes were identified with `FindMarkers` in Seurat.
616 To curate a set of growth-related genes to examine expression changes across cell types,
617 we used the definition of insulin-like receptor signaling pathway in FlyBase (Gene group
618 FBgg0000910). Other genes were curated from (Choi et al., 2015), (Welte, 2015), (Heier and
619 Kühnlein, 2018), and (Heier et al., 2021).
620 The snRNA-seq reads are deposited at ArrayExpress (EMBL-EBI) under experiment no.
621 E-MTAB-12068.
622
623 **Triglycerides quantification assay**
624 The concentration of TGs in embryos was measured using the Triglyceride
625 Quantification Colorimetric Kit from Sigma (Cat. #MAK266). 50 stage5 embryos were
626 homogenized in Eppendorf tubes on a Nonidet P40 Substitute (Sigma, Cat. #74385) 5% solution.
627 The triglycerides concentration in each homogenate was then quantified following the
628 instructions provided by the manufacturer. Absorbance was measured at 570 nm.
629
630 **MALDI-imaging mass spectrometry on sectioned ovaries**
631 Ovaries needed to be cryo-sectioned to prepare the tissue for MALDI-imaging mass spec.
632 Breifly, a small number of ovaries was embedded in a previously heated 5% m/v
633 carboxymethylcellulose (Sigma) solution. This solution then solidifies at room temperature, and
634 the resulting molds were sectioned in a Leica CM1950 cryostat at -20°C, producing slices with a
635 thickness of 20 µm. These slices were then mounted on regular glass slides.
636 The samples were then coated with a microcrystalline matrix of 2,5-dihydroxybenzoic
637 acid dissolved in 70% acetonitrile to 15 mg/ml, with the help of a TM-Sprayer robotic sprayer
638 (HTX Technologies, Carrboro, NC, USA). The sprayer operated at a spray temperature of 80°C,
639 flow rate of 0.01 ml/min, track spacing of 3 mm and 10 passes, and the estimated surface
640 concentration was 3µg/mm^2. The glass slides were then mounted onto a custom adaptor and
641 loaded into the MS imaging ion source (AP-SMALDI5, TransMIT GmbH, Giessen, Germany).

642 Generated ions were co-axially transferred to a high mass-resolution mass spectrometer
643 (QExactive Plus mass spectrometer, ThermoFisher Scientific).
644 Metabolite annotation was performed using the METASPACE cloud software
645 (Alexandrov et al., 2019) with SwissLipids database (Aimo et al., 2015) (version 2018-02-02).
646 The Principal Component Analysis of these results was performed on R using the FactoMineR
647 and factoextra packages (http://factominer.free.fr/). Enrichment analysis were carried out using
648 LION/web (Molenaar et al., 2019).
649
650 **Dissection of ovarioles**
651 Flies were reared in uncrowded cages with apple juice plates supplied with yeast paste for
652 48h prior to dissection. 10-12 female flies were dissected for ovaries, which were kept on ice in
653 PBT with 4% PFA until all samples were processed. The ovaries were then fixed in PBT/PFA
654 for 30 min, washed twice in PBT and placed in Prolong Gold with DAPI. They were then further
655 dissected to separate the ovarioles and mounted on glass slides. The slides were imaged on a
656 Zeiss 880 confocal microscope and scored for ovariole number and oocyte length.
657
658 **Larval behavior**
659 Larvae (3rd instar, 5 days after egg laying) were harvested from food vials using a 10%
660 glucose solution and placed on agar plates, where their movement was recorded using a FL3-U3-
661 13Y3M-C CMOS camera (https://www.flir.de/products/flea3-usb3/) for two minutes. Then,
662 positional information as a function of time was automatically extracted from the videos for each
663 individual larvae using FIMtrack (Risse et al., 2017). Behavior-related parameters (speed,
664 bending, etc) were then calculated using this dataset.
665
666 **Acknowledgements**
667 We thank Phillip Oel, Leslie Pan, Nikolaos Papadopoulos, Blanca Pijuan-Sala and Xuefei Yuan
668 for their help and advice in single-nuclei transcriptomics. We thank Ching-Ho Chang for his
669 advice on genomic analysis. We thank Dimitri Kromm and Lars Hufnagel for their help in light-
670 sheet imaging. We thank Pinar Onal and Stephen Small for sharing the Bicoid antibody, and
671 Nicolas Gompel, Artyom Kopp, and Eileen Furlong for sharing *Drosophila* stocks. We also
672 thank Martijn Molenaar for discussions on the interpretation of the metabolomics data. We thank

Pinar Onal and members of the Crocker group for their input on the project, in particular Natalia Misunou, Mindy Liu Perkins, Albert Tsai, Rafael Galupa, Noa Ottilie Borst, and Gilberto Alvarez Canales for providing feedback on the manuscript. Mattew A. Benton, Claire Standley and Xitong Liang for feedback on the manuscript. X.C.L. and L.G. are supported by fellowships from the European Molecular Biology Laboratory Interdisciplinary Postdoc Programme (EIPOD) under Marie Skłodowska-Curie Actions COFUND (664726 and 847543, respectively). Research in the Crocker lab is supported by the European Molecular Biology Laboratory (EMBL).

Author contributions

Conceptualization: X.C.L, L.G., J.C. Investigation: X.C.L., L.G., M.E., K.R., J.C. Methodology: X.C.L., L.G., M.E., T.A., J.C. Formal analysis: X.C.L., L.G., J.C. Data curation: X.C.L., L.G., J.C. Visualization: X.C.L., L.G., J.C. Software: X.C.L., L.G., J.C. Resources: T.A. Supervision: T.A., J.C. Project administration: X.C.L., J.C. Funding acquisition: J.C. Writing, original draft: X.C.L., L.G., J.C. Writing, review & editing: X.C.L., L.G., M.E., K.R,, T.A., J.C.

Competing interests

The authors declare no competing interests.

Supplementary Material

Fig. S1. Mutagenesis, experimental evolution and sampling scheme.

Fig. S2. Response of different populations.

Fig. S3. Changes in embryo length in evolved populations.

Fig. S4. Changes in allele frequency in evolved populations.

Fig. S5. Embryonic phenotypes of the evolved line 2-6-1A.

Fig. S6. Single-nuclei transcriptomes of the evolved line 2-6-1A.

Fig. S7. Metabolic alterations in oocytes from the evolved line 2-6-1A.

Fig. S8. Quantification of crawling behavior of 3rd-instar larvae from 2-min videos.

Fig. S9. Cross *bicoid* transgenes into Ind and Canton-S.

Fig. S10. Ovaries in different *Drosophila* species.

Table S1. Viability of stocks carrying 2x-to-6x-*bicoid* (prior to selection).

Table S2. Marker genes and cell types for clusters in single nuclei RNA-seq.

Supplemental Data File 1 (video). Light-sheet imaging of eGFP-tagged Bicoid throughout embryonic development of 4x*bcd* embryos.

Supplemental Data File 2 (Microsoft Excel format). Variants with recurrent changes across multiple populations between Generation 3 and 7.

Supplemental Data File 3 (Microsoft Excel format). Marker genes of yolk cluster that are differentially expressed in the evolved line.

Supplemental Data File 4 (Microsoft Excel format). Differentially expressed genes between the evolved line and the other two samples.

Supplemental Data File 5 (Microsoft Word format). Map of eGFP-Bicoid construct used to generate the 4x*bcd* line in this study.

Reference

Aimo, L., Liechti, R., Hyka-Nouspikel, N., Niknejad, A., Gleizes, A., Götz, L., Kuznetsov, D., David, F.P.A., van der Goot, F.G., Riezman, H., et al. (2015). The SwissLipids knowledgebase for lipid biology. Bioinformatics *31*, 2860–2866. https://doi.org/10.1093/bioinformatics/btv285.

Alexandrov, T., Ovchinnikova, K., Palmer, A., Kovalev, V., Tarasov, A., Stuart, L., Nigmetzianov, R., Fay, D., Contributors, K.M., Gaudin, M., et al. (2019). METASPACE: A community-populated knowledge base of spatial metabolomes in health and disease. BioRxiv 539478. https://doi.org/10.1101/539478.

Azevedo, R.B.R., French, V., and Partridge, L. (1996). Thermal Evolution of Egg Size in Drosophila melanogaster. Evolution (N. Y). *50*, 2338. https://doi.org/10.2307/2410702.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature *461*, 1243–1247. https://doi.org/10.1038/nature08480.

Bergelson, J., Kreitman, M., Petrov, D.A., Sanchez, A., and Tikhonov, M. (2021). Functional biology in its natural context: a search for emergent simplicity. Elife *10*, 1–12. https://doi.org/10.7554/eLife.67646.

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., and Nüsslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. EMBO J. *7*, 1749–1756. https://doi.org/10.1002/j.1460-2075.1988.tb03004.x.

Bershtein, S., and Tawfik, D.S. (2008). Advances in laboratory evolution of enzymes. Curr. Opin. Chem. Biol. *12*, 151–158. https://doi.org/https://doi.org/10.1016/j.cbpa.2008.01.027.

Böhni, R., Riesgo-Escovar, J., Oldham, S., Brogiolo, W., Stocker, H., Andruss, B.F., Beckingham, K., and Hafen, E. (1999). Autonomous Control of Cell and Organ Size by CHICO, a Drosophila Homolog of Vertebrate IRS1–4. Cell *97*, 865–875. https://doi.org/10.1016/S0092-8674(00)80799-0.

Bökel, C. (2008). EMS Screens. In Methods in Molecular Biology (Clifton, N.J.), pp. 119–138.

Bothma, J.P., Garcia, H.G., Esposito, E., Schlissel, G., Gregor, T., and Levine, M. (2014). Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. Proc. Natl. Acad. Sci. U. S. A. *111*, 10598–10603. https://doi.org/10.1073/pnas.1410022111.

Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell *169*, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038.

Briscoe, J., and Small, S. (2015). Morphogen rules: design principles of gradient-mediated embryo patterning. Development *142*, 3996–4009. https://doi.org/10.1242/dev.129452.

Busturia, A., and Lawrence, P.A. (1994). Regulation of cell number in Drosopfiila. Nature *370*, 561–563. https://doi.org/10.1038/370561a0.

Caprioli, R.M., Farmer, T.B., and Gile, J. (1997). Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. Anal. Chem. *69*, 4751–4760. .

Cheung, D., and Ma, J. (2015). Probing the impact of temperature on molecular events in a developmental system. Sci. Rep. *5*, 13124. https://doi.org/10.1038/srep13124.

Cheung, D., Miles, C., Kreitman, M., and Ma, J. (2014). Adaptation of the length scale and amplitude of the Bicoid gradient profile to achieve robust patterning in abnormally large Drosophila melanogaster embryos. Development *141*, 124–135. https://doi.org/10.1242/dev.098640.

Choi, S., Lim, D.-S., and Chung, J. (2015). Feeding and Fasting Signals Converge on the LKB1-SIK3 Pathway to Regulate Lipid Metabolism in Drosophila. PLOS Genet. *11*, e1005263. .

Church, S.H., Donoughe, S., de Medeiros, B.A.S., and Extavour, C.G. (2019). Insect egg size and shape evolve with ecology but not developmental rate. Nature *571*, 58–62. https://doi.org/10.1038/s41586-019-1302-4.

Church, S.H., de Medeiros, B.A.S., Donoughe, S., Márquez Reyes, N.L., and Extavour, C.G.

766 (2021). Repeated loss of variation in insect ovary morphology highlights the role of development
767 in life-history evolution. Proc. R. Soc. B Biol. Sci. *288*, 20210150.
768 https://doi.org/10.1098/rspb.2021.0150.
769 Clark, E., Battistara, M., and Benton, M.A. (2022). A timer gene network is spatially regulated
770 by the terminal system in the Drosophila embryo. BioRxiv 2022.01.26.477848.
771 https://doi.org/10.1101/2022.01.26.477848.
772 Crocker, J., Ilsley, G.R., and Stern, D.L. (2016). Quantitatively predictable control of Drosophila
773 transcriptional enhancers in vivo with engineered transcription factors. Nat. Genet. *48*, 292–298.
774 https://doi.org/10.1038/ng.3509.
775 Davidson, E.H. (2010). The Regulatory Genome: Gene Regulatory Networks In Development
776 And Evolution - Eric H. Davidson - Google Books (Academic Press).
777 Davies, J. (2017). Using synthetic biology to explore principles of development. Dev. *144*,
778 1146–1158. https://doi.org/10.1242/dev.144196.
779 Dragosits, M., and Mattanovich, D. (2013). Adaptive laboratory evolution – principles and
780 applications for biotechnology. Microb. Cell Fact. *12*, 64. https://doi.org/10.1186/1475-2859-12-
781 64.
782 Driever, W., and Nüsslein-Volhard, C. (1988). The bicoid protein determines position in the
783 Drosophila embryo in a concentration-dependent manner. Cell *54*, 95–104.
784 https://doi.org/10.1016/0092-8674(88)90183-3.
785 Dubuis, J.O., Samanta, R., and Gregor, T. (2013). Accurate measurements of dynamics and
786 reproducibility in small genetic networks. Mol. Syst. Biol. *9*, 639.
787 https://doi.org/https://doi.org/10.1038/msb.2012.72.
788 Engstrom, L.E. (1971). Studies of the effects of two-way selection for ovariole number in
789 Drosophila melanogaster (University of Illinois at Urbana-Champaign).
790 Frankel, N., Erezyilmaz, D.F., McGregor, A.P., Wang, S., Payre, F., and Stern, D.L. (2011).
791 Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. Nature
792 *474*, 598–603. https://doi.org/10.1038/nature10200.
793 Fuqua, T., Jordan, J., van Breugel, M.E., Halavatyi, A., Tischer, C., Polidoro, P., Abe, N., Tsai,
794 A., Mann, R.S., Stern, D.L., et al. (2020). Dense and pleiotropic regulatory information in a
795 developmental enhancer. Nature *587*, 235–239. https://doi.org/10.1038/s41586-020-2816-5.
796 Galupa, R., Alvarez-Canales, G., Borst, N.O., Fuqua, T., Gandara, L., Misunou, N., Richter, K.,

797 Alves, M.R.P., Karumbi, E., Perkins, M.L., et al. (2022). Enhancer architecture and chromatin
798 accessibility constrain phenotypic space during development. BioRxiv *38*, 2022.06.02.494376.
799 https://doi.org/10.1101/2022.06.02.494376.
800 Gandara, L., Tsai, A., Ekelöf, M., Galupa, R., Noon, E.P.-B., Alexandrov, T., and Crocker, J.
801 (2022). Developmental phenomics suggests that H3K4 monomethylation catalyzed by Trr
802 functions as a phenotypic capacitor. BioRxiv 2022.03.15.484407.
803 https://doi.org/10.1101/2022.03.15.484407.
804 Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read
805 sequencing. ArXiv 1207.3907. https://doi.org/10.48550/arxiv.1207.3907.
806 Gilbert, S.F., Bosch, T.C.G., and Ledón-Rettig, C. (2015). Eco-Evo-Devo: developmental
807 symbiosis and developmental plasticity as evolutionary agents. Nat. Rev. Genet. *16*, 611–622.
808 https://doi.org/10.1038/nrg3982.
809 Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., and Desai, M.M. (2017). The dynamics
810 of molecular evolution over 60,000 generations. Nature *551*, 45–50.
811 https://doi.org/10.1038/nature24287.
812 Green, D.A. (2014). Developmental and Genetic Mechanisms of Ovariole Number Evolution in
813 Drosophila. Harvard University.
814 Gregor, T., Bialek, W., De Ruyter Van Steveninck, R.R., Tank, D.W., and Wieschaus, E.F.
815 (2005). Diffusion and scaling during early embryonic pattern formation. Proc. Natl. Acad. Sci. U.
816 S. A. *102*, 18403–18407. https://doi.org/10.1073/pnas.0509483102.
817 Gregor, T., Wieschaus, E.F., McGregor, A.P., Bialek, W., and Tank, D.W. (2007). Stability and
818 Nuclear Dynamics of the Bicoid Morphogen Gradient. Cell *130*, 141–152.
819 https://doi.org/10.1016/j.cell.2007.05.026.
820 Hammonds, A.S., Bristow, C.A., Fisher, W.W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis,
821 M., Yu, B., Frise, E., and Celniker, S.E. (2013). Spatial expression of transcription factors in
822 Drosophilaembryonic organ development. Genome Biol. *14*, R140. https://doi.org/10.1186/gb-
823 2013-14-12-r140.
824 Heier, C., and Kühnlein, R.P. (2018). Triacylglycerol metabolism in drosophila melanogaster.
825 Genetics *210*, 1163–1184. https://doi.org/10.1534/genetics.118.301583.
826 Heier, C., Klishch, S., Stilbytska, O., Semaiuk, U., and Lushchak, O. (2021). The Drosophila
827 model to interrogate triacylglycerol biology. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids

1866, 158924. https://doi.org/10.1016/j.bbalip.2021.158924.

Houchmandzadeh, B., Wieschaus, E., and Leibler, S. (2002). Establishment of developmental precision and proportions in the early Drosophila embryo. Nature *415*, 798–802. https://doi.org/10.1038/415798a.

Houle, D., Govindaraju, D.R., and Omholt, S. (2010). Phenomics: the next challenge. Nat. Rev. Genet. *11*, 855–866. https://doi.org/10.1038/nrg2897.

Huang, A., Rupprecht, J.-F., and Saunders, T.E. (2020). Embryonic geometry underlies phenotypic variation in decanalized conditions. Elife *9*, 1–21. https://doi.org/10.7554/eLife.47380.

Jacob, F. (1982). The possible and the actual (University of Washington Press).

Jha, A.R., Miles, C.M., Lippert, N.R., Brown, C.D., White, K.P., and Kreitman, M. (2015). Whole-Genome Resequencing of Experimental Populations Reveals Polygenic Basis of Egg-Size Variation in Drosophila melanogaster. Mol. Biol. Evol. *32*, 2616–2632. https://doi.org/10.1093/molbev/msv136.

Karaiskos, N., Wahle, P., Alles, J., Boltengagen, A., Ayoub, S., Kipar, C., Kocks, C., Rajewsky, N., and Zinzen, R.P. (2017). The Drosophila embryo at single-cell transcriptome resolution. Science (80-.). *358*, 194–199. https://doi.org/10.1126/science.aan3235.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. ru, and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

Lenski, R.E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME J. *11*, 2181–2194. https://doi.org/10.1038/ismej.2017.69.

Lenski, R.E., Ofria, C., Pennock, R.T., and Adami, C. (2003). The evolutionary origin of complex features. Nature *423*, 139–144. https://doi.org/10.1038/nature01568.

Levy, S.F., Blundell, J.R., Venkataram, S., Petrov, D.A., Fisher, D.S., and Sherlock, G. (2015). Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature *519*, 181–186. https://doi.org/10.1038/nature14279.

Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell *177*, 1022-1034.e6. https://doi.org/https://doi.org/10.1016/j.cell.2019.04.014.

Lobell, A.S., Kaspari, R.R., Serrano Negron, Y.L., and Harbison, S.T. (2017). The genetic architecture of Ovariole number in Drosophila melanogaster: Genes with major, quantitative, and pleiotropic effects. G3 Genes, Genomes, Genet. *7*, 2391–2403. https://doi.org/10.1534/g3.117.042390.

Lott, S.E., Kreitman, M., Palsson, A., Alekseeva, E., and Ludwig, M.Z. (2007). Canalization of segmentation and its evolution in Drosophila. Proc. Natl. Acad. Sci. *104*, 10926–10931. https://doi.org/10.1073/pnas.0701359104.

Mallard, F., Nolte, V., Tobler, R., Kapun, M., and Schlötterer, C. (2018). A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in Drosophila. Genome Biol. *19*, 1–15. https://doi.org/10.1186/s13059-018-1503-4.

Mensch, J., Di Battista, C., De Majo, M.S., Campos, R.E., and Fischer, S. (2021). Increased size and energy reserves in diapausing eggs of temperate Aedes aegypti populations. J. Insect Physiol. *131*. https://doi.org/10.1016/J.JINSPHYS.2021.104232.

Miles, C.M., Lott, S.E., Hendriks, C.L.L., Ludwig, M.Z., Manu, Williams, C.L., and Kreitman, M. (2011). Artificial selection on egg size perturbs early pattern formation in Drosophila melanogaster. Evolution (N. Y). *65*, 33–42. https://doi.org/10.1111/j.1558-5646.2010.01088.x.

Miyazawa, H., and Aulehla, A. (2018). Revisiting the role of metabolism during development. Development *145*. https://doi.org/10.1242/dev.131110.

Moczek, A.P., Sultan, S., Foster, S., Ledón-Rettig, C., Dworkin, I., Nijhout, H.F., Abouheif, E., and Pfennig, D.W. (2011). The role of developmental plasticity in evolutionary innovation. Proc. R. Soc. B Biol. Sci. *278*, 2705–2713. https://doi.org/10.1098/rspb.2011.0971.

Molenaar, M.R., Jeucken, A., Wassenaar, T.A., van de Lest, C.H.A., Brouwers, J.F., and Helms, J.B. (2019). LION/web: a web-based ontology enrichment tool for lipidomic data analysis. Gigascience *8*, giz061. https://doi.org/10.1093/gigascience/giz061.

Namba, R., Pazdera, T.M., Cerrone, R.L., and Minden, J.S. (1997). Drosophila embryonic pattern repair: How embryos respond to bicoid dosage alteration. Development *124*, 1393–1403. .

Němec, V. (2002). Quantitative changes in protein, glycogen and fat content in the eggs of the locusts, Locusta migratoria migratorioides and Schistocerca gregaria (Orthoptera), during

embryogenesis. Eur. J. Entomol. *99*, 557–559. https://doi.org/10.14411/eje.2002.072.

Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. Nature *287*, 795–801. https://doi.org/10.1038/287795a0.

Oldham, S., Stocker, H., Laffargue, M., Wittwer, F., Wymann, M., and Hafen, E. (2002). The Drosophila insulin/IGF receptor controls growth and size by modulating PtdInsP3 levels. Development *129*, 4103–4109. https://doi.org/10.1242/DEV.129.17.4103.

Orr, H. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution (N. Y). *52*, 935–949. .

Payne, F. (1920). Selection for high and low bristle number in the mutant strain "reduced." Genetics *5*, 501–542. https://doi.org/10.1093/genetics/5.6.501.

Perkins, M.L., Gandara, L., and Crocker, J. (2022). A synthetic synthesis to explore animal evolution and development. Philos. Trans. R. Soc. B Biol. Sci. *377*, 20200517. https://doi.org/10.1098/rstb.2020.0517.

Risse, B., Berh, D., Otto, N., Klämbt, C., and Jiang, X. (2017). FIMTrack: An open source tracking and locomotion analysis software for small animals. PLOS Comput. Biol. *13*, e1005530. .

Rockman, M. V (2012). The QTN Program and the Alleles That Matter for Evolution: All That's Gold Does Not Glitter. Evolution (N. Y). *66*, 1–17. https://doi.org/10.1111/j.1558-5646.2011.01486.x.

Rudman, S.M., Greenblum, S.I., Rajpurohit, S., Betancourt, N.J., Hanna, J., Tilk, S., Yokoyama, T., Petrov, D.A., and Schmidt, P. (2022). Direct observation of adaptive tracking on ecological time scales in Drosophila. Science (80-.). *375*. https://doi.org/10.1126/science.abj7484.

Sandberg, T.E., Salazar, M.J., Weng, L.L., Palsson, B.O., and Feist, A.M. (2019). The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. Metab. Eng. *56*, 1–16. https://doi.org/https://doi.org/10.1016/j.ymben.2019.08.004.

Schlötterer, C., Tobler, R., Kofler, R., and Nolte, V. (2014). Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. Nat. Rev. Genet. *15*, 749–763. https://doi.org/10.1038/nrg3803.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data.

Cell *177*, 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

Suvorov, A., Kim, B.Y., Wang, J., Armstrong, E.E., Peede, D., D'Agostino, E.R.R., Price, D.K., Waddell, P., Lang, M., Courtier-Orgogozo, V., et al. (2022). Widespread introgression across a phylogeny of 155 Drosophila genomes. Curr. Biol. *32*, 111-123.e5. https://doi.org/10.1016/j.cub.2021.10.052.

Teleman, A.A., Chen, Y.W., and Cohen, S.M. (2005). Drosophila melted modulates FOXO and TOR activity. Dev. Cell *9*, 271–281. https://doi.org/10.1016/j.devcel.2005.07.004.

Tennessen, J.M., Bertagnolli, N.M., Evans, J., Sieber, M.H., Cox, J., and Thummel, C.S. (2014). Coordinated Metabolic Transitions During Drosophila Embryogenesis and the Onset of Aerobic Glycolysis. G3 Genes|Genomes|Genetics *4*, 839–850. https://doi.org/10.1534/g3.114.010652.

Tu, M.-P., and Tatar, M. (2003). Juvenile diet restriction and the aging and reproduction of adult Drosophila melanogaster. Aging Cell *2*, 327–333. https://doi.org/https://doi.org/10.1046/j.1474-9728.2003.00064.x.

Turner, T.L., Stewart, A.D., Fields, A.T., Rice, W.R., and Tarone, A.M. (2011). Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. PLoS Genet. *7*. https://doi.org/10.1371/journal.pgen.1001336.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164–e164. https://doi.org/10.1093/nar/gkq603.

Weber, M. (2022). Philosophy of Developmental Biology (Cambridge: Cambridge University Press).

Welte, M.A. (2015). As the fat flies: The dynamic lipid droplets of Drosophila embryos. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids *1851*, 1156–1185. https://doi.org/https://doi.org/10.1016/j.bbalip.2015.04.002.

White, C.R., Alton, L.A., Bywater, C.L., Lombardi, E.J., and Marshall, D.J. (2022). Metabolic scaling is the product of life-history optimization. Science (80-.). *377*, 834–839. https://doi.org/10.1126/SCIENCE.ABM7649.

Wittkopp, P.J., and Kalay, G. (2012). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet. *13*, 59–69. https://doi.org/10.1038/nrg3095.

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. *8*, 281-291.e9. https://doi.org/10.1016/j.cels.2018.11.005.

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173. .

Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., and Visscher, P.M. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum. Mol. Genet. *27*, 3641–3649. https://doi.org/10.1093/hmg/ddy271.

Zhang, W., Reeves, G.R., and Tautz, D. (2021). Testing Implications of the Omnigenic Model for the Genetic Analysis of Loci Identified through Genome-wide Association. Curr. Biol. *31*, 1092-1098.e6. https://doi.org/https://doi.org/10.1016/j.cub.2020.12.023.

Zirin, J., Hu, Y., Liu, L., Yang-Zhou, D., Colbeth, R., Yan, D., Ewen-Campen, B., Tao, R., Vogt, E., VanNest, S., et al. (2020). Large-Scale Transgenic Drosophila Resource Collections for Loss- and Gain-of-Function Studies. Genetics *214*, 755–767. https://doi.org/10.1534/genetics.119.302964.

Fig. S1.

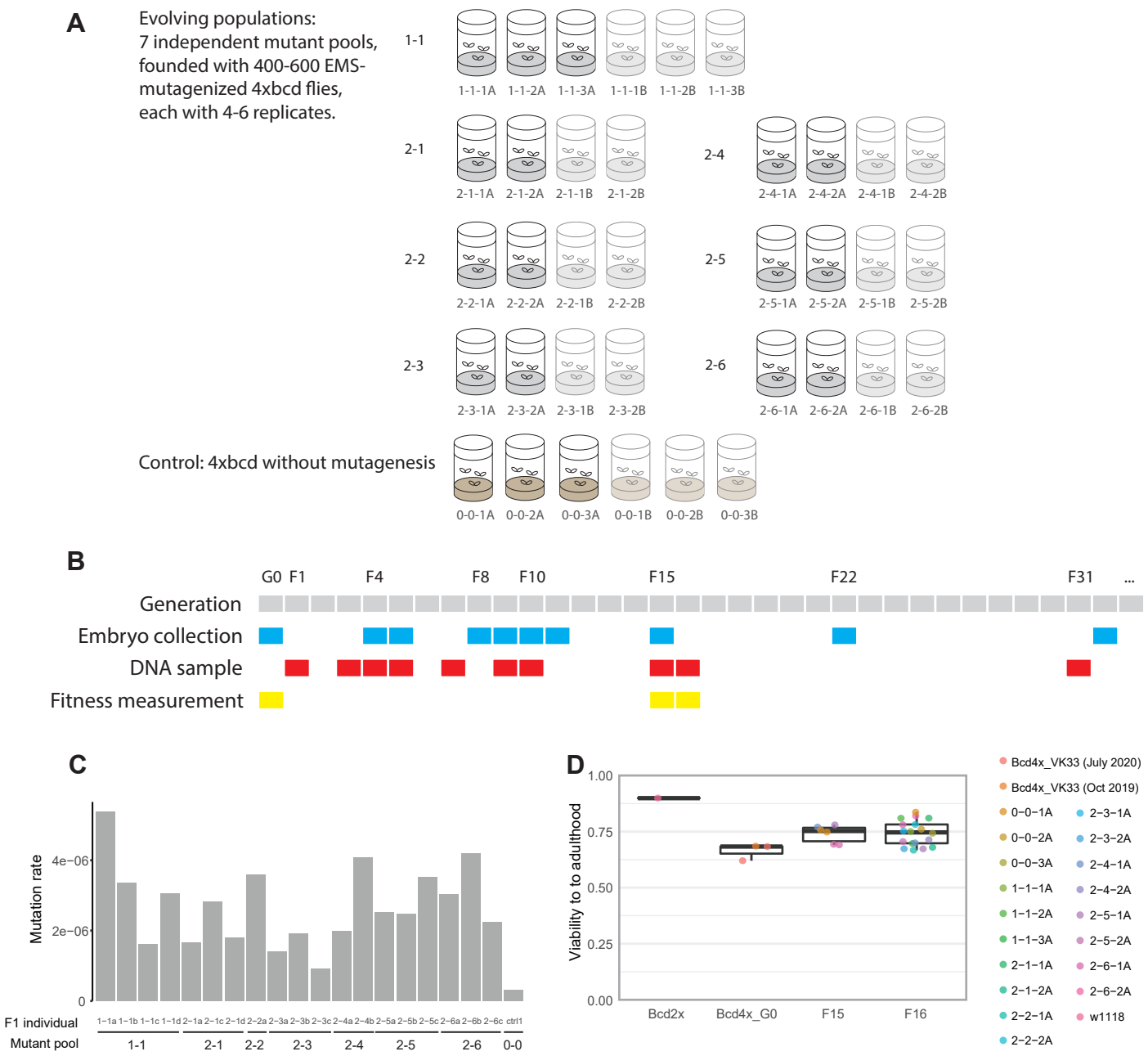


Fig. S1. Mutagenesis, experimental evolution and sampling scheme. (**A**) The evolving populations were from 7 independent mutant pools each with 4-6 replicates. 4xbcd flies without mutagenesis were evolved in parallel as control for standing variation in the lab stock. (**B**) Sampling time for embryo collection, DNA and fitness measurement. (**C**) Mutation rate of 19 mutagenized F1 individuals from the seven mutant pools and one individual from the background population, measured by the number of mutations divided by number of bases covered by whole-genome sequencing in each invidual. (**D**) Viability to adulthood for the evolving populations, compared to wild-type (Bcd2x, w1118) and the ancestral stock (Bcd4x_VK33, measured at two time points).

Fig. S2.

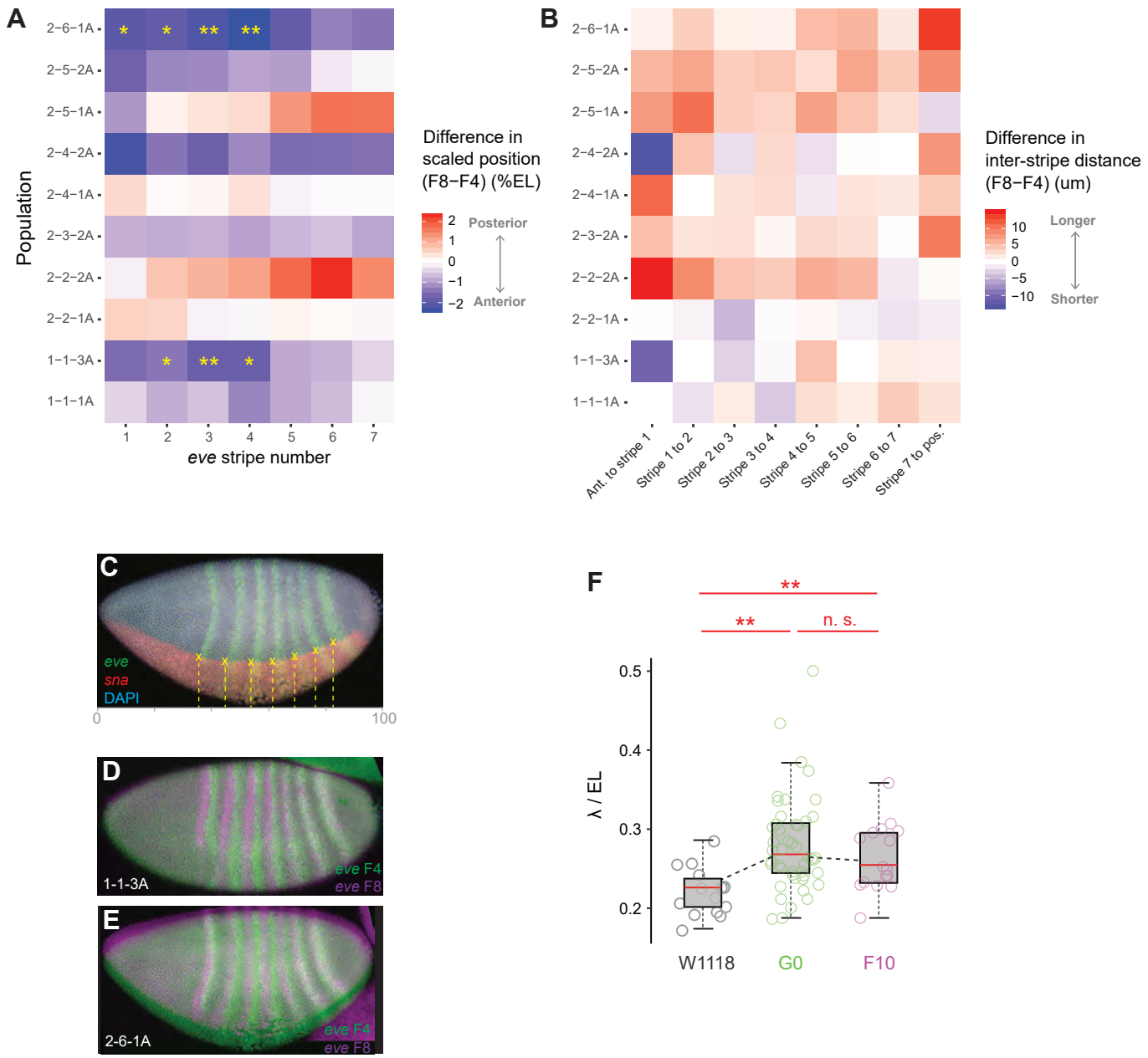


Fig. S2. Response of different populations. (**A**) Heatmap shows the difference in average *eve* stripe positions between F4 and F8 in each population. Asterisks indicate statistically different distributions between F4 and F8 (Wilconxon test). *, FDR-adjusted $p < 0.05$; **, FDR-adjusted $p < 0.01$. (**B**) The difference in average inter-stripe distance between generations. (**C**) A representative embryo, showing that the *eve* stripe positions were measured at the intersection between *eve* and *sna* expression, detected by *in situ* hybridization. (**D-E)** Representative embryos from 1-1-3A (**D**) and 2-6-1A (**E**), with F4 and F8 embryos scaled together. (**F**) The decay constant λ of Bicoid gradient, scaled to egg length (EL). The shape of Bicoid gradient was not significantly changed across generations in Population 2-6-1A. Asterisks indicate p values from Wilcoxon test. **, $p < 0.01$. n. s., not significant.

Fig. S3.

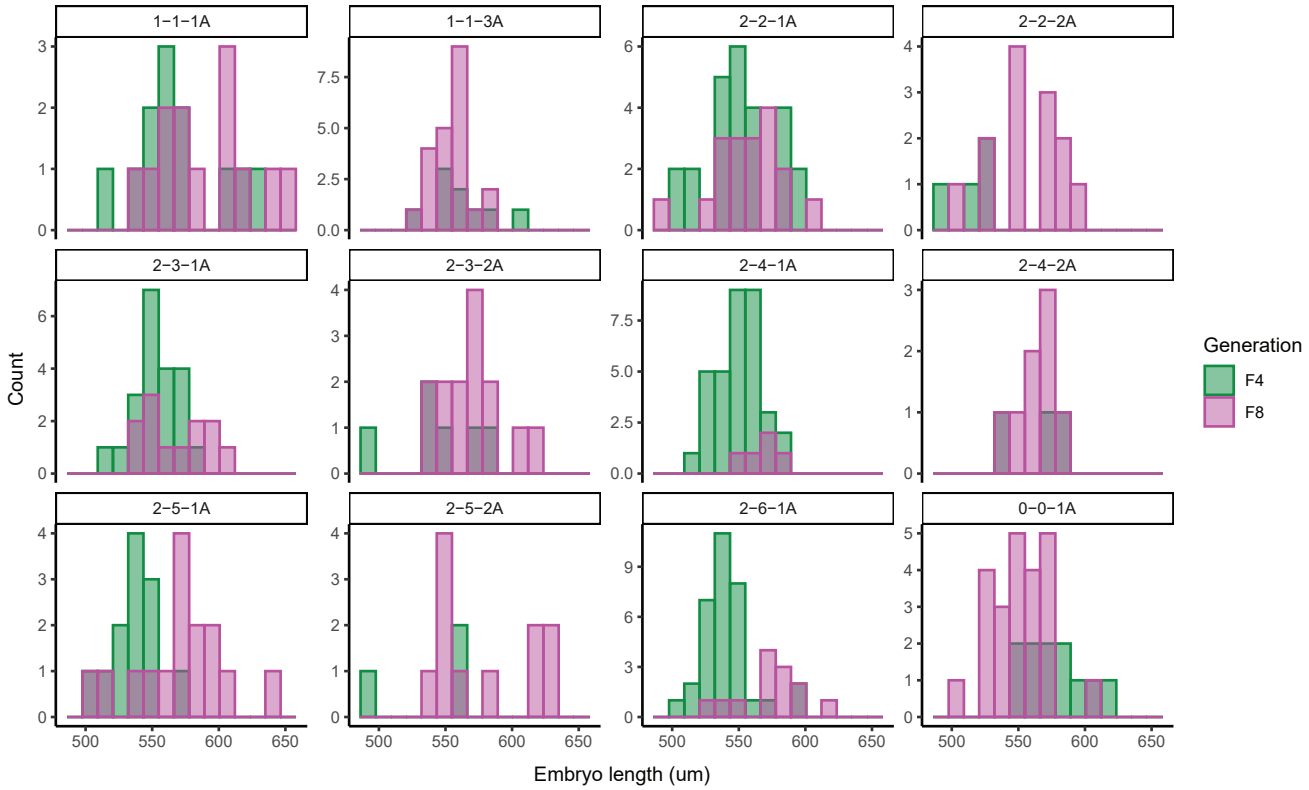


Fig. S3. Changes in embryo length in evolved populations. F8 embryos in 2-2-2A, 2-5-1A, and 2-6-1A were significantly larger than F4 ($p = 0.01$, 0.003, and 0.0003, respectively, Wilcoxon test).

Fig. S4.

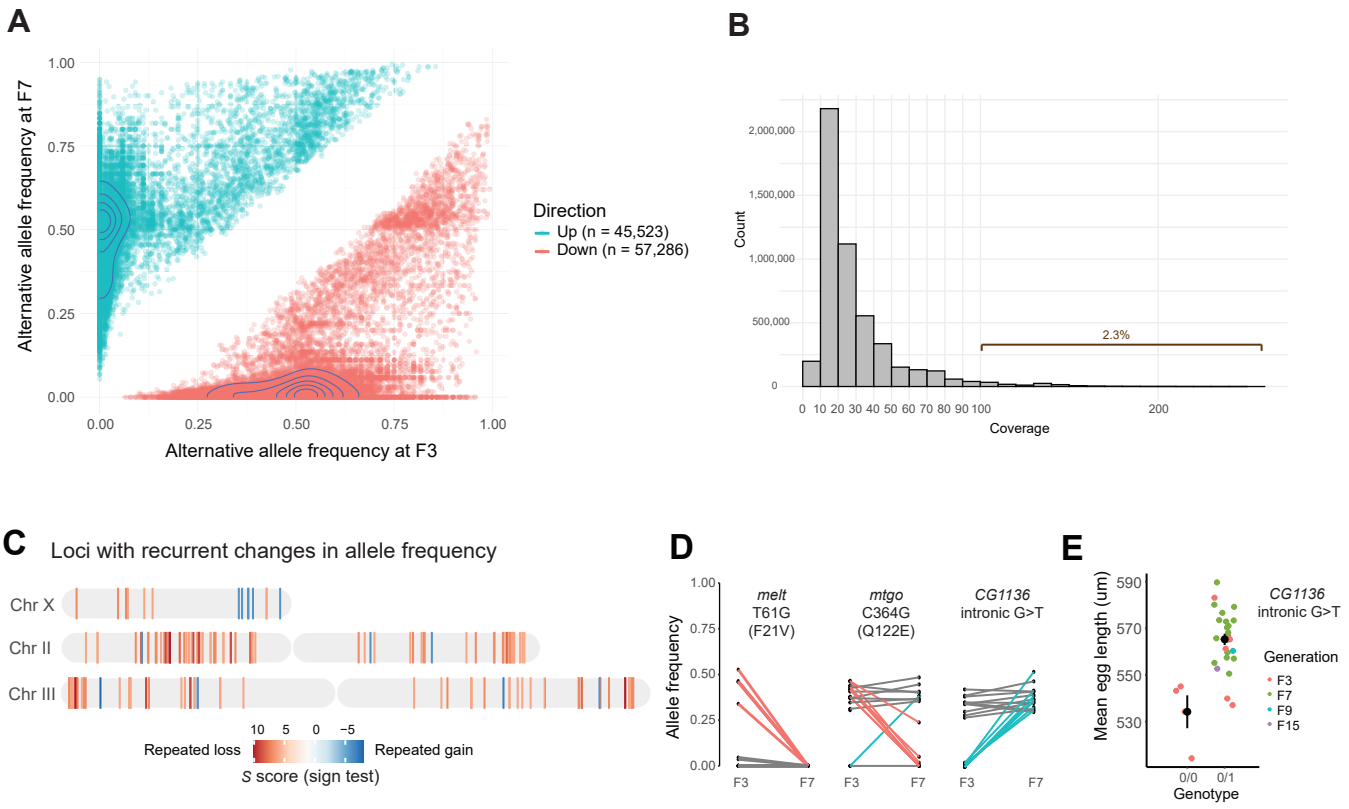


Fig. S4. Changes in allele frequency in evolved populations. (**A**) Alternative allele frequency at Generation 3 (F3) and Generation 7 (F7). Each point represents a comparison between two generations for a given variant in a given population. 102,809 significant changes on 54,045 sites across 18 populations (Fisher's exact test, FDR-adjusted $p < 0.05$) are shown. Colors represent the direction of changes (blue, increase in alternative allele frequency; red, decrease in alternative allele frequency). Contour lines mark high-density areas on the plot, showing that most changes were transitions between heterozygous (allele frequency around 0.5) and homozygous (allele frequency of 1) states. (**B**) Distribution of sequencing depth on all the variants surveyed with Fisher's test (450,739 variants across two generations and 18 populations). The median was 21 and the mean was 29. (**C**) Variants with consistent increased (blue) or decreased (red) allele frequency in six or more populations, with examples in (**D**). Colored lines indicate significant changes based on Fisher's exact test (red, decrease; blue, increase, FDR-adjusted $p<0.05$). (**E**) Genotype of the SNP in CG1136 was correlated with embryo length ($p=8.36e-05$). Black points represent the average phenotype of populations with the same genotype, and error bars represent standard errors.

Fig. S5.

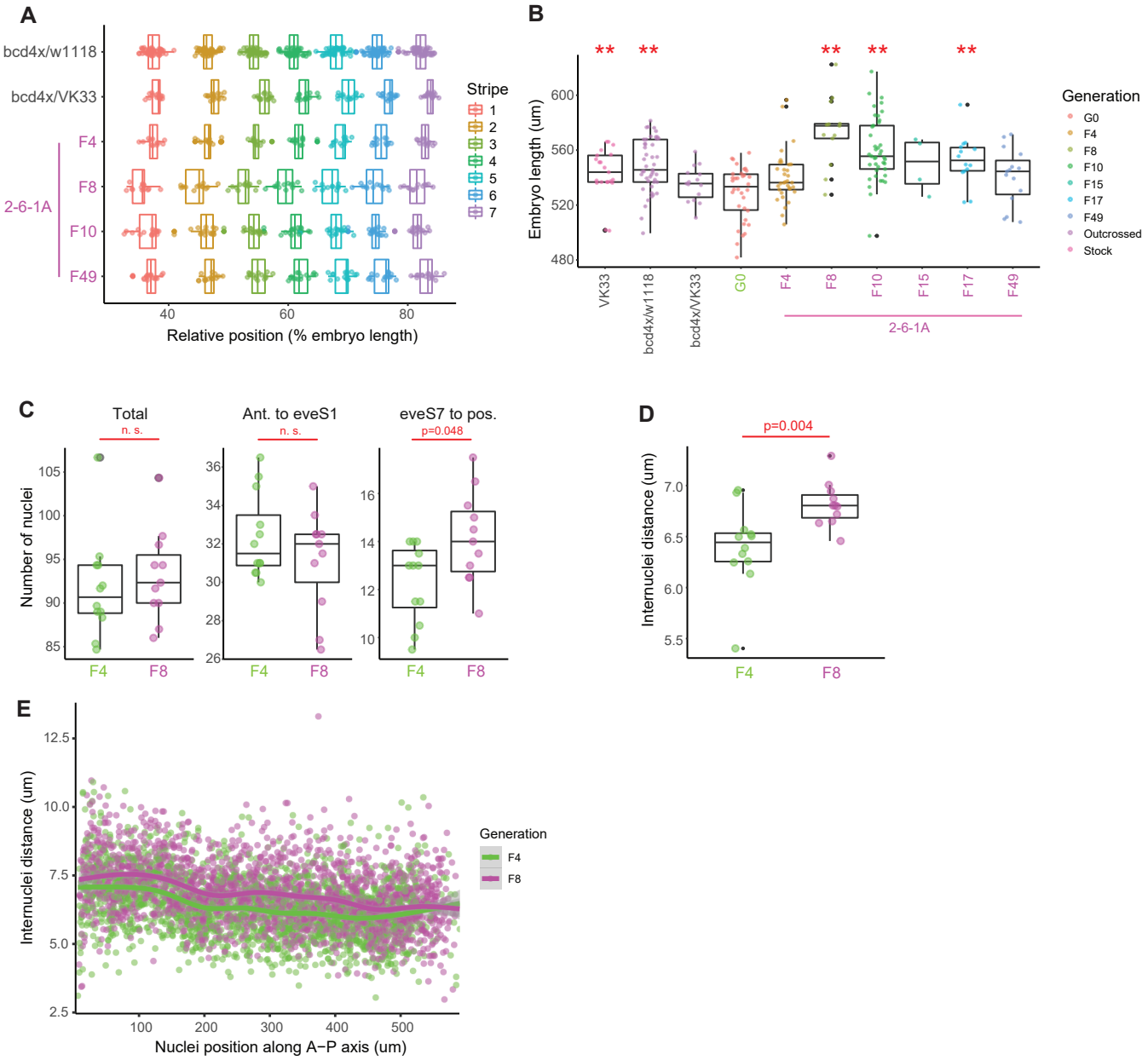


Fig. S5. Embryonic phenotypes of the evolved line 2-6-1A. *eve* stripe positions (**A**) and embryo length (**B**) changed dynamically across time. "bcd4x/w1118" and "bcd4x/VK33" are 4xbcd lines established from outcrosses between 2-6-1A (Generation 40) and wild-type w1118 or VK33, to represent a baseline phenotype without any evolutionary changes. Asterisks in (**B**) represent comparisons to G0 by Wilcoxon tests. (**C**) Number of nuclei along the *sna* border across the A-P axis. For each embryo, counts from two independent experimenters were averaged (the counts from the two experimenters were not significantly different). Particularly, the data of total number of nuclei were averaged across two measurements by X.C.L. and 1 measurement by L.G. Other data were only measured once by each experimenter. (**D**) Average internuclei distance. The data were averaged across all nuclei counted along the A-P axis and across two experimenters' measurements for each embryo. (**E**) Internuclei distance (the distance between the center of a given nuclei and that of its neighboring nuclei at the anterior side) as a function of nuclei position along the A-P axis, with data from both experimenters. Trend lines were generated with generalized additive mode smoothing ("gam" method), showing that F8 embryos had larger nuclei in the anterior and in the middle. P-values are from Wilcoxon tests. **, $p < 0.01$. n. s., not significant.

Fig. S6.

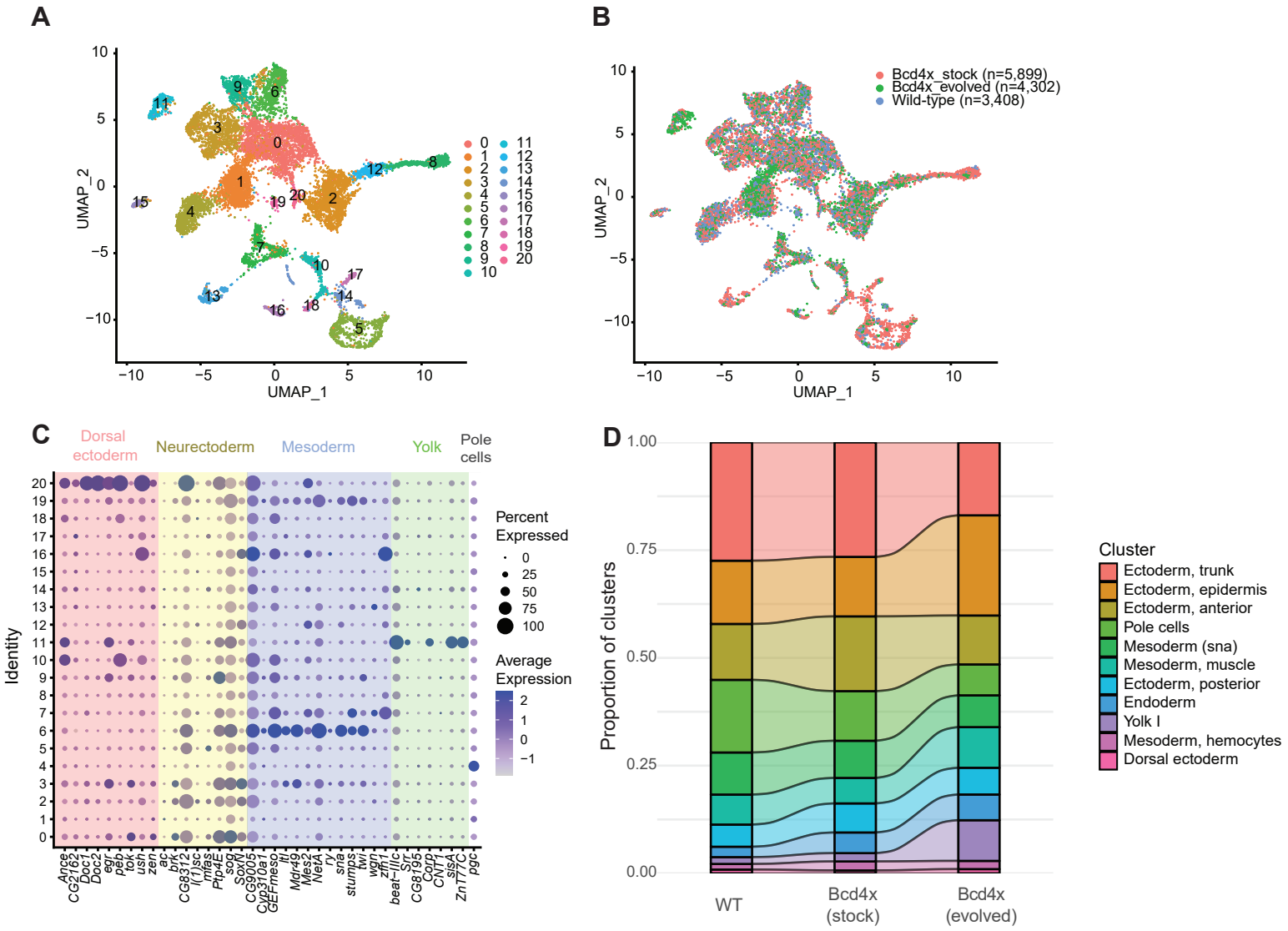


Fig. S6. Single-nuclei transcriptomes of the evolved line 2-6-1A (Generation 20). (**A-B**) UMAP of single-nuclei transcriptomes across three samples, with colors representing cell types in (**A**) and sample identity in (**B**). n represents the number of nuclei in each sample. Wild-type is VK33. (**C**) Dot plot showing expression of marker genes for primordium assignment used in Karaiskos, Wahle et al. 2017. (**D**) Proportion of clusters in three samples, only including early embryonic cell types (see **Table S2**).

Fig. S7.

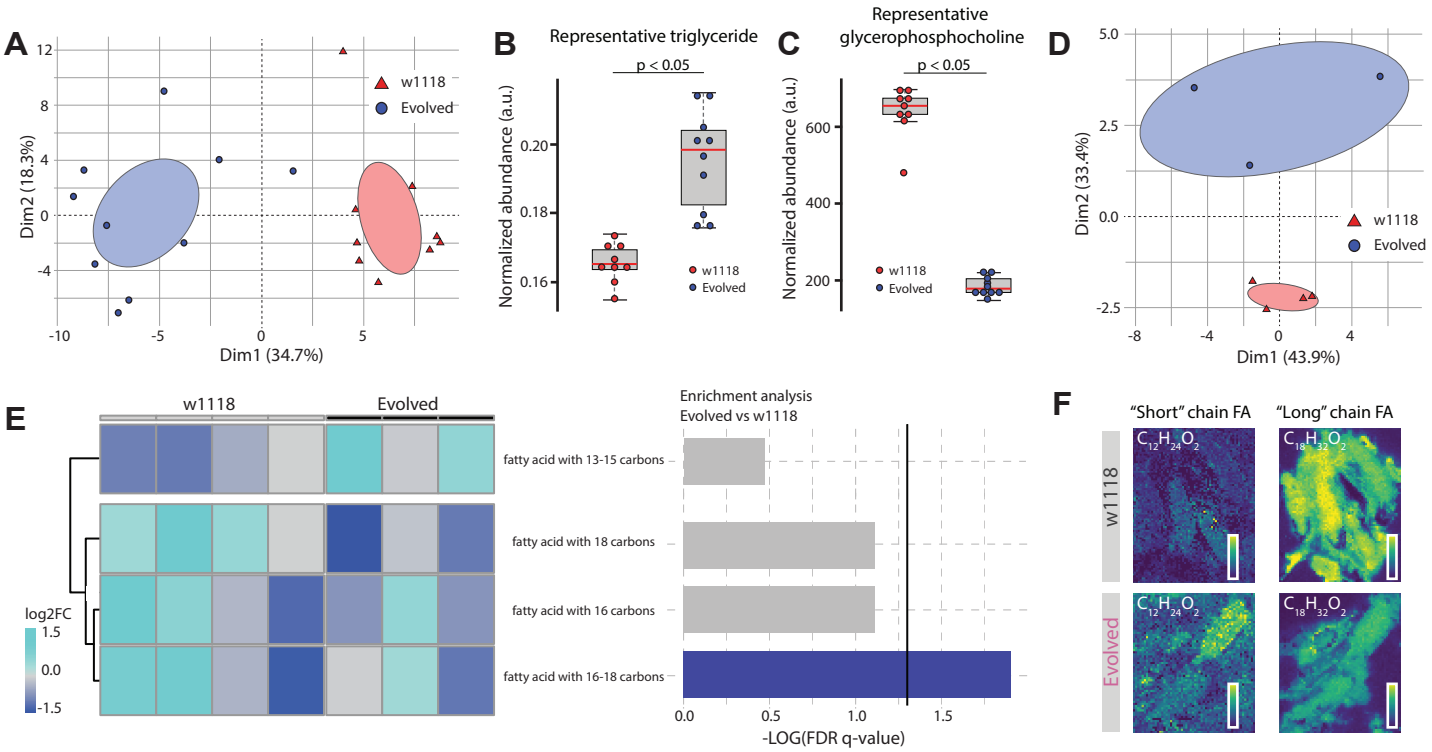


Fig. S7. Metabolic alterations in oocytes from the evolved line 2-6-1A. Related to **Fig. 3F-G**.
(**A**) Principal Components Analysis (PCA) based on single-oocyte abundance values for 122 different lipids identified by MALDI imaging mass spec. Each dot represents an individual oocyte. n=9 for w1118, n=10 for 2-6-1A. The ellipses delimit the 95% confidence interval.
(**B** & **C**) Abundance values of a representative triglyceride (**B**) and glycerophosphocholine (**C**) from single oocytes obtained through MALDI-imaging mass spec. Each dot represents an individual oocyte with n=9 for w1118 and n=10 for 2-6-1A. The boxed region is one s.d. and the tails are two s.d. (95 %). $p<0.05$, Student's T-test.
(**D**) Principal Components Analysis (PCA) based on single-oocyte abundance values for 21 different fatty acids identified through tandem mass spectrometry followed by MALDI imaging. Each dot represents an individual oocyte. n=4 for w1118, n=3 for 2-6-1A. The ellipses delimit the 95% confidence interval.
(**E**) Relative abundance of fatty acids (left panel, rows) across oocytes (left panel, columns) and their enrichment (right panel), based on 21 fatty acids detected through tandem mass spectrometry coupled with MALDI-imaging. The vertical line represents a threshold of q=0.05.
(**F**) Spatial distribution of a representative C12 (m/z= 199.17031; left) or C18 fatty acid (m/z= 269.24856; right) in sectioned ovaries from the w1118 and 2-6-1A lines.

Fig. S8

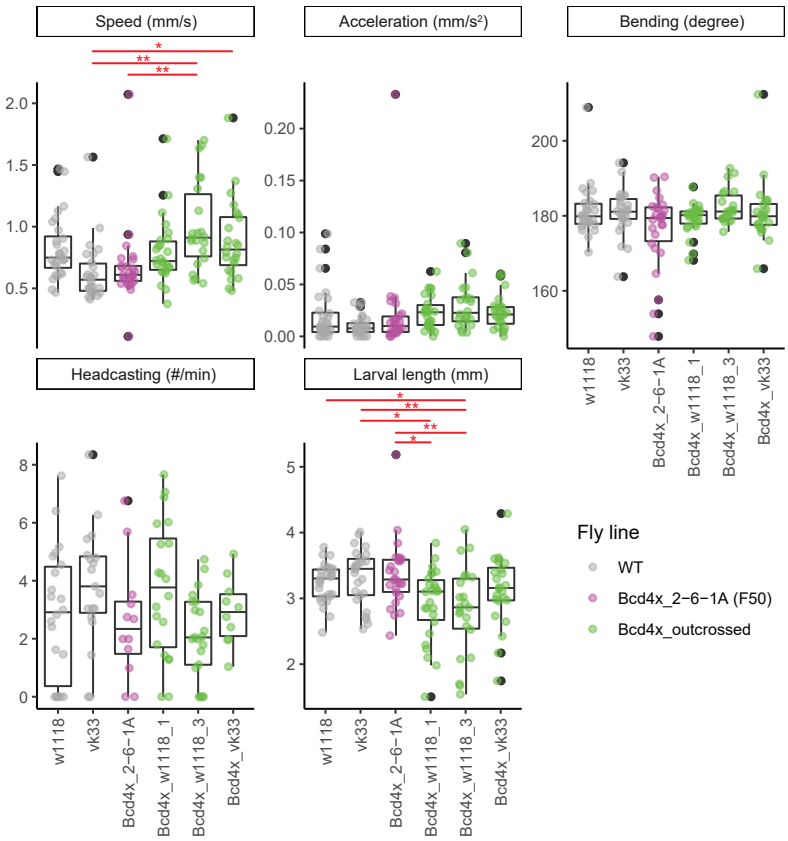


Fig. S8. Quantification of crawling behavior of 3rd-instar larvae from 2-min videos. Laval length was also extracted from the videos. The evolved line 2-6-1A is from Generation 50. Bcd4x_w1118_1 and Bcd4x_w1118_3 are two biological replicates of Bcd4x lines, outcrossed to the w1118 background. Bcd4x_vk33 is the bcd4x line outcrossed to VK33 background (see **Methods** and **Fig. S5** for details of the outcrosses). Asterisks indicate p-values from pair-wise comparisons using Tukey's test. *, $p < 0.05$. **, $p<0.01$. The evolved line did not show any consistent differences from the wild-type or the outcrossed lines.

Fig. S9.

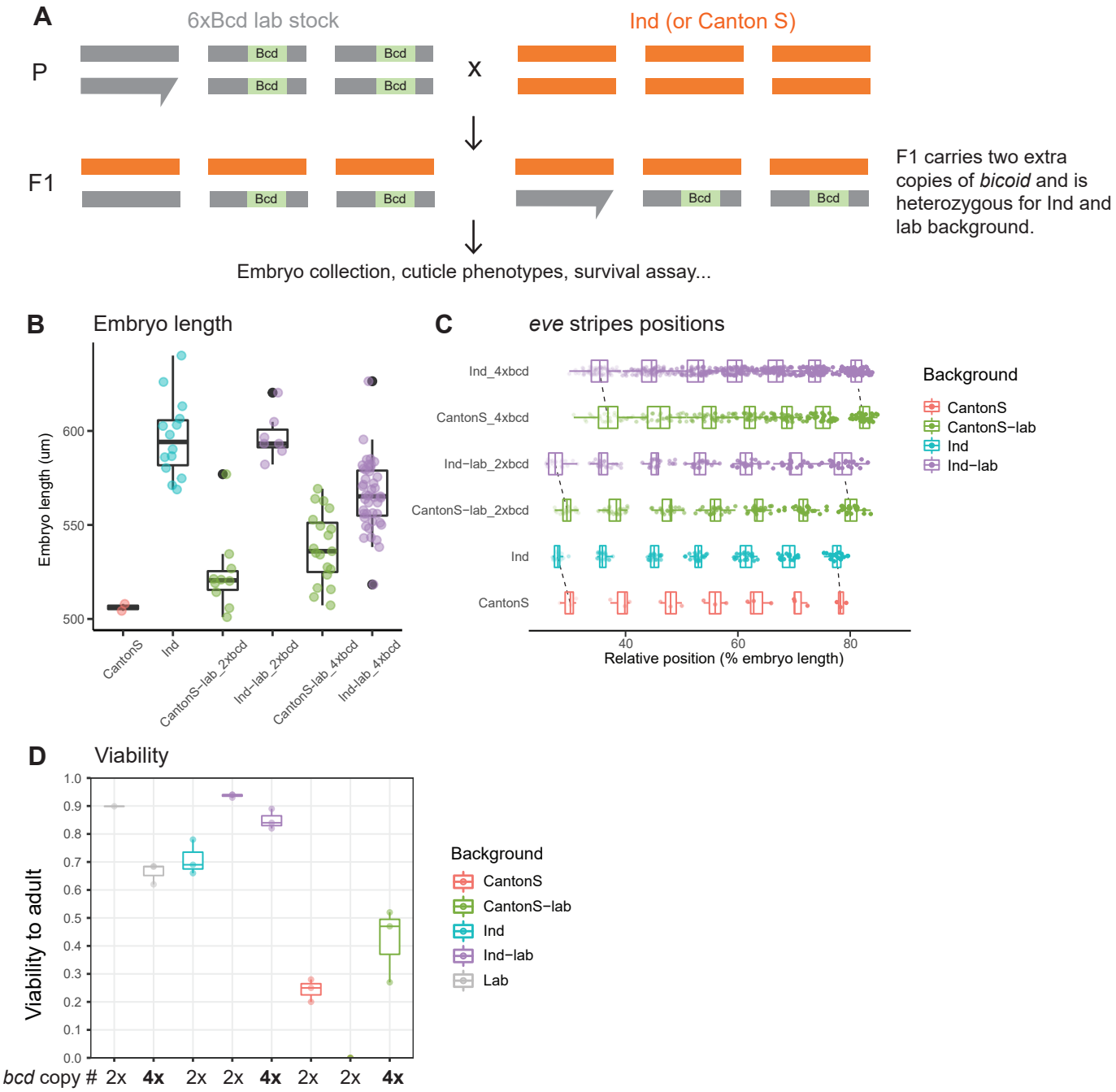
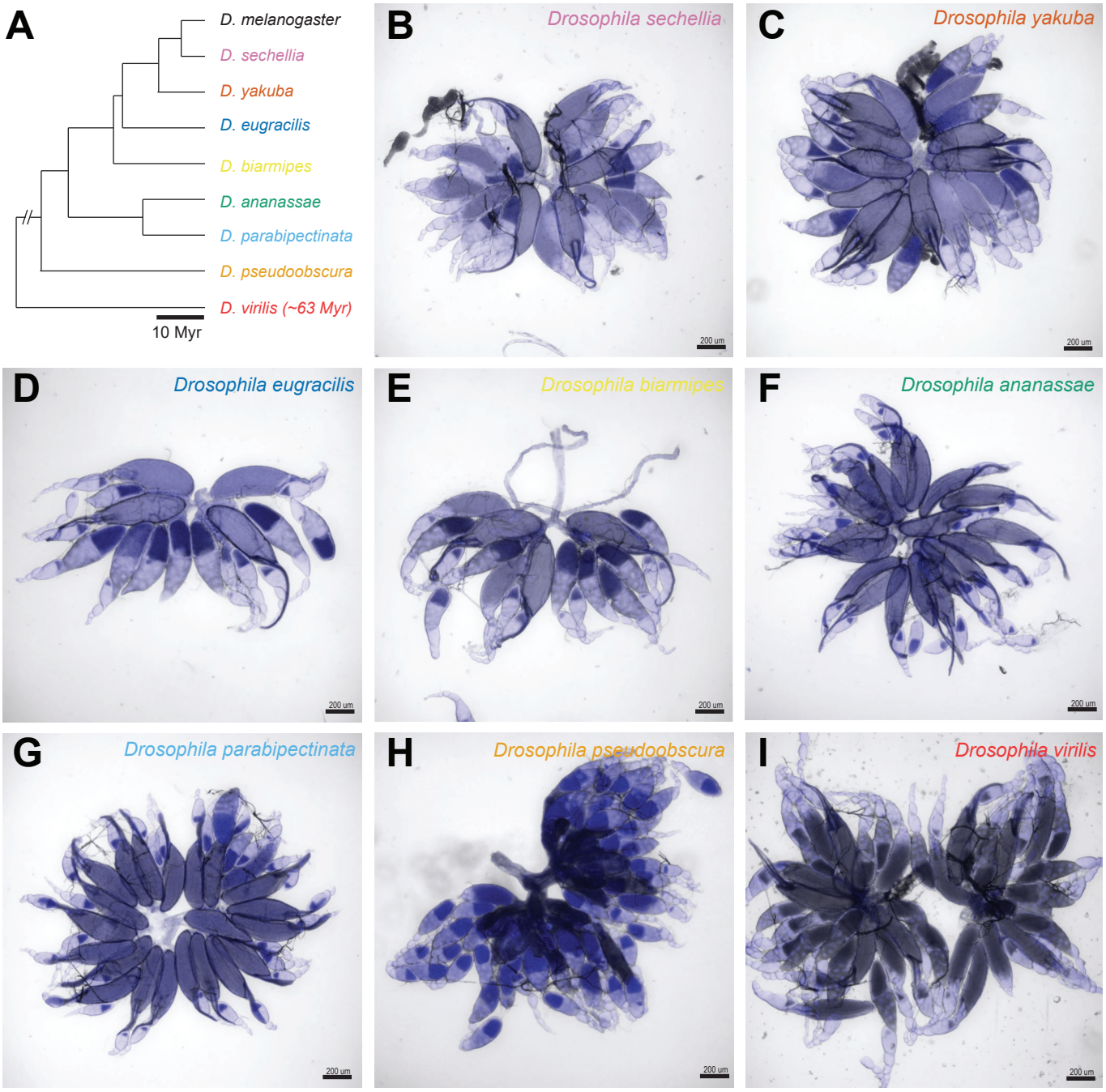


Fig. S9. Cross *bicoid* transgenes into Ind and Canton-S. (**A**) The crossing scheme. (**B**) Embryo length of Canton-S, Ind, and the F1 flies from crossing the two lines to a lab strain VK33 (CantonS-lab_2xbcd, Ind-lab_2xbcd) or a 6xBcd line (CantonS-lab_4xbcd, Ind-lab_4xbcd). (**C**) *eve* stripe positions, detected by *in situ* hybridization. The *eve* stripes of Ind-derived lines were located further to the anterior than those of Canton-S-derived lines under both *bicoid* dosages. (**D**) Viability to adult of the above-mentioned lines, along with the lab strains (from the data in **Table S1**, with the 4xBcd line in VK33 background). CantonS-lab control (2xbcd) embryos had an extremely low viability, probably due to genomic incompatibility between VK33 and Canton-S.

Fig. S10.



Extended Data Fig. 10. Ovaries in different *Drosophila* species. (**A**) Phlyogeny of the species explored in Fig. 5 with representative images (**G-I**).

Table S1. Viability of stocks carrying 2x-to-6x-*bicoid* (prior to selection).

Stock	Dosage of *bicoid*	Hatch rate (24h)	Viability to adult
w1118	2x	0.73	0.899
4xbcd_VK18	4x	0.68	0.853
4xbcd_VK33	4x	0.72	0.685
6xbcd	6x	0.49	0.389

Table S2. Marker genes and cell types for clusters in single nuclei RNA-seq.

Cluster	Selected top marker genes	Inferred cell type	Early embryonic cells (stage 4-6)?	% of total nuclei (across three samples)
0	*lov, opa, hth*	Ectoderm, trunk	Y	15.6
1	*Ribosomal genes*	Cytosolic debris		14.4
2	*ab, ds, ovo, heph, aop, ft, Wdr62*	Ectoderm, related to epidermis	Y	10.9
3	*oc, knrl, Lim1, jigr1*	Ectoderm, anterior	Y	9.3
4	*pgc, exu, gnu, nos*	Pole cells	Y	7.6
5	*nahoda, verm, exp*	Tracheal system		6.0
6	*sna, GEFmeso, twi*	Mesoderm	Y	5.6
7	*CadN, zfh1, Him, meso18E, tin*	Mesoderm, related to muscle	Y	4.7
8	*Rdl, para, nAChRalpha6, dysc, slo*	Neuron		4.6
9	*byn, fkh, tll, disco*	Ectoderm, posterior	Y	4.0
10	*peb, GATAe*	Endoderm	Y	2.9
11	*ZnT77C, sisA, beat-IIIc*	Yolk I	Y	2.6
12	*scrt, wor, insc, nerfin-1, ase, insb, stan, elav*	Central nervous system		2.1
13	*bt, Mhc, sls*	Muscle		2.1
14	*AhcyL2, su(r), Gs2, CG9932, Apoltp, Tps1*	Yolk II		2.0
15	*bcd, CG14229, CG6418, TfIIFbeta, GCC185, GFP*	Unknown		1.2
16	*Pxn, CG4829, Sr-Cl, ham, vri, gcm2, ct, pnr, RhoL, Pvr, zfh1, gcm, C1GalTA*	Mesoderm, precursors of hemocytes and lymph glands	Y	1.2

17	*prc, Apoltp, Hayan, Tig, CG10026, fon, daw, vkg, CG9932, AkhR, CG7149, svp, CG2736*	Fat body		1.1
18	*CG4839, CG11453, CG8907, CG8925, Vha100-5, CG4301, Drip, wdp, Traf-like*	Intestine and Malpighian tubules		0.9
19	*svp, CG15646, CG10663, HGTX, CG1673, CG7191, htl, ham*	Mesoderm, related to glia development		0.8
20	*C15, pnr, Doc2, peb, Doc3, tup, ush, Doc1*	Dorsal ectoderm, amnioserosa	Y	0.5