The mechanistic basis of genetic assimilation in natural fly populations

Giacomo Cavalli

Giacomo.Cavalli@igh.cnrs.fr

Institute of Human Genetics https://orcid.org/0000-0003-3709-3469

Gonzalo Sabarís

Institut de Génétique Humaine https://orcid.org/0000-0002-7797-6824

Bernd Schuettengruber

Institute of Human Genetics

Giorgio Papadopoulos

Institut de Génétique Humaine

Marta Coronado-Zamora

Institute of Evolutionary Biology, CSIC, UPF

Maximilian Fitz-James

Department of Biochemistry, University of Oxford https://orcid.org/0000-0002-6084-5887

Josefa González

Institute of Evolutionary Biology, CSIC, UPF https://orcid.org/0000-0001-9824-027X

Article

Keywords:

Posted Date: May 10th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4258726/v1

Additional Declarations: There is **NO** Competing Interest.

The mechanistic basis of genetic assimilation in natural fly populations

Authors

Gonzalo Sabarís[1], Bernd Schuettengruber[1], Giorgio L. Papadopoulos[1], Marta Coronado-Zamora[2], Maximilian H. Fitz-James[1], Josefa González[2], Giacomo Cavalli[1#]

Affiliations

[1] Institute of Human Genetics, CNRS, University of Montpellier, Montpellier, France

[2] Institute of Evolutionary Biology, CSIC, UPF, Barcelona, Spain

[#]corresponding author: giacomo.cavalli@igh.cnrs.fr

Keywords: Waddington, genetic assimilation, standing genetic variation, epigenetic inheritance, stress-induced variation, transposable elements, phenocopy, ectopic veins, *Drosophila*.

Summary

Genetic assimilation is a process by which a trait originally driven by the environment becomes independent of the initial cue and is expressed constitutively in a population. More than seven decades have passed since Waddington's pioneering demonstration of the acquisition of morphological traits through genetic assimilation, but the underlying mechanism remains unknown. Here, we address this gap by performing a multi-omic analysis of Waddington's genetic assimilation using the ectopic veins phenocopy in *Drosophila* as a model. Our study reveals the assimilation of ectopic veins in both outbred and inbred fly natural populations, despite their initially limited genetic diversity. The assimilation of ectopic veins is driven by the selection of regulatory alleles already present in the ancestral populations, including downregulation of the *Cad96Ca* tyrosine kinase gene by the insertion of a transposable element in its 3' untranslated region. The genetic variation at this locus in the inbred population is maintained by a large chromosomal inversion. In outbred populations, the evolution of ectopic veins results from a polygenic response shaped by the selective environment. Our results support a model in which selection for multiple pre-existing alleles in the ancestral population, rather than stress-induced genetic or epigenetic variation, drives the evolution of ectopic veins in natural fly populations.

Main Text

The origin of phenotypic novelty constitutes a central focus of evolutionary research. Environmental stresses can perturb normal development, inducing phenocopies — phenotypes that occur without apparent changes to the DNA but which replicate those generated by known genetic alleles[1]. Notably, heat shock (HS) applied during the pupal stage of *Drosophila* species induces various phenocopies in the wing venation pattern, including the loss of crossveins ("crossveinless") or the appearance of ectopic veins[2–4]. In a landmark study, Conrad Waddington demonstrated that the HS-induced crossveinless phenocopy can increase in frequency within a population through cycles of stress and selection over multiple generations, eventually becoming constitutive in the absence of stress[3]. This phenomenon, which he defined as genetic assimilation, is envisioned to facilitate rapid phenotypic evolution under the plasticity-first hypothesis[5,6].

Since Waddington's pioneering work, genetic assimilation has been generalized in laboratory experiments for several other *Drosophila* traits, from various wing venation phenocopies including ectopic veins[4,7], bristle, eye color and melanotic nodule mutant phenotypes[8], to more extreme traits such as haltere to wing homeotic transformations by ether exposure[9,10]. Genetic assimilation has also been studied in the laboratory for the evolution of polyphenism in the tobacco hornworm[11] and for the evolution of wing color in butterflies[12]. There is ample evidence suggesting that genetic assimilation also occurs in nature[13–16], and mathematical modelling has been used to understand its dynamics in populations[17–19]. Despite this wealth of evidence for genetic assimilation, there is still heated debate about its importance in evolution and the underlying mechanism[13,20,21]. Ironically, while Waddington hypothesized that genetic assimilation had a genetic basis, his experiments are often seen as evidence for an epigenetic component to heritability, arguing for the inheritance of acquired traits. He rather proposed that genetic assimilation occurs through the accumulation of rare genetic variants present at low frequencies in the ancestral population. Under normal developmental conditions, this genetic variation would not induce phenotypes (and could therefore be defined as "cryptic")[22] due to developmental canalization that funnels the phenotype closer to the norm, tending towards a standard phenotype. Stress might act as a catalyst, overcoming canalization and exposing this cryptic variation to selective pressure[5,23], until a point where it is expressed under

normal conditions, leading to genetic assimilation[24,25]. At the core of this model is the assumption that genetic assimilation is a polygenic response that relies on a substantial amount of standing genetic variation in the ancestral population, although compelling direct evidence is missing. Furthermore, there has been renewed interest based on emerging evidence, suggesting that mechanisms such as epigenetic inheritance[26–28] or stress-induced variation via transposable elements[8] may be other sources of variation that play a role in the genetic assimilation process. To discriminate between epigenetic regulation, polygenic responses or stress-induced mutation is a challenge of extraordinary difficulty because the available experimental paradigms that have been established involve *in vivo* studies on populations and it is difficult to separate genetic versus non-genetic contributions in these systems. Studies in laboratory models are needed as they allow a more direct exploration of the mechanistic basis of this phenomenon.

Results

Rapid assimilation of the ectopic veins trait in outbred and inbred natural fly populations

To test for phenotypic assimilation, we used four *wild type* inbred fly populations (D208, D437, D820 and D907) from the *Drosophila Genetic Reference Panel* (*DGRP*) library[29] and a *wild type* outbred fly population (called Mix) obtained from the genetic admixture of the four inbred lines into a single population (Figure 1a and Methods). We chose these lines and the ectopic veins (EV) phenocopy because they showed higher phenocopy induction upon pupal heat shock than common laboratory strains (Extended Data Fig. 1a-b). To recapitulate the Waddington's selection experiment, we generated "heat shock (HS) selection" lines by giving a heat-shock at the pupal stage and selecting flies hatching with ectopic veins (EV) phenotypes (Figure 1a). We assessed ectopic veins assimilation by measuring its penetrance in flies derived from the HS selection but reared under normal conditions, without heat shock, in the last generation (Figure 1a: "assimilated flies"). These lines were compared with control (C) flies reared under normal conditions for the same number of generations. Once the ectopic veins were found to persist in the assimilated flies, i.e. in the absence of stress, the EV phenotype was further selected in the absence of heat shock for several generations in the "assimilated (A) selection" (Figure 1a) lines. To distinguish between the phenotypic changes induced by heat shock and those that arise spontaneously under normal

conditions, we also selected for the ectopic vein phenotype in the absence of heat shock in lines derived from the control line, which we designated "non-assimilated (NA) selection" (Figure 1a) in order to be consistent with Bateman's nomenclature[4]. This experimental design allowed us to assess the response to EV artificial selection in the presence or absence of stress treatment, and to analyze the molecular mechanism underlying EV evolution in each regime.

In the Mix outbred population, the EV phenocopy was induced by the heat shock treatment from the first generation, but increased only when a higher selective pressure was applied after the seventh generation by intercrossing only flies with the strongest EV phenocopies (Figure 1b). Similarly, the EV trait persisted in the assimilated flies from the eighth generation onwards, becoming more penetrant with additional generations (Figure 1b and Extended Data Fig. 2a). Further selection of the assimilated flies in the absence of heat shock continued to increase the EV penetrance until eventual fixation (Figure 1c). The Mix outbred population also evolved the EV trait in the absence of stress treatment in the non-assimilated selection line (Figure 1d). Consistent results were obtained in a second independent selection experiment, where a strong EV selection pressure was applied from the start and the EV phenocopy reached almost full penetrance in the population in only three generations under heat shock selection (Extended Data Fig. 2b) and was assimilated after a return to normal conditions (Extended Data Fig. 2c-d). Likewise, the EV trait evolution in the non-assimilated selection without heat shock was also reproducible (Extended Data Fig. 2e).

Of the four inbred lines tested, only one (D907) assimilated the EV trait in the population. The EV phenocopy was induced by heat shock in each generation, however in this case it did not increase as generations progressed (Figure 1e). The EV trait was assimilated after the eighth generation (Figure 1e and Extended Data Fig. 2f) and further selection of the assimilated flies in the absence of heat shock continued to increase the EV penetrance to high levels (Figure 1f). Strikingly, the EV trait also responded to selection without stress in the non-assimilated selection line (Figure 1g). In contrast, the D208 inbred line never assimilated the EV phenocopy despite several generations of phenocopy induction and selection under heat shock (Extended Data Fig. 2g) and further selection in the absence of stress in the assimilated lines (Extended Data Fig. 2h).

Finally, repeated stress affected the viability in the inbred lines D438 and D820 and they became extinct after three generations of heat shock treatment (Extended Data Fig. 2i).

In summary, the EV trait was assimilated in both the Mix outbred population and the D907 inbred population but not in the other inbred lines tested. Notably, the phenotype was stronger in the Mix lines, where it is fully penetrant and the extra veins branch from multiple locations on the wing (Figure 1h). In the D907 lines, the EV phenotype is highly but not fully penetrant and the extra veins are short and arise mainly from a single site on the wing (Figure 1h).

The evolution of the ectopic veins in the Mix selection lines is driven by changes in the expression of multiple genes

Changes in the expression patterns of developmental genes frequently underlie morphological evolution[30,31]. In our efforts to delineate the genetic determinants involved in the evolution of the EV trait within the Mix selection lines, we performed transcriptomic analysis at two developmental time points relevant for wing vein development: the third instar larval wing disc (WD) and the pupal wing (PW) (24-26h after puparium formation)[32] (Figure 2a). We identified differentially expressed genes (DEGs) for the Mix selection lines relative to control flies in WD and PW (Extended Data Fig. 3 and Extended Data Table 1) and considered only those that were consistently deregulated between the two selection experiments for each line (Extended Data Fig. 4a-g and Extended Data Table 1a-b). We then shortlisted these candidate genes, some of which were shared between Mix assimilated and non-assimilated lines and others that were specific to one line (Figure 2a and Extended Data Fig. 4c,g), based on their functional annotation related to wing vein development (FlyBase). We also considered uncharacterized genes that were functionally related to other candidate genes (GeneMANIA)[33] and/or had a high level of deregulation.

Among the candidate genes, the tyrosine kinase receptor gene *Cad96Ca* was consistently downregulated in the WD and PW of Mix non-assimilated in both selection experiments and in a single experiment for Mix assimilated (Extended Data Figure 4c,g), with expression reduced close to the fold change cut-off for significance in the second experiments. We confirmed *Cad96Ca* downregulation in both experiments in the wing discs of Mix assimilated lines by RT-qPCR

(Extended Data Fig. 4d). Since *Cad96Ca* as well as most other selected candidate genes were downregulated (with the exception of *CG4250*), we performed gene knockdowns of these genes in the wings using the *nubbin-GAL4* driver to test their function. Among the common deregulated genes for both selection lines, knockdowns of *Cad96Ca*, *CG4250* and *Pxd* resulted in ectopic vein phenotypes or, in the case of *Pxd*, even more severe wing mutant phenotypes (Figure 2b,d). Knockdown of the *tx* gene resulted in pupal lethality. As for the Mix assimilated specific candidate genes, knockdown of *CG5144* caused a significant increase in ectopic veins, while *Vajk1* led to severe wing deformities (Figure 2b,d). Among the specific candidate genes in Mix non-assimilated, *CG8907*, *Cyp9c1* and *kibra* resulted in significant induction of ectopic veins (Figure 2b,d). Since *CG4250* was found upregulated, in this case we also analyzed its overexpression in the wing and found that it led to a significant induction of ectopic veins over control flies (Figure 2c,e). Remarkably, the wing positions from which the extra veins arise in the gene knockdowns are very similar to the EV phenotypes in the Mix selection lines (Figure 1h). Nevertheless, the EV phenotypes in the Mix selection lines are stronger than in each individual gene knockdown, suggesting that the EV phenotype in the selection lines arises from the additive effect of the deregulation of multiple genes.

Loss of a *cis*-regulatory element contributes to EV evolution

The changes in gene expression underlying EV evolution suggest that *cis*-regulatory elements controlling the expression of the candidate genes may have gained or lost activity in the Mix selection lines. To identify these changes, we performed CUT&RUN assays to analyze the differential distribution of the active enhancer mark H3K27ac between the selection lines[34]. The overall genome-wide enrichment of H3K27ac is remarkably similar between lines (Extended Data Fig. 5a), yet we found 24 differentially enriched H3K27ac peaks in Mix assimilated and 35 in non-assimilated compared to control flies (Extended Data Fig. 5b). We focused on those genes that are both associated with the differential H3K27ac peaks and that are consistently differentially expressed in each selection line. In Mix assimilated flies, the downregulation of the *SK* gene correlates with the loss of an H3K27ac peak close to its promoter. In Mix non-assimilated flies,

downregulation of the *Pxd*, *Ect3* and *CG31262* genes correlated with a loss of H3K27ac peaks at these loci (Figure 2f and Extended Data Fig. 5b-e).

The *SK* gene, which encodes a subunit of a potassium channel, and the *Ect3* gene, which is predicted to be involved in carbohydrate metabolism, are unlikely to be associated with EV evolution in the respective flies where they were found deregulated. On the other hand, the *Pxd* gene was downregulated in both Mix assimilated and non-assimilated lines and is located in the genome next to another candidate gene, *CG8907*, which was downregulated in Mix non-assimilated flies only. Both *Pxd* and *CG8907* were shown to be involved in EV assimilation in these flies, as their wing knockdowns resulted in ectopic veins and wing mutant phenotypes (Figure 2b,d). We hypothesized that the DNA sequences overlapping the H3K27ac peak in the first intron of the *Pxd* gene may act as *cis*-regulatory signals controlling the expression of the *Pxd* and/or *CG8907* genes, and may have evolved exclusively in Mix non-assimilated flies, leading to the downregulation of these genes in the wings. To test this hypothesis, we performed a CRISPR/Cas9 targeted deletion of a 480bp region overlapping the H3K27ac peak in the first intron of *Pxd* (*Pxd_i1_deletion*, Figure 2f). The deletion resulted in a significant induction of wing mutant phenotypes such as mild ectopic veins and other wing shape mutant phenotypes (Figure 2g). Remarkably, the wing regions from which the extra veins are induced by the deletion are very similar to the positions of the ectopic veins in the Mix selection lines (Figure 1h). We then analyzed whether the deletion caused changes in the expression levels of the *Pxd* and *CG8907* genes in the wings. We found that the deletion did not alter the expression levels of the *Pxd* gene, but significantly decreased the levels of *CG8907* in both the wing discs and pupal wing (Figure 2h). This suggests that there may be an enhancer in this DNA sequence that controls *CG8907* expression in wing discs and pupal wings and has lost its function, contributing to EV evolution, in Mix non-assimilated flies. Sequence analysis in this region did not identify any *cis*-mutations exclusively associated with the Mix non-assimilated lines, suggesting that the loss of *CG8907* enhancer activity might be caused by changes in the function of *trans*-acting factors acting via this element (Extended Data Fig. 6).

The evolution of ectopic veins in the D907 selection lines is explained by changes in the expression of a few shared genes

We performed transcriptomic analysis in wing discs and pupal wings of the D907 selection lines to identify putative genes involved in EV evolution. In contrast to the Mix lines, we found a small number of differentially expressed genes (DEGs) for assimilated and non-assimilated lines relative to the control in WD and PW (Figure 3a, Extended Data Fig. 7a and Extended Data Table 2). The *Cad96Ca* gene, which was downregulated in the wing discs of all Mix selection lines, was also found to be downregulated in the wing discs of both D907 selection lines. *Cad96Ca* knockdown in the wing disc induced ectopic veins arising near the L2 and L5 veins, which are remarkably similar to the ectopic vein phenotypes of the D907 selection lines (Figure 3b). These results position the downregulation of the *Cad96Ca* gene as a major player in the evolution of the EV trait in both the Mix and D907 selection lines. Another candidate is the *Hsp83* gene, which was downregulated in the pupal wings of D907 assimilated flies. *Hsp83* also showed lower levels in D907 non-assimilated flies, although not significantly different from the control in RNA sequencing analysis, and we confirmed its downregulation in both D907 assimilated and non-assimilated flies by RT-qPCR (Extended Data Fig. 7b). Interestingly, *Hsp83* dysfunction using the loss-of-function allele $Hsp83^{e6A35}$ in heterozygotes caused ectopic vein phenotypes and knockdown in the wing disc resulted in deformed wings in females (Figure 3b) and lethality in males, confirming its function in wing development.

A large chromosomal inversion maintains the genetic variation associated with the EV evolution in the D907 inbred population

The observation of such a strong response to selection in an inbred line, with ostensibly limited genetic variability, was somewhat surprising[36]. To identify the genomic responses to EV selection, we performed Pool-seq experiments[37] from D907 derived (assimilated, non-assimilated, and control) and ancestral (P0) flies. We annotated single nucleotide polymorphisms (SNPs) and small inversions and deletions (InDels) for each fly population and analyzed significant frequency changes between the derived lines and the ancestral D907 population (Figure 1a). Strikingly, we found that most of the significant associated mutations in D907 assimilated and non-assimilated

flies occur in a large region of chromosome 3R (~14Mb in length), extending from the middle part of the chromosome arm toward the subtelomeric region (Figure 4a). There are no significantly associated alleles in the control flies in this region, suggesting that the associated mutations in D907 assimilated and non-assimilated flies are not caused by genetic drift.

Genetic variation can accumulate in DNA sequences spanned by large chromosomal inversions because homologous recombination during meiosis is greatly reduced, impairing genetic homogenization in this region[38]. We therefore hypothesized that this pattern of linked mutations, spanning several megabases of the genome, may be the consequence of selection for a large chromosomal inversion, and that the mutations are inherited as haplotypes due to limited recombination with the non-inverted chromosome. Interestingly, the paracentric chromosomal inversion *In(3R)C*, which is rarely found in wild cosmopolitan fly populations[39], but is one of the inversions of the TM balancer chromosome series (TM3, TM6 and TM6B)[40], overlaps with this region. *In(3R)C* was recently mapped by Hi-C assays, identifying the chromosome-proximal breakpoint site at the 20.3Mb and the subtelomeric breakpoint site at ~32Mb of chromosome 3R[41]. Therefore, we performed Hi-C experiments in the D907 lines to analyze whether *In(3R)C* was present in these populations. We found a significant enrichment of the interaction frequency at the breakpoint site for *In(3R)C* at 20.3Mb in chromosome 3R in D907 control flies, which is absent in D907 assimilated and non-assimilated flies (Figure 4b). This result suggests that *In(3R)C* segregated in the ancestral D907 population and was counter-selected, resulting in fixation of the standard chromosomal variant in D907 assimilated and non-assimilated lines. We estimated the frequency of *In(3R)C* in the D907 control and ancestral populations (23% and 30%, respectively) using as a proxy the average frequency of mutations that were counter-selected (frequency =0) in D907 A and NA in the region of chromosome 3R covered by the inversion (Figure 4b and Extended Data Fig. 8).

Segregating transposable element variation, rather than de novo variation induced by heat shock, contributes to the evolution of EV

It was previously proposed that transposable elements (TEs) could contribute to the apparent assimilation of an acquired character by providing *de novo* genetic variation induced by

heat shock[8]. We tested whether this phenomenon, or the selection of segregating TEs already present in the ancestral population, might contribute to EV evolution in the Mix and D907 selection lines. To this end, we used the Pool-seq data to annotate the TE insertions in the genomes of the ancestral and derived populations. There are equivalent numbers of total TEs (Figure 5a) and *de novo* TE insertions (TE mobilizations) (Figure 5b) in the genome of the D907 and Mix assimilated lines compared to the other populations. We reasoned that if heat shock-induced *de novo* TE insertions played a role in EV evolution, they should be highly penetrant in the assimilated selections. We found no high frequency (> 0.5) of *de novo* TE insertions in the assimilated selection lines D907 and Mix, suggesting that this was not the case (Figure 5c). Only five *de novo* inserted TEs were found in D907 NA flies and two in a single replicate of Mix NA, but none of them falls close (within a 1Kb window) to genes with known functions in wing vein development (Figure 5c and Extended Data Table 3b). Taken together, these results suggest that heat shock-induced variation through TE mobilization was not the cause of EV evolution in the Mix and D907 assimilated selection lines.

We then investigated whether TEs segregating in the ancestral populations were selected and contributed to the evolution of EV in the Mix and D907 selection lines. We searched for TEs that had significantly increased in frequency in the selected lines compared to the parental and control populations (see Methods). We analyzed the overlap between the significantly enriched TE-associated genes and the candidate genes identified in the transcriptome analysis for each selection line (Extended Data Table 3c-d). Among all the selection lines, we found twelve genes whose deregulation correlates with a nearby TE insertion that has significantly increased in frequency in the respective selection. Remarkably, among these genes we found a TE of the *I-element* family inserted into the 3'UTR of the *Cad96Ca* gene (*Cad96Ca*$^{[I\text{-}element]}$), which was significantly associated in all selected lines (Figure 5d). *Cad96Ca*$^{[I\text{-}element]}$ was present at intermediate penetrance in the D907, and at much lower frequency in the Mix parental and control populations. Notably, it increased significantly in frequency in all D907 and Mix assimilated and non-assimilated selections, showing a strong signature of selection in all EV selection lines (Figure 5e). This signature of selection of *Cad96Ca*$^{[I\text{-}element]}$ correlated with the downregulation of *Cad96Ca*

in all D907 and Mix selection lines, strongly suggesting that the *I-element* insertion may play a role in the downregulation of *Cad96Ca*.

Transposable elements inserted in euchromatic regions can affect the expression of nearby genes by the spreading of repressive heterochromatin marks (H3K9me3)[42] or by post-transcriptional mechanisms[43]. We found no differential enrichment of H3K9me3 in D907 assimilated or non-assimilated lines at the *Cad96Ca* locus in wing discs, where *Cad96Ca* is downregulated in these flies compared to control flies (Extended Data Fig. 9). This suggests that the *I-element* insertion does not affect *Cad96Ca* expression by altering the chromatin landscape at this locus. We then performed Rapid Amplification of cDNA Ends at the 3' end (3' RACE) PCR amplification specific for the *Cad96Ca* mRNA in the wing discs of D907 lines. We found three different mRNA isoforms of *Cad96Ca* in D907 control flies, of which the shorter one (Figure 5f, band C in the gel) corresponds to the *wild type* isoform and two larger ones (Figure 5f, bands A and B, respectively) correspond to isoforms containing the *I-element*. In D907 assimilated and non-assimilated flies, the *wild type* isoform of *Cad96Ca* mRNA was replaced by the larger isoforms containing the *I-element* in their 3'UTR (Figure 5f). Taken together, these results suggest that *Cad96Ca*$^{[I\text{-}element]}$ has been selected in D907 and Mix assimilated and non-assimilated selection lines and may contribute to dampening *Cad96Ca* expression by altering the 3'UTR mRNA composition, leading to its destabilization by a TE-dependent post-transcriptional mechanism[43].

Discussion

Despite decades of empirical evidence for genetic assimilation, its role in evolution remains a subject of debate, largely because its mechanistic basis is unclear[13,20,21]. It is generally assumed that genetic assimilation is contingent on a large amount of standing genetic variation in the initial population, as some inbred laboratory populations of *Drosophila* have failed to exhibit genetic assimilation[3,4,24]. Here, we demonstrate the assimilation of the ectopic veins trait in both an outbred (Mix) and an inbred (D907) fly population characterized by limited initial genetic diversity. Our analysis reveals standing genetic variation explaining EV assimilation in both inbred and outbred populations. The higher level of standing variation in the outbred Mix population leads to a polygenic response to selection, orchestrating changes in the expression of multiple genes.

Conversely, when lower levels of standing variation occur in the D907 inbred population, the evolution of EV is driven by the dysregulation of a few developmental genes.

Common regulatory changes underlie EV evolution in the D907 assimilated and non-assimilated selections, such as the downregulation of the *Cad96Ca* and *Hsp83* genes. In addition, the EV trait was never assimilated in the D208 inbred population, arguing against a prominent role for stress-induced genetic[8] or epigenetic variation[26,27]. The strong response to selection in the presence and absence of stress was a striking result, given that D907 is a highly inbred population with low expected levels of genetic variability, and it raised two key questions that needed to be addressed: i) what genetic variants underlie the evolution of ectopic veins in D907 flies, and ii) how is this genetic variation maintained within the inbred D907 fly population?

Downregulation of the *Cad96Ca* gene emerged as a major contributor to the EV evolution in both the D907 and Mix populations. *Cad96Ca* dysfunction was not previously associated with ectopic vein phenotypes, but functional analysis confirmed its role in wing vein development, possibly mediated by the *grainy head* pathway[44,45]. We identified a regulatory allele of the *Cad96Ca* gene, a transposable element inserted in its 3'UTR (*Cad96Ca*$^{[I\text{-}element]}$), which showed a strong signature of selection in all D907 and Mix selection lines. TE insertions in the 3'UTR of coding genes are associated with reduced levels of gene expression in flies[46], humans and mice[47], which may be mediated by a post-transcriptional TE silencing mechanism[43]. We found that the genetic variation spanning the *Cad96Ca* locus was maintained in the D907 ancestral population by the presence of the large chromosomal inversion *In(3R)C*, probably by preventing genetic homogenization within this chromosomal segment[38]. Counter-selection of *In(3R)C* in D907 assimilated and non-assimilated lines resulted in the fixation of *Cad96Ca*$^{[I\text{-}element]}$, contributing to EV evolution in these fly populations.

The Mix selection lines have evolved a much stronger ectopic vein phenotype than the D907 lines. Consistently, in addition to *Cad96Ca* downregulation, several other gene deregulations underlie EV evolution in the Mix lines. While some of these genes are shared between Mix assimilated and non-assimilated lines, others have evolved their expression exclusively for each line. A prominent example is the exclusive loss of a putative enhancer element controlling *CG8907* expression in the wing discs and pupal wings of Mix non-assimilated flies,

together with the different chromosomal contributions to the EV phenotype in each line (Extended Data Fig. 10). Notably, Bateman also observed different allelic contributions between assimilated and non-assimilated selection lines, although her interpretation assumed selection for identical segregating alleles in both scenarios[4]. In contrast, we interpret our results differently, whereby EV evolution occurred through fixation of distinct alleles in response to selection in the presence or absence of stress in Mix assimilated and non-assimilated lines, respectively. Certain alleles, such as *Cad96Ca*$^{[I\text{-}element]}$, have a significant effect on the EV trait, inducing expression even under non-stressful conditions. Selection without stress can increase its penetrance within the population, as seen in the non-assimilated lines D907 and Mix. In addition, selection under stress can shape trait evolution by exposing cryptic trait-modifying alleles (or masking others expressed under non-stressful conditions) to selective pressure[5,23], probably through a form of genotype by environment interaction[48]. Consequently, trait selection in different environments can lead to genetic canalization of the adaptive trait through alternative developmental pathways (Figure 6).

In conclusion, our study provides insights into the mechanistic basis of Waddington's genetic assimilation in natural animal populations. Our results support a model in which trait assimilation is driven by selection for multiple alleles present in the ancestral population, with no evidence for stress-induced genetic variation. To determine the role of epigenetic inheritance, we studied EV assimilation in genetically constrained populations and found that genetic variability was still contingent for EV evolution. This challenges the notion that epigenetic inheritance alone drives trait assimilation in inbred populations. However, in populations with larger genetic variance, distinguishing between epigenetic and genetic allelic contributions is challenging and remains to be investigated.

Extended Data

This manuscript includes ten Extended Data Figures and seven Extended Data Tables.

Methods

Drosophila strains and handling

Flies were reared in standard cornmeal-yeast extract medium at 25°C unless otherwise specified. For gene knockdowns in the wing, virgin females of the *nubbin-GAL4* driver (BDSC_25754) were outcrossed with the RNAi lines targeting the corresponding gene and more than 50 heterozygous progeny females were scored per line. RNAi against GFP gene (BDSC_9331) was used as control. For the knockdown of *Hsp83* genes, *nubbin-GAL4* driver caused pupal lethality so we used *A9-GAL4* instead (BDSC_8761). A list of all the fly lines and genotypes used is reported in Extended Data Table 4.

Recapitulation of Waddington selection experiment using ectopic veins trait

We used the inbred lines D208 (DGRP-208), D437 (DGRP-437), D820 (DGRP-820) and D907 (DGRP-907) belonging to the *Drosophila Genetic Reference Panel* (DGRP) library. These fly population lines were derived from a natural population and inbreeding through 20 generations of full-sib mating[29]. We also used an outbred fly population (Mix), which is consisted in the mixture of these four inbred populations into a single population. For that, we took 20-mated females from each inbred line and bred them together for five generations in large population size to allow for genetic admixture. The sixth generation was the parental population (P0) for the first replicate of the heat-shock and artificial selection experiment and the twenty-one generations for the second replicate. We performed the pupal heat shock treatment by incubating pupae (0-27h after puparium formation) for 4h at 40°C. This pupal period is sensitive for wing phenocopies induction (Extended Data Fig. 1A), and resistant to pupal lethality after stress[2]. After the heat-shock, flies continued their life cycle at 25°C. Adult flies were collected during the first two days of adult emergency from the pupal case for scoring of ectopic veins phenocopy. Phenocopies induction can differs between sexes[4], being the ectopic veins phenocopy slightly more frequent in females than in males. We always reported the EV in females to avoid misinterpretation of the results. Control fly lines for each population were maintained in large population sizes in normal conditions at 25°C without selection. Non-assimilated selections started from control fly populations at the matched generation of the assimilated selection lines. The number of flies scored per generation for each selection line, together with details of the statistical analysis, are reported in Extended Data Table 5.

Imaging of fly adult wings

Adults were collected and frozen until used. Adult wings were dissected and mounted in a microscope slide with a drop of Euparal mounting medium. After overnight drying at 50°C, the wings were imaged with bright-field microscopy. The images were processed using Adobe Photoshop.

Ectopic veins expressivity score

Ectopic vein scoring has been always performed on females to maintain consistency between experiments. The collected adult females were frozen until analyzed. We implemented a scoring system to analyze the expressivity of the ectopic veins as follows. We analyzed both wings per individual, each wing can have an EV score from 0 (no EV) to 4, with a maximum EV score of 8 per fly. Score 1: Mild ectopic vein or thickening of a vein in a single spot on the wing. Score 2: Mild ectopic veins arising from two different areas of the wing. Score 3: At least two regions of the wing with ectopic veins and at least one of these with ramifications. Score 4: At least three regions of the wing with ectopic veins and at least one with ramifications.

Chromosomal replacements

Females of the Mix assimilated and non-assimilated lines were outcrossed to males of a double balancer line for the second and third chromosomes (*+/+; CyO/Sp; TM6B/Sb*). Heterozygous F1 progeny of males carrying the *Cyo* and *TM6B* balancer chromosomes and the *Sp* and *Sb* chromosomes were selected and backcrossed to the corresponding females of the selection line, and the F2 progeny were analyzed. This F2 progeny consisted of four different genotypes for both balancer and mutant chromosome backcrosses: i. the parental genotype (selection line) restoration; ii. heterozygous replacement of the second chromosome; iii. heterozygous replacement of the third chromosome; iv. double heterozygous replacement of the second and third chromosomes. At least 50 females of each F2 genotype combination were analyzed for ectopic vein penetrance and expressivity.

RNA-seq experiments

Third instar larvae and prepupae were collected and sexed, with only females selected for RNA preparation. 20 third-instar larval wing discs and 15-20 pupal wings (24-27h APF) per replicate were rapidly dissected in Schneider medium on ice. We performed four biological replicates per condition (selection lines). Total RNA was extracted with TRIzol reagent and purified using the *RNA Clean & Concentrator™-5 kit* (*Zymo Research*, cat. R1015) according to the manufacturer's instructions and DNase I treatment (*QIAGEN*, cat. 79254). Purified RNA was quantified on a NanoDrop equipment and >1µg of total RNA was sent to Novogene (https://en.novogene.com/) for strand-specific poly-A enrichment mRNA-seq library preparation. Paired-end sequencing of the libraries was performed on Novogene's NovaSeq 6000 PE150 platform.

Transcriptome analysis

RNA-seq data quality was assessed using FastQC (v 0.12.1). Stranded RNA-seq data were mapped to the *Drosophila melanogaster* dm6 genome using STAR (v 2.7.0) with default parameters. Aligned sequencing reads mapped to gene transcripts were counted using featureCounts (Subread v 2.0.6) with –s 2 (reverse stranded) and default parameters. The gene transcript annotation file was obtained from FlyBase (release 6.34). Prior to statistical analysis, genes with fewer than 10 reads (cumulating all samples analyzed) were removed. Differentially expressed genes (DEGs) were identified using the DESeq2 R package[50]. Genes with adjusted p-value<0.05 (using the Benjamini-Hochberg FDR method) and |log2FC|>0.58 were considered differentially expressed. Volcano plots were generated using the "EnhancedVolcano" R package (DOI: 10.18129/B9.bioc.EnhancedVolcano).

Reverse transcription quantitative PCR experiments (RT-qPCR)

About 20 third-instar larval wing discs and 15-20 pupal wings (24-27h APF) from female flies were rapidly dissected in Schneider medium on ice. Total RNA was extracted using the same procedure as described above for RNA-seq experiments. The purified RNA was quantified on a NanoDrop and a minimum of 250ng of total RNA was used for reverse transcription using the *SuperScript IV Reverse Transcriptase Kit* (*Thermo Fisher Scientific*) according to the manufacturer's

recommendations. Finally, quantitative PCR was performed on a *LightCycler480* instrument using SYBR Green I Master Mix (*Roche*). PCR primers are listed in Extended Data Table 6.

CRISPR/Cas9 genome editing

To generate the targeted deletion of the H3K27ac peak in the first intron of *Pxd* (*Pxd_i1_deletion*), we used the stock *nos-Cas9* attP40 (*y,sc,v; nos-Cas9/CyO; +/+*) as the recipient strain. Microinjection was performed using a two-guide RNA strategy designed (gRNA1: 5'-GTGAGATCGACCGACAAAGACGG-3', gRNA2: 5'-CCACTGGTCACCTAGAGAAGTGG-3'), PCR screened (fw: 5'-CCCGCCATTCACCTGGTGGTCT-3', rv: 5'-TTGTTTAATTCGCTCAGGTAATTGC-3'), and implemented by *Rainbowgene Transgenic Flies Inc*. Deletion phenotypic effects were compared with the recipient fly strain as a control.

CUT&RUN assays

CUT&RUN assays were performed according to Kami Ahmad's protocol implemented for *Drosophila* tissues (https://dx.doi.org/10.17504/protocols.io.umfeu3n) with minor modifications. 20 female third-instar larval wing discs were dissected in Schneider medium at room temperature, centrifuged for 3 min at 700g and washed twice with wash+ buffer before addition of Concanavalin A-coated beads. MNase digestion (pAG-MNase Enzyme from *Cell Signaling*) was performed for 30 min on ice. After ProteinaseK digestion, DNA was recovered using *SPRIselect beads* and eluted in 0.1X 50µl TE. DNA libraries were prepared using the *NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB)* following instructions, but with the following changes: (i) adaptors were diluted 1:10 in water for adaptor ligation (step 2), (ii) the size selection of the adaptor-ligated DNA in step 3A was omitted (we proceeded directly to step 3B) and (iii) we performed 14 cycles of PCR with 10 seconds of annealing/extension for enrichment of short DNA fragments. Libraries were sequenced on a NovaSeq 6000 system with 150 bp PE reads by Novogene. We performed three biological replicates for the histone H3K27ac antibody (*Active Motif,* Cat. 39134) for each Mix selection lines, two replicates for the histone H3K9me3 antibody (*Abcam,* Cat. AB8898) and one control replicate for the Normal Rabbit IgG antibody (*Cell Signaling,* Cat. 2729S) for D907 and Mix populations. All antibodies were used at a 1:100 dilution.

CUT&RUN data analysis

The quality of the reads was assessed using FastQC. Fastq files were aligned to the *D. melanogaster* reference genome dm6 using Bowtie 2 (v 2.4.2)[51] with the following parameters: --local ---very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I 10 -X 700. SAM files were compressed into BAM files using SAMtools (v 1.16.1)[52] and reads with low mapping quality (Phred score <30) were discarded. Duplicate reads were removed using sambamba markdup (v 1.0.0)[53] with the following parameters: -r --hash-table-size 500000 --overflow-list-size 500000. Peak calling was performed with each replicate as a separate input file and IgG as the control library using MACS2[54] with the following parameters: -g dm -f BAMPE -q 0.01. For visualization, reads per kilo base per million mapped reads (RPKM)-normalized bigWig binary files were generated using the bamCoverage function from deepTools2 (v 3.5.5)[55] with the following parameters: --normalizeUsing RPKM --ignoreDuplicates -e 0 -bs 10. Finally, replicates were merged using samtools merge with default parameters. Genome browser plots were generated using the pyGenomeTracks package (v 3.8)[56] and heatmaps using the plotHeatmap function from deepTools2. Differential enrichment in peak analysis was performed using the "DiffBind" R package (v 3.12.0)[57] with default normalization by sequencing library depth and the edgeR method for differential analysis (the significance cut-off was FDR<0.05 and |log2FC|>0.58). Assignment of genes to enriched regions was made to the nearest gene TSS within a 10Kb window using the "ChIPseeker" R package (v 1.38.0)[58].

Pool-seq assays

We performed whole-genome sequencing of pools of individuals (Pool-seq) from D907 and Mix (experiment 1 and 2) assimilated, non-assimilated, control and parental (P0) flies. We extracted high-quality genomic DNA for sequencing from 100 adult females from each population using the *Gentra Puregene Cell Kit* (*QIAGEN*). gDNA pellet was purified twice by phenol:chloroform:isoamyl alcohol precipitation and the pellet was resuspended in 100µl TE buffer. High-molecular weight gDNA was analyzed by gel electrophoresis, quantified on NanoDrop and >2µg of the mass was sent to Novogene for Illumina whole-genome sequencing

(WGS) library preparation. Paired-end sequencing of the libraries was performed on Novogene's NovaSeq 6000 PE150 platform, targeting ~100x reference genome coverage (14Gb per library).

Pool-seq analysis

Mapping to the reference genome and calling of variants (SNPs and InDels) was performed by Novogene's bioinformatics facility. Briefly, the raw sequencing data were aligned to the *D. melanogaster* reference genome dm6 using the BWA software[59] (parameters: mem -t 4 -k 32 -M). Library coverage ranged from 96.69% to 97.1%. The GATK pipeline[60] was used to call individual SNPs and InDels from the BAM files with the following parameters: '-T HaplotypeCaller --gcpHMM 10 -stand_emit_conf 10 -stand_call_conf 30 -ERC GVCF'. The ANNOVAR software[61] was used to annotate the detected variants. Significant changes in variant frequency between the derived lines (D907 control, assimilated and non-assimilated) and the base population (D907 P0) were tested by a Fisher test with Bonferroni's correction for multiple comparisons. Significant variants were filtered keeping only minimum coverage > 50, maximum coverage <250, coverage ratio threshold =0.5 and q-value threshold =0.0001. Manhattan plots were performed using the "qqman" R package[62], with a genome bin size of 10Kb for the averaged q-value.

Hi-C experiments

Hi-C experiments were performed using the *EpiTect Hi-C Kit* (*Quiagene* Cat. 59971). We used 50 third-instar larval wing imaginal discs per sample. Briefly, discs were homogenized and fixed in activated Buffer T and 2% Formaldehyde using Tissue Masher tubes (Biomasher II (EOG-sterilized) 320103 Funakoshi). Tissue was digested by adding 25µl Collagenase I and II (40 mg/ml) for 1 hour at 37°C. Samples were centrifuged and supernatant was carefully aspirated, leaving ~250µl of solution in the tube. Then 250µl QIAseq Beads equilibrated to room temperature were added to bind nuclei to the beads and all subsequent reactions were performed on the beads according to the manufactures protocol. Single replicate libraries made from D907 control, assimilated and non-assimilated selection lines were multiplexed and sequenced in a single lane on a DNBseq-G400 100bp paired-end platform from BGI (https://www.bgi.com/).

Hi-C analysis

Hi-C samples were analyzed using the TADbit pipeline[63], which was used to (i) assess the quality of the reads; (ii) map the paired-end reads to the *D. melanogaster* dm6 reference genome using Bowtie 2 (v.2.3.5.1), taking into account the restriction enzyme DpnII religation sites (fragment-based mapping); (iii) remove non-informative reads using the default TADbit filtering options. All valid pair interactions were then processed using the Cooler package (v. 0.9.1)[64] to generate .cool files at 100bp resolution. Subsequently, .mcool files at different resolutions (100bp to 40Kb) were obtained and normalized using the Iterative Correction and Eigenvector Decomposition algorithm (ICE)[65] with default parameters. HiGlass software[66] was used for matrix visualization.

Transposable elements genome-wide annotation analysis

In order to detect TE insertions that appeared de novo in the derived populations (not present in parental nor control populations: TE mobilizations), and TEs segregating in the parental population that could have an effect in the derived populations, we used the combination of the results of two TE caller programs: PoPoolationTE2 (v1.10.03)[67] and TEMP2 (v0.1.4)[68] to obtain reliable TE insertion calls. First, raw paired-end reads were trimmed using the FASTQ preprocessor Fastp (v.0.12.4)[69] with a minimum quality Phred score ≥ 20 (-q 20) and minimum read length of 20 bp (-l 20). Read quality was assessed using FastQC (v0.11.9). For detecting TE insertions with PoPoolationTE2, we first created a "TE-merged-reference" for *D. melanogaster*, which consists of the repeatmasked *D. melanogaster* reference genome (r.6.31) and the TE consensus sequences[70]. For creating the repeatmasked reference genome, we used RepeatMasker (v.4.1.2-p1) with options -gccalc -s -cutoff 200 -no_is -nolow -norna -gff -u with the high quality TE library available for *D. melanogaster*[70]. Next, we created the "TE hierarchy file" by using an ad hoc bash script. The raw reads of each sample were mapped to the TE-merged-reference, by using the local alignment algorithm BWA bwasw (v0.7.17-r1188)[59] with option -M. Both read pairs were mapped separately to the TE-merged-reference, and the paired end information was restored subsequently with module se2pe of PoPoolationTE2. A ppileup was generated for each sample with the PoPoolationTE2 ppileup function (with option --map-qual 15). Finally, TE insertions were identified with the modules identifySignatures (--min-count 2) and frequency. The final set of insertion per

sample was identified with the module pairupSignatures. We used BEDTools intersect with option -v (v2.30.0)[71] to remove the TE insertions detected in heterochromatic regions (defined in Rech et al. 2022). To detect TE insertions we used the insertion module of TEMP2. Since TEMP2 needs the TE annotations in bed format, we used the RepeatMasker annotation that was created for PoPoolationTE2 and transformed it to a bed file using rmsk2bed (from BEDOPS, v.2.4.39)[72]. We used BWA mem (v.0.7.17-r1188) with options -Y and -T 20 to map the paired-end reads to the *D. melanogaster* reference genome. For calculating the fragment length of the samples sequencing library needed for TEMP2, we used Picard's CollectInsertSizeMetrics module (v.2.26.11)[73] and used the mean insert size of each sample. Then, we used TEMP2 with parameter -m 5 (percentage of mismatch allowed when mapping to TEs) to detect TE insertions. TE insertions that were annotated in heterochromatic regions were removed with BEDTools intersect -v (v2.30.0). To combine the TE call information of both softwares, we used BEDOPS (v.2.4.39)[72], with the argument --everything, to combine in a single bed the annotations of PoPoolationTE2 and TEMP2 for each sample. We next used BEDTools (v2.30.0) with options merge -i and -d 20 to collapse insertions overlapping or allowing a maximum distance of 20bp into a single call. To consider a TE insertion detected both programs as the same, they must be from the same TE family and a maximum distance of the predicted insertion positions of 20bp. Finally, we combined all TEs insertions from the different lines (D907, Mix and Mix 2) into a single annotation file using the same strategy as above. The total number of insertions detected in each sample is in Extended Data Table 3a. For the frequency estimations, we only relied on the estimates of TEMP2. To calculate if the TE frequency shift between the parental/control and derived lines was significant, we performed a Fisher's exact test on R, using the frequencies in the parental and/or control line and the derived lines (N and or A) and considered as significant if the p-value was smaller than 0.01. Statistical analyses were performed in R (R Core Team 2022).

3' RACE-PCR assays

3' RACE-PCR assays were performed using *3' RACE System for Rapid Amplification of cDNA Ends kit* (*Thermo Fisher Scientific*, Cat. 18373-019). 20 third-instar larval wing discs were dissected in Schneider medium on ice, and total RNA was extracted using TRIzol reagent following purification

using the *RNA Clean & Concentrator™-5* kit (*Zymo Research*, cat. R1015) following manufacturer instructions and DNase I treatment (*QIAGEN*, cat. 79254). About 1µg of total RNA per sample was used for the reverse transcriptase reaction to synthesize first strand cDNA at the poly(A) region of the mRNAs using oligo(dT)-containing adapter primer. PCR amplification of Cad96Ca cDNA was done using the universal amplification primer provided with the kit (AUAP: 5'-GGCCACGCGTCGACTAGTAC-3') and the Cad96Ca-specific primer (5'-CGAGAGCGGTTTCCCGATCACAA-3'). Nested PCR was carried out in a second amplification reaction with the AUAP primer and another Cad96Ca-specific internal primer (5'-CGACCAATGCACTGAACCCGAAC-3'). PCR amplicons were resolved by an agarose 1% gel electrophoresis and DNA bands were recovered from the gel using the *NucleoSpin® Gel and PCR Clean-up* kit (MACHEREY-NAGEL) for Sanger sequencing by Eurofins Genomics.

Data availability

Newly generated data used in this study are listed in Extended Data Table 7 and the private record GSE255496 has been created in the Gene Expression Omnibus until it is made publicly available after publication. The following secure token has been created to allow review of data set GSE255496 while it remains in private status: otgxqougzfmnxch.

Code availability

All scripts used for detecting TE insertions and frequency analyses are available at: https://github.com/GonzalezLab/waddington-transposons.

Acknowledgments

We thank Marco Di Stefano, Michael Szalay and Flora Paldi for assistance with Hi-C experiments. We thank the BioCampus *Drosophila* facility (Montpellier, France) for technical support. G.S. was supported by the grant from the European Research Council (Advanced Grant 3DEpi, under grant agreement number 788972). The laboratory of G.C. was supported by grants from the European Research Council (Advanced Grant 3DEpi, under grant agreement number 788972), the European Union (CHROMDESIGN Project, under the Marie Skłodowska- Curie grant agreement number

813327), the Fondation pour la Recherche Médicale (EQU202303016280), the MSDAVENIR foundation (project GENE- IGH), the INSERM, the Centre National pour la Recherche Scientifique, the Agence Nationale de la Recherche (E- RARE project 'IMPACT' and "PLASMADIFF3D" project, ANR-18-CE15-0010) and by the French National Cancer Institute (INCa PLBIO18-362).

Author contributions

G.S. and G.C. conceived of and led the project. G.S. designed and performed the experiments. G.S. performed the *Drosophila* selection experiments supported by M.H.F-J. G.S. and G.C. interpreted the data. G.S. performed the bioinformatics analysis with the support of G.L.P. B.S. performed cloning experiments for *Drosophila* transgenesis microinjection. M.C-Z. and J.G. carried out the bioinformatics analysis of the transposable elements. G.S. wrote the manuscript with editorial input from G.C., J.G., B.S., M.C-Z., M.H.F-J. All authors reviewed and commented on the manuscript.

Competing interest statement

The authors declare no competing interests.

References

1. Petersen, N. S. & Mitchell, H. K. *Environmentally Induced Developmental Defects in Drosophila*. *In: Hightower L, Nover L, editors. Heat Shock and Development.* (New York: Springer-Verlag, 1991). doi:10.1017/CBO9781107415324.004.
2. Mitchell, H. K. & Petersen, N. S. Developmental abnormalities in Drosophila induced by heat shock. *Dev. Genet.* **3**, 91–102 (1982).
3. Waddington, C. H. Genetic Assimilation of an Acquired Character. *Evolution (N. Y).* **7**, 118–126 (1953).
4. Bateman, K. G. The Genetic Assimilation of Four Venation Phenocopies. *J. Genet.* **56**, 443–474 (1959).
5. Pfennig, D. W. *Phenotypic Plasticity and Evolution*. (CRC Press, 2021). doi:https://doi.org/10.1201/9780429343001.
6. West-Eberhard, M. J. *Developmental Plasticity and Evolution*. *Oxford University Press* (Oxford University Press, 2003). doi:https://doi.org/10.1093/oso/9780195122343.001.0001.
7. Marzec, S. R., Pelletier, K., Chang, A. H.-P. & Dworkin, I. Reexamining Waddington: Canalization and new mutations are not required for the evolution of genetic assimilation. *bioRxiv* 2022.01.09.475581 (2022).
8. Fanti, L., Piacentini, L., Cappucci, U., Casale, A. M. & Pimpinelli, S. Canalization by Selection of de Novo Induced Mutations. *Genetics* **206**, 1995–2006 (2017).
9. Waddington, C. H. Genetic Assimilation of the Bithorax Phenotype. *Evolution (N. Y).* **X**, 1–13 (1956).
10. Ho, M. W., Tucker, C., Keeley, D. & Saunders, P. T. Effects of successive generations of ether

673 treatment on penetrance and expression of the bithorax phenocopy in Drosophila melanogaster. *J.*
674 *Exp. Zool.* **225**, 357–368 (1983).
675 11. Suzuki, Y. & Nijhout, H. F. Evolution of a Polyphenism by Genetic Accommodation. *Science (80-.).*
676 **311**, 650–652 (2006).
677 12. van der Burg, K. R. L. *et al.* Genomic architecture of a genetically assimilated seasonal color pattern.
678 *Science (80-.).* **725**, 721–725 (2020).
679 13. Pigliucci, M. & Murren, C. J. Perspective: Genetic assimilation and a possible evolutionary paradox:
680 Can macroevolution sometimes be so fast as to pass us by? *Evolution (N. Y).* **57**, 1455–1464 (2003).
681 14. Schlichting, C. D. & Wund, M. A. Phenotypic plasticity and epigenetic marking: An assessment of
682 evidence for genetic accommodation. *Evolution (N. Y).* **68**, 656–672 (2014).
683 15. Ehrenreich, I. M. & Pfennig, D. W. Genetic assimilation: A review of its potential proximate causes
684 and evolutionary consequences. *Ann. Bot.* **117**, 769–779 (2016).
685 16. Gibert, J. M. The flexible stem hypothesis: evidence from genetic data. *Dev. Genes Evol.* **227**, 297–
686 307 (2017).
687 17. Nishikawa, K. & Kinjo, A. R. Mechanism of evolution by genetic assimilation: Equivalence and
688 independence of genetic mutation and epigenetic modulation in phenotypic expression. *Biophys.*
689 *Rev.* **10**, 667–676 (2018).
690 18. Masel, J. Genetic assimilation can occur in the absence of selection for the assimilating phenotype,
691 suggesting a role for the canalization heuristic. *J. Evol. Biol.* **17**, 1106–1110 (2004).
692 19. Raju, A., Xue, B. K. & Leibler, S. A theoretical perspective onWaddington's genetic assimilation
693 experiments. *Proc. Natl. Acad. Sci. U. S. A.* **120**, 1–10 (2023).
694 20. Pigliucci, M. Phenotypic plasticity and evolution by genetic assimilation. *J. Exp. Biol.* **209**, 2362–
695 2367 (2006).
696 21. De Jong, G. Evolution of phenotypic plasticity: Patterns of plasticity and the emergence of
697 ecotypes. *New Phytol.* **166**, 101–118 (2005).
698 22. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* **396**,
699 336–342 (1998).
700 23. Levis, N. A. & Pfennig, D. W. Plasticity-led evolution: A survey of developmental mechanisms and
701 empirical tests. *Evol. Dev.* **22**, 71–87 (2020).
702 24. Waddington, C. H. Genetic assimilation. *Adv. Genet.* **10**, 257–293 (1961).
703 25. Nijhout, H. F., Kudla, A. M. & Hazelwood, C. C. *Genetic assimilation and accommodation: Models*
704 *and mechanisms*. *Current Topics in Developmental Biology* vol. 141 (Elsevier Inc., 2021).
705 26. Sollars, V. *et al.* Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for
706 morphological evolution. *Nat. Genet.* **33**, 70–74 (2003).
707 27. Ruden, D. M., Garfinkel, M. D., Sollars, V. E. & Lu, X. Waddington's widget: Hsp90 and the
708 inheritance of acquired characters. *Semin. Cell Dev. Biol.* **14**, 301–310 (2003).
709 28. Sabarís, G., Fitz-James, M. H. & Cavalli, G. Epigenetic inheritance in adaptive evolution. *Ann. N. Y.*
710 *Acad. Sci.* **1524**, 22–29 (2023).
711 29. MacKay, T. F. C. *et al.* The Drosophila melanogaster Genetic Reference Panel. *Nature* **482**, 173–178
712 (2012).
713 30. Carroll, S. B. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of
714 Morphological Evolution. *Cell* **134**, 25–36 (2008).
715 31. Stern, D. L. & Orgogozo, V. The loci of evolution: How predictable is genetic evolution? *Evolution*
716 *(N. Y).* **62**, 2155–2177 (2008).
717 32. Blair, S. S. Wing Vein Patterning in *Drosophila* and the Analysis of Intercellular Signaling. *Annu. Rev.*
718 *Cell Dev. Biol.* **23**, 293–319 (2007).
719 33. Franz, M. *et al.* GeneMANIA update 2018. *Nucleic Acids Res.* **46**, W60–W64 (2018).
720 34. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts

developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
35. Cutforth, T. & Rubin, G. M. Mutations in Hsp83 and cdc37 impair signaling by the sevenless receptor tyrosine kinase in Drosophila. *Cell* **77**, 1027–1036 (1994).
36. Menezes, B. F. *et al.* An attempt to select non-genetic variation in resistance to starvation and reduced chill coma recovery time in Drosophila melanogaster. *J. Exp. Biol.* **222**, (2018).
37. Schlötterer, C., Kofler, R., Versace, E., Tobler, R. & Franssen, S. U. Combining experimental evolution with next-generation sequencing: A powerful tool to study adaptation from standing genetic variation. *Heredity (Edinb).* **114**, 431–440 (2015).
38. Durmaz, E., Kerdaffrec, E., Katsianis, G., Kapun, M. & Flatt, T. How Selection Acts on Chromosomal Inversions. *eLS* **1**, 307–315 (2020).
39. Inoue, Y. & Igarashi, Y. On the category of naturally occurring inversions of Drosophila melanogaster. *Japanese J. Genet.* **69**, 105–118 (1994).
40. Miller, D. E. *et al.* Identification and characterization of breakpoints and mutations on drosophila melanogaster balancer chromosomes. *G3 Genes, Genomes, Genet.* **10**, 4271–4285 (2020).
41. Ghavi-Helm, Y. *et al.* Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
42. Lee, Y. C. G. & Karpen, G. H. Pervasive epigenetic effects of drosophila euchromatic transposable elements impact their evolution. *Elife* **6**, 1–31 (2017).
43. Saito, K. & Siomi, M. C. Small RNA-Mediated Quiescence of Transposable Elements in Animals. *Dev. Cell* **19**, 687–697 (2010).
44. Wang, S. *et al.* The tyrosine kinase Stitcher activates Grainy head and epidermal wound healing in Drosophila. *Nat. Cell Biol.* **11**, 890–895 (2009).
45. Lee, H. & Adler, P. N. The grainy head transcription factor is essential for the function of the frizzled pathway in the Drosophila wing. *Mech. Dev.* **121**, 37–49 (2004).
46. Coronado-Zamora, M. & González, J. Transposons contribute to the functional diversification of the head, gut, and ovary transcriptomes across Drosophila natural strains. *Genome Res.* **33**, 1541–1553 (2023).
47. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **41**, 563–571 (2009).
48. Huang, W., Carbone, M. A., Lyman, R. F., Anholt, R. R. H. & Mackay, T. F. C. Genotype by environment interaction for gene expression in Drosophila melanogaster. *Nat. Commun.* **11**, 1–10 (2020).
49. Buratovich, M. A., Phillips, R. G. & Whittle, J. R. S. Genetic relationships between the mutations spade and Sternopleural and the wingless gene in Drosophila development. *Dev. Biol.* **185**, 244–260 (1997).
50. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
51. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
52. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
53. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
54. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol. 2008 99* **9**, 1–9 (2008).
55. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
56. Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).

57. Stark, R. & Brown, G. *DiffBind: Differential binding analysis of ChIPSeq peak data*. (2021).
58. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
59. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
60. Depristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–501 (2011).
61. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, 1–7 (2010).
62. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *J. Open Source Softw.* **3**, 731 (2018).
63. Serra, F. *et al.* Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* **13**, 1–17 (2017).
64. Abdennur, N. & Mirny, L. A. Cooler: Scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
65. Imakaev, M. *et al.* Iterative Correction of Hi-C Data Reveals Hallmarks of Chromosome Organization. *Nat. Methods* **9**, 999–1003 (2012).
66. Kerpedjiev, P. *et al.* HiGlass: Web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 1–12 (2018).
67. Kofler, R., Gómez-Sánchez, D. & Schlötterer, C. PoPoolationTE2: Comparative Population Genomics of Transposable Elements Using Pool-Seq. *Mol. Biol. Evol.* **33**, 2759–2764 (2016).
68. Yu, T. *et al.* A benchmark and an algorithm for detecting germline transposon insertions and measuring de novo transposon insertion frequencies. *Nucleic Acids Res.* **49**, E44 (2021).
69. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
70. Rech, G. E. *et al.* Population-scale long-read sequencing uncovers transposable elements associated with gene expression variation and adaptive signatures in Drosophila. *Nat. Commun.* **13**, 1–16 (2022).
71. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
72. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919 (2012).
73. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

Figure 1

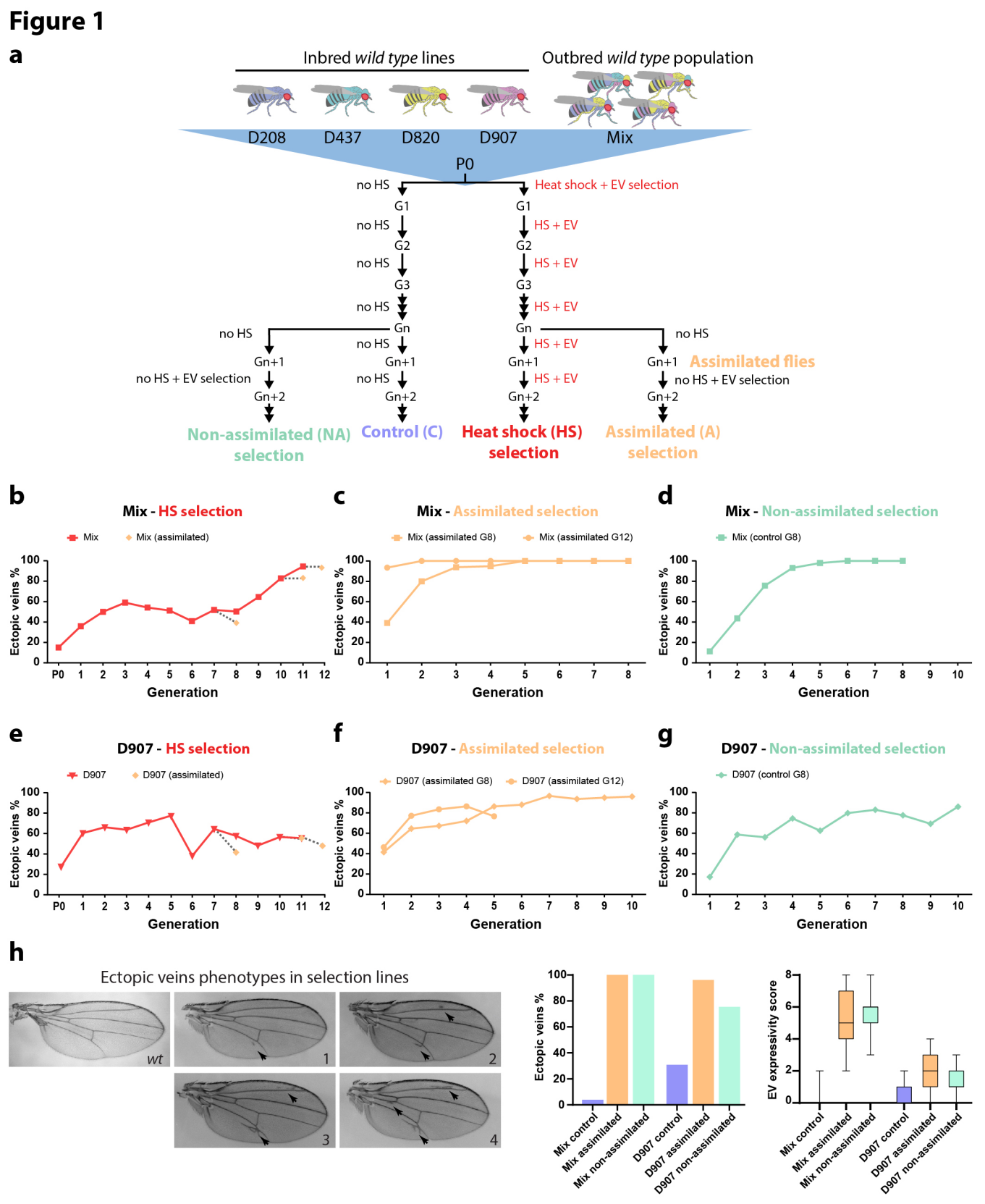

a
Inbred wild type lines
Outbred wild type population
D208
D437
D820
D907
Mix
P0
no HS
Heat shock + EV selection
G1
HS + EV
G2
G3
Gn
Gn+1
Gn+2
no HS + EV selection
Assimilated flies
Non-assimilated (NA) selection
Control (C)
Heat shock (HS) selection
Assimilated (A) selection
b
Mix - HS selection
Mix
Mix (assimilated)
c
Mix - Assimilated selection
Mix (assimilated G8)
Mix (assimilated G12)
d
Mix - Non-assimilated selection
Mix (control G8)
e
D907 - HS selection
D907
D907 (assimilated)
f
D907 - Assimilated selection
D907 (assimilated G8)
D907 (assimilated G12)
g
D907 - Non-assimilated selection
D907 (control G8)
Ectopic veins %
Generation
h
Ectopic veins phenotypes in selection lines
wt
1
2
3
4
Ectopic veins %
EV expressivity score
Mix control
Mix assimilated
Mix non-assimilated
D907 control
D907 assimilated
D907 non-assimilated

Figure 1: Rapid assimilation of ectopic veins trait in outbred and inbred natural fly populations.
a- Experimental design used to recapitulate the Waddington assimilation experiment for the ectopic veins (EV) phenocopy. EV selection was applied to four inbred fly lines (D208, D437, D820 and D907) and an outbred population (Mix) consisting of the genetic admixture of the four inbred lines. Heat shock (HS) and EV selection was performed over several generations in the HS selection lines. Further EV selection in the absence of heat shock was derived from the HS selection in the assimilated selection line. Control flies were maintained under normal conditions and EV selection without heat shock was derived from these flies in the non-assimilated selection. **b-** Solid line shows the ectopic veins penetrance as a response to heat shock induction and EV artificial selection in the Mix outbred population. The individual diamond dots connected with dashed lines indicate the EV penetrance of the assimilated flies (progeny of flies from the heat shock selection without pupal heat shock in the current generation) tested in alternate generations. **c-** EV penetrance response in Mix assimilated selection lines derived from the eighth (squared dots) and twelfth (circled dots) generations. **d-** EV penetrance response in Mix non-assimilated selection line derived from control flies from the eighth generation. **e-** Solid line shows the ectopic veins penetrance as a response to heat shock induction and EV artificial selection in the D907 inbred population. The individual diamond dots connected with dashed lines indicate the EV penetrance of the assimilated flies (flies from the heat shock selection without pupal heat shock in the current generation) tested in alternate generations. **f-** EV penetrance response in D907 assimilated selection lines derived from the eighth (diamonds dots) and twelfth (circled dots) generations. **g-** EV penetrance response in D907 non-assimilated selection line derived from control flies from the eighth generation. **h-** Representative images of the ectopic veins phenotypes in the selected flies showing the different expressivity score values per wing. Penetrance and expressivity of ectopic veins in adult females of the Mix and D907 selection lines. The two wings were analyzed in each individual, where each wing can have an EV expressivity score from 0 (no EV) to 4, with a maximum EV score of 8 per fly.

Figure 2

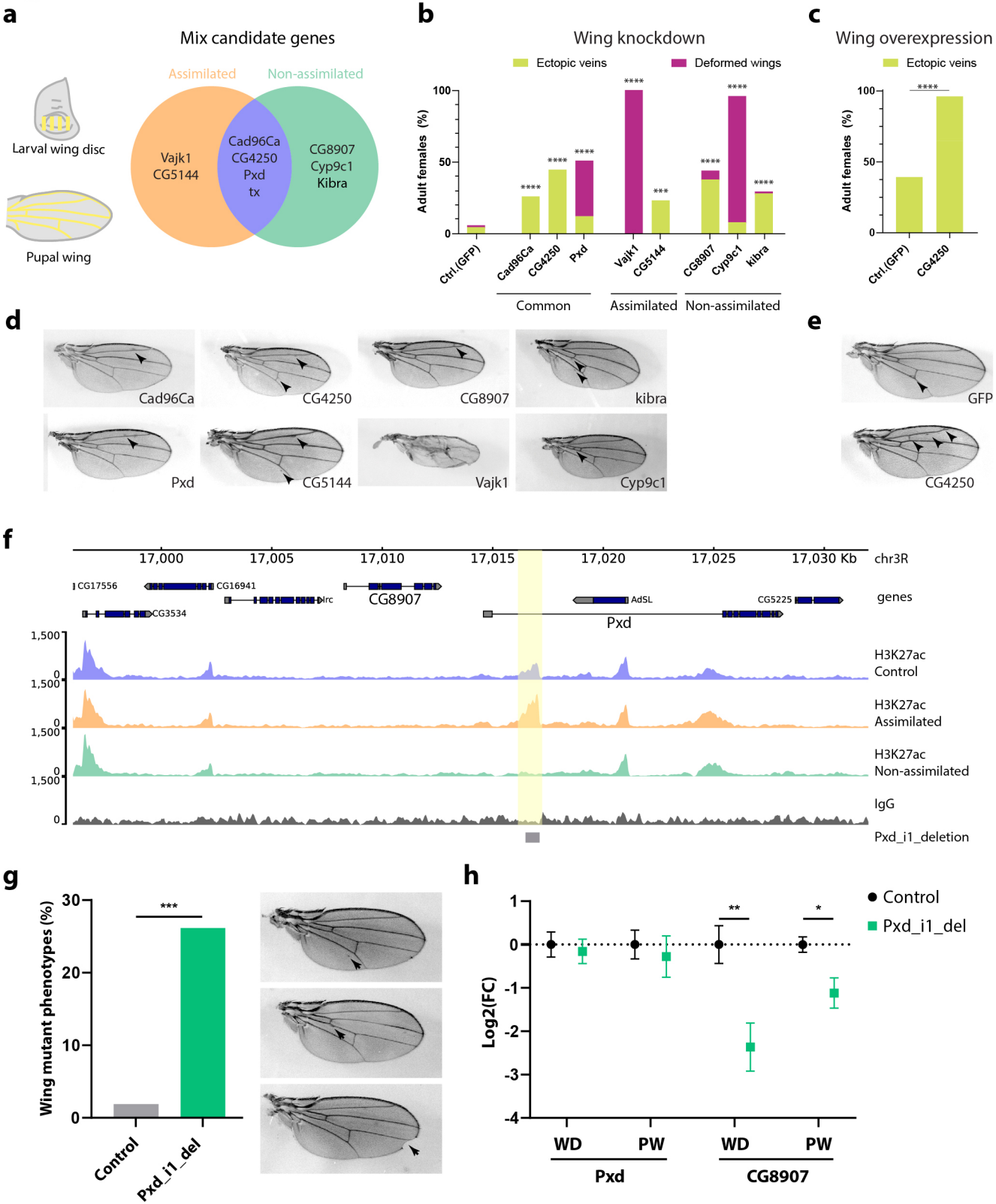
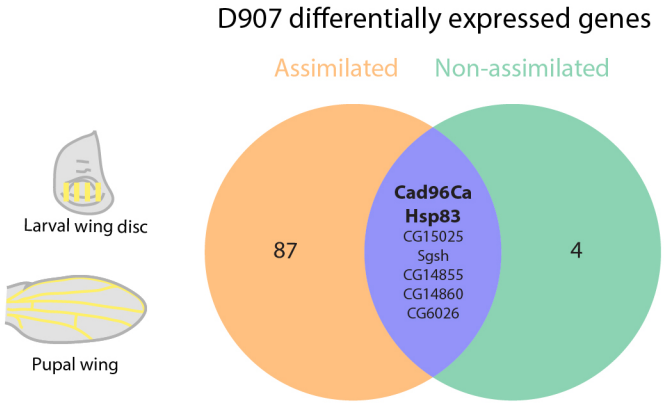

a
Mix candidate genes
Assimilated
Non-assimilated
Larval wing disc
Pupal wing
Vajk1
CG5144
Cad96Ca
CG4250
Pxd
tx
CG8907
Cyp9c1
Kibra
b
Wing knockdown
Ectopic veins
Deformed wings
Adult females (%)
Ctrl.(GFP)
Cad96Ca
CG4250
Pxd
Vajk1
CG5144
CG8907
Cyp9c1
kibra
Common
Assimilated
Non-assimilated
c
Wing overexpression
Ectopic veins
Adult females (%)
Ctrl.(GFP)
CG4250
d
Cad96Ca
CG4250
CG8907
kibra
Pxd
CG5144
Vajk1
Cyp9c1
e
GFP
CG4250
f
17,000
17,005
17,010
17,015
17,020
17,025
17,030 Kb
chr3R
CG17556
CG16941
Irc
CG8907
AdSL
CG5225
CG3534
Pxd
genes
H3K27ac Control
H3K27ac Assimilated
H3K27ac Non-assimilated
IgG
Pxd_i1_deletion
g
Wing mutant phenotypes (%)
Control
Pxd_i1_del
h
Log2(FC)
Control
Pxd_i1_del
WD
PW
WD
PW
Pxd
CG8907

Figure 2: Multiple gene deregulations underlie the evolution of the ectopic veins trait in the Mix selection lines

a- Shortlist of confident candidate genes associated with the ectopic veins evolution in Mix assimilated and non-assimilated selection lines obtained from the transcriptome analysis in the larval wing disc and pupal wing. **b-** Quantification of the wing phenotypes caused by the functional characterization of the confident candidate genes by gene knockdown in the wing using the *nub-GAL4* driver. *GFP* knockdown was used as control. **c-** Overexpression of *GFP* (control) and *CG4250* genes in the wing using *nub-GAL4* driver. Statistical differences in **b** and **c** were analyzed relative to control flies using two-sided Fisher's exact test (***$p < 0.001$, and ****$p < 0.0001$). **d,e-** Representative images of wing mutant phenotypes induced by gene knockdowns (**d**) or gene overexpression (**e**) in the wing. Black arrowheads in the images indicate ectopic veins. **f-** CUT&RUN tracks for H3K27Ac (merge of three replicates) and IgG (control) in the wing disc for the Mix lines. Yellow shading marks the position of the H3K27ac differential peak in Mix non-assimilated in the first intron of the *Pxd* gene. The 480bp deleted by the CRISPR/Cas9 system is shown in the grey box. (*Pxd_i1_del*). **g-** Quantification and representative images of the wing mutant phenotypes induced by *Pxd_i1_del* (Two-sided Chi-square test, ***$p<0.001$). **h-** Expression level of the *Pxd* and *CG8907* genes in the wing disc (WD) and pupal wing (PW) of *Pxd_i1_del* and control (*nos-Cas9*) flies analyzed by RT-qPCR. The plot shows the averaged log2 fold change normalized to control flies and error bars represent the standard error of the mean (SEM) from six biological replicates for the WD and three for the PW. Significance was calculated using unpaired t-test (*$p < 0.05$ and **$p < 0.01$).

Figure 3

a



b

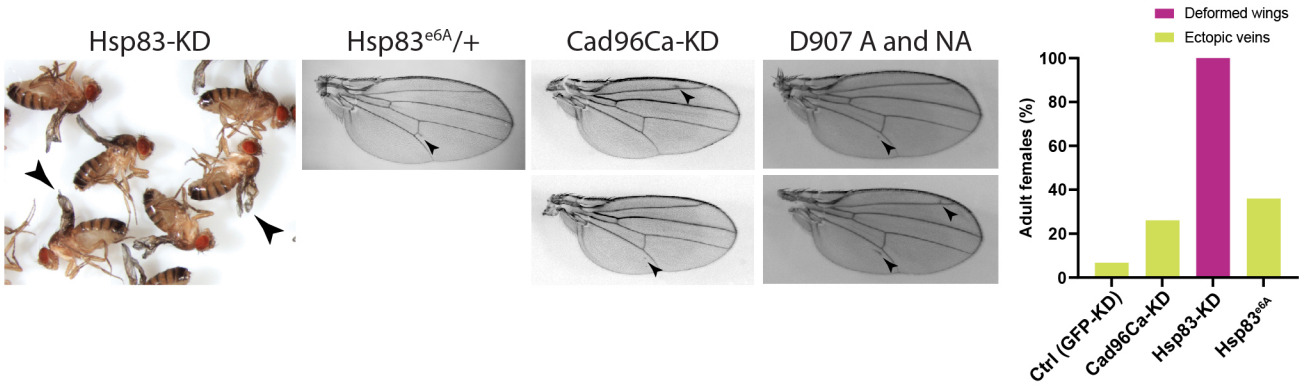


Figure 3: Downregulation of Cad96Ca and Hsp83 genes underlie the evolution of ectopic veins in D907 selection lines

a- Overlap of differentially expressed genes in D907 assimilated and non-assimilated lines in wing disc and pupal wing. *Cad96Ca* and *Hsp83* are downregulated in wing disc and pupal wings respectively in both selection lines. **b-** Functional characterization of the *Hsp83* and *Cad96Ca* candidate genes by knockdown in the wing using the *A9-GAL4* driver (Hsp83-KD) and *nub-GAL4* (Cad96Ca-KD) and the null allele $Hsp83^{e6A}$ in heterozygous flies. Representative images of the phenotypes are shown together with the ectopic veins in D907 assimilated (A) and non-assimilated (NA) lines. Penetrance values refer to adult females.

Figure 4

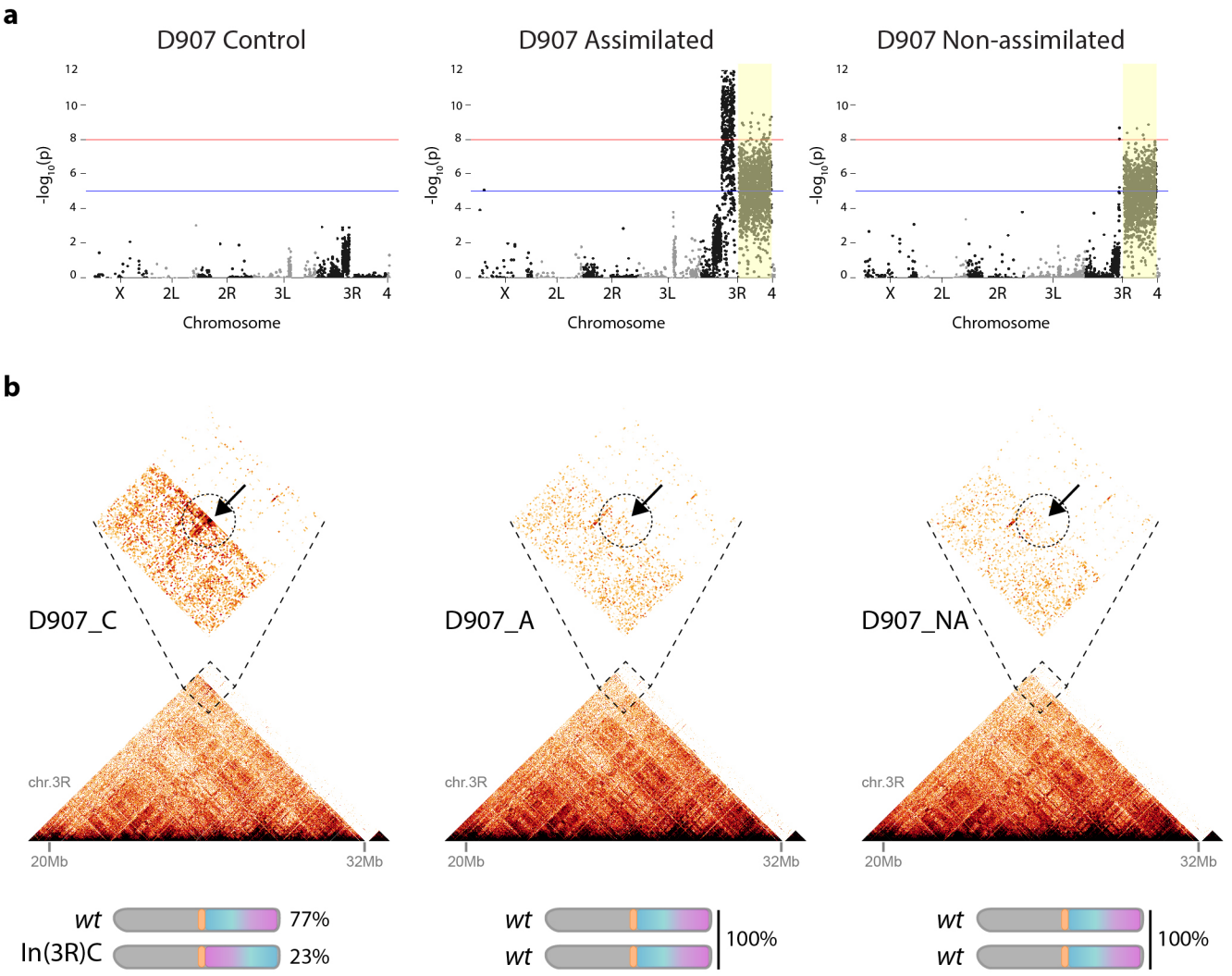


Figure 4: A large chromosomal inversion maintains the genetic variation associated with the EV evolution in the D907 selection lines

a- Manhattan plots showing the genome-wide allele associations from the Pool-seq analysis in D907 control, assimilated and non-assimilated lines relative to the ancestral population (P0). Yellow shading marks the position of the chromosome 3R spanned by the chromosomal inversion *In(3R)C*. **b-** Hi-C maps in D907 control (C), assimilated (A) and non-assimilated (NA) flies, centered on the region spanned by the chromosomal inversion *In(3R)C* (chr3R: ~20.3-32Mb). Dashed lines boxes show magnifications of the breakpoint site interaction. Cartoons show the estimated frequency of each chromosomal variant in each population.

Figure 5

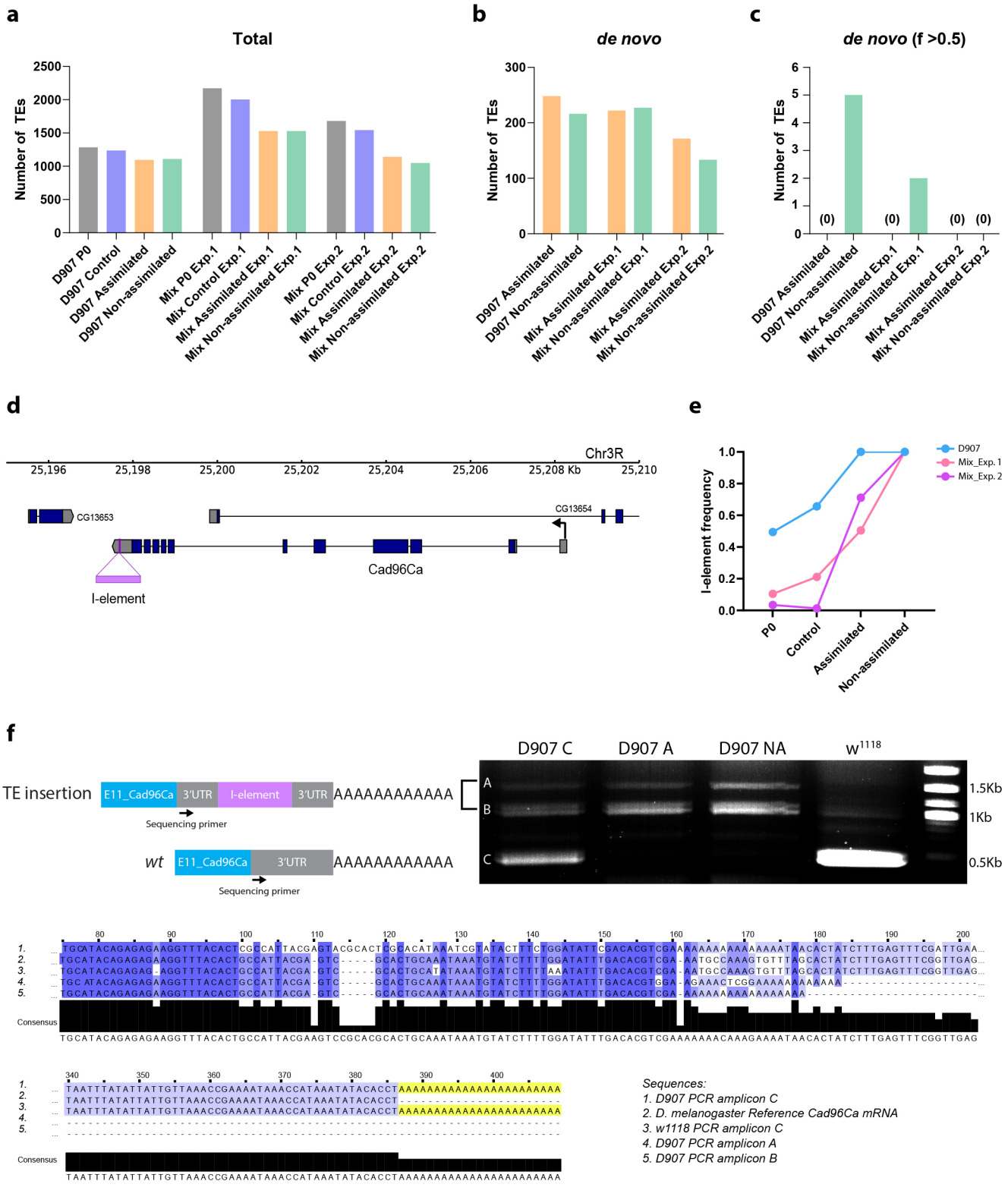

a
Total
Number of TEs
2500
2000
1500
1000
500
0
D907 P0
D907 Control
D907 Assimilated
D907 Non-assimilated
Mix P0 Exp.1
Mix Control Exp.1
Mix Assimilated Exp.1
Mix Non-assimilated Exp.1
Mix P0 Exp.2
Mix Control Exp.2
Mix Assimilated Exp.2
Mix Non-assimilated Exp.2
b
de novo
Number of TEs
300
200
100
0
D907 Assimilated
D907 Non-assimilated
Mix Assimilated Exp.1
Mix Non-assimilated Exp.1
Mix Assimilated Exp.2
Mix Non-assimilated Exp.2
c
de novo (f >0.5)
Number of TEs
6
5
4
3
2
1
0
(0)
(0)
(0)
(0)
D907 Assimilated
D907 Non-assimilated
Mix Assimilated Exp.1
Mix Non-assimilated Exp.1
Mix Assimilated Exp.2
Mix Non-assimilated Exp.2
d
Chr3R
25,196
25,198
25,200
25,202
25,204
25,206
25,208 Kb
25,210
CG13653
CG13654
Cad96Ca
I-element
e
I-element frequency
1.0
0.8
0.6
0.4
0.2
0.0
P0
Control
Assimilated
Non-assimilated
D907
Mix_Exp.1
Mix_Exp.2
f
TE insertion
E11_Cad96Ca
3'UTR
I-element
3'UTR
AAAAAAAAAAAA
Sequencing primer
wt
E11_Cad96Ca
3'UTR
AAAAAAAAAAAA
Sequencing primer
D907 C
D907 A
D907 NA
w1118
A
B
C
1.5Kb
1Kb
0.5Kb
Consensus
Consensus
Sequences:
1. D907 PCR amplicon C
2. D. melanogaster Reference Cad96Ca mRNA
3. w1118 PCR amplicon C
4. D907 PCR amplicon A
5. D907 PCR amplicon B

Figure 5: Segregating transposable element variation contributes to the evolution of EV in D907 and Mix selection lines

a- Total number of transposable element (TE) insertions mapped to the genome of the Mix and D907 lines. **b-** Number of *de novo* TE insertions mapped to the genome of the Mix and D907 lines. **c-** Number of *de novo* TE insertions with a frequency greater than 0.5 mapped to the genome of the Mix and D907 lines. **d-** Scheme of the *Cad96Ca* locus showing the *I-element* insertion mapped to the 3'UTR region. **e-** *I-element* frequency in D907 and Mix (Experiment 1 and 2) ancestral and derived populations. **f-** Agarose gel electrophoresis resolution of 3'RACE-PCR amplicons specific for *Cad96Ca* mRNA isoforms in w^{1118} (*wild type*), D907 control (C), assimilated (A) and non-assimilated (NA). Clustal Omega multiple sequence alignment of the DNA amplification products (A, B and C bands in gel) and the *D. melanogaster Cad96Ca* reference mRNA sequence (NCBI RefSeq database: NM_143092.3). Black arrow in cartoons illustrate the position of the sequencing primer specific for the *Cad96Ca* 3'UTR.

Figure 6

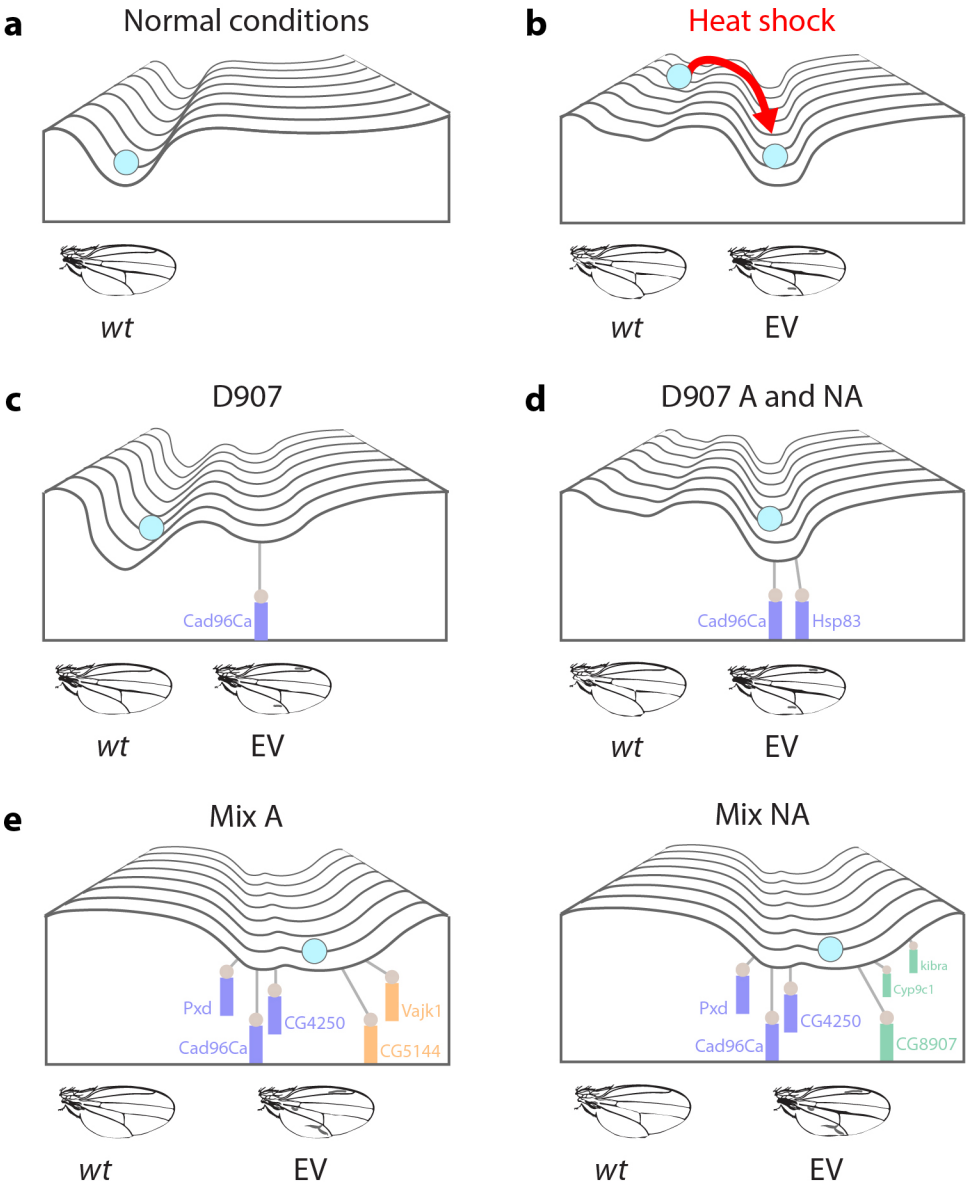
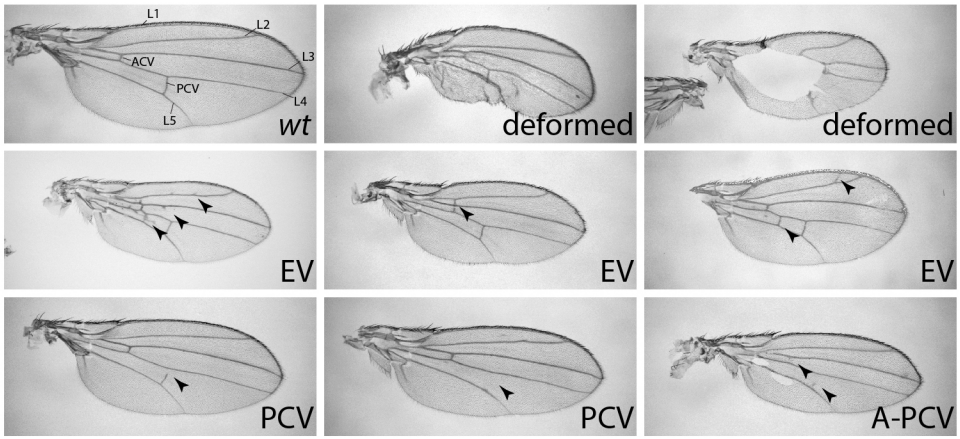


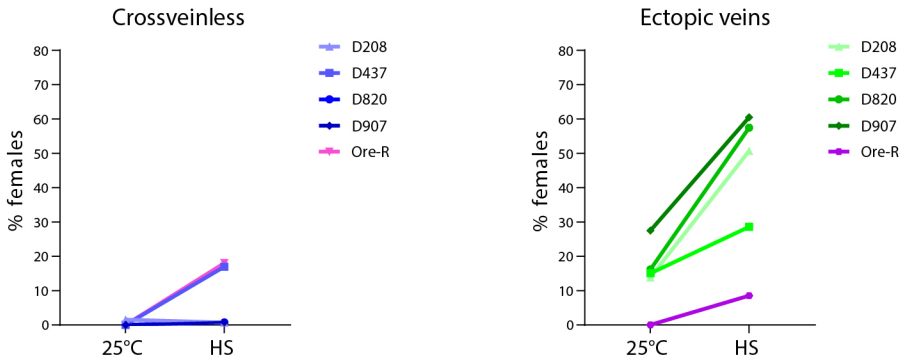
Figure 6: Genetic canalization of the ectopic veins trait in inbred and outbred natural fly populations
a- Under normal conditions development is canalized towards the *wild type* wing vein pattern. **b-** Developmental stress such as heat shock can induce decanalization of normal development leading to the ectopic veins phenocopy. **c-** In D907 flies, development is partially decanalized under normal conditions by expression of the *Cad96Ca*$^{[l\text{-}element]}$ allele, resulting in ectopic vein phenotypes in some individuals of the population. **d-** Selection for the ectopic veins trait in D907 assimilated (A) and non-assimilated (NA) lines resulted in canalization of the trait by downregulation of the *Cad96Ca* and *Hsp83* genes. **e-** Mix assimilated (A) and non-assimilated (NA) selection lines canalized the ectopic veins trait through a multigenic response. Some alleles were selected in both lines, leading to deregulation of common genes (such as *Cad96Ca*, *CG4250* and *Pxd*), while other trait modifier alleles were selected exclusively in each line under the alternative selection environment.

Extended Data Fig. 1

a



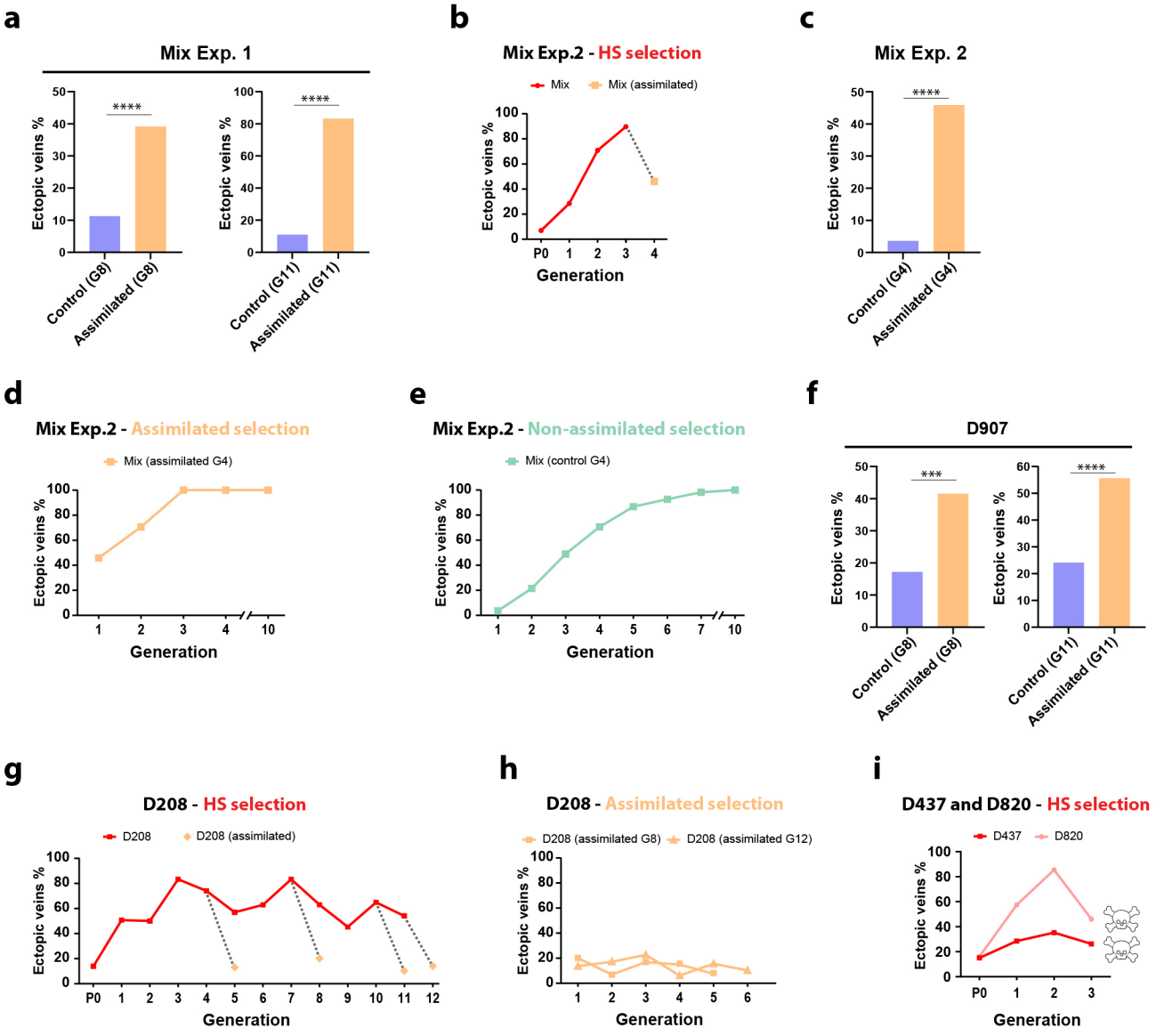
b



Extended Data Fig. 1: Pupal heat shock induces diverse wing phenocopies

a- Representative images of the *Drosophila wild type* wing vein pattern indicating the five longitudinal veins (L1-5) and the anterior (ACV) and posterior crossveins (PCV). The different wing phenocopies induced by pupal heat shock are also shown, such as different ectopic veins (EV), breakage or loss of the anterior or posterior crossvein, defined as "crossveinless" (A-PCV), and deformed wings. **b-** Plasticity of crossveinless and ectopic veins phenocopies upon pupal heat shock (HS) in different *wild type* laboratory strains.

Extended Data Fig. 2

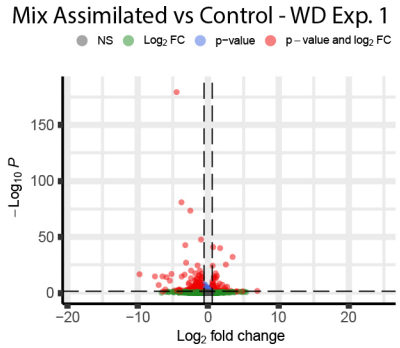


Extended Data Fig. 2: Recapitulation of Waddington genetic assimilation experiment for ectopic veins in inbred and outbred natural fly populations

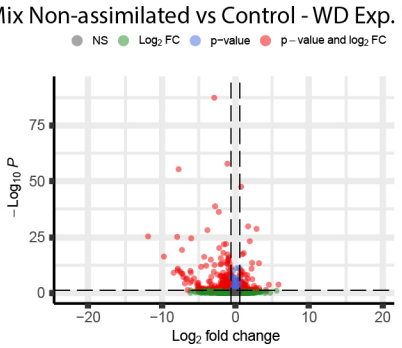
a- Comparison of ectopic vein penetrance between Mix control and assimilated flies in the eighth and eleventh generations for the first selection experiment (two-sided Chi-square test: ****$p<0.0001$). **b-** Solid line shows the ectopic veins penetrance as a response to heat shock induction and EV artificial selection in the Mix population for the second independent selection experiment. The individual squared dot connected with a dashed line indicates the EV penetrance of the assimilated flies. **c-** Comparison of ectopic vein penetrance between Mix control and assimilated flies in the fourth generation for the second selection experiment (two-sided Chi-square test: ****$p<0.0001$). **d-** EV penetrance response in Mix assimilated selection for the second selection experiment in a line derived from the fourth generation. **e-** EV penetrance response in Mix non-assimilated selection for the second selection experiment in a line derived from control flies from the fourth generation. **f-** Comparison of ectopic vein penetrance between D907 control and assimilated flies in the eighth and eleventh generations (two-sided Chi-square test: ***$p<0.001$, ****$p<0.0001$). **g-** Solid line shows the ectopic veins penetrance as a response to heat shock induction and EV artificial selection in the D208 population. The individual diamond dots connected with a dashed line indicates the EV penetrance of the assimilated flies tested in alternate generations. **h-** EV penetrance response in D208 assimilated selection in lines derived from the eighth and twelfth generations. **i-** EV penetrance response to heat shock induction and EV artificial selection in the D437 and D820 populations.

Extended Data Fig. 3

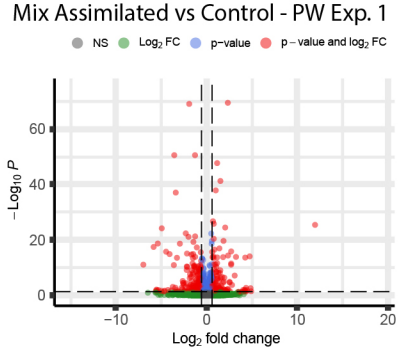
DEGs: 252



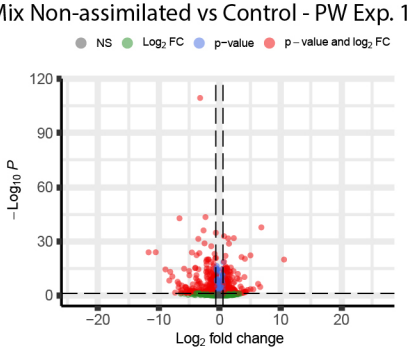
DEGs: 301



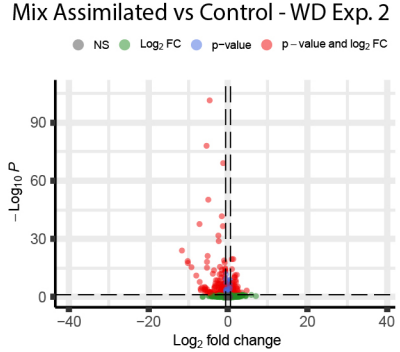
DEGs: 302



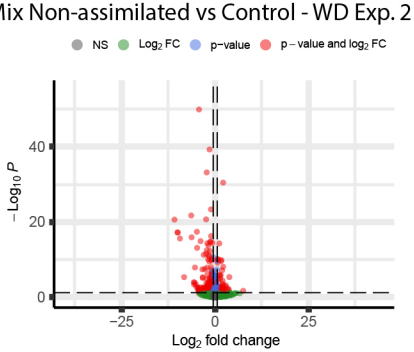
DEGs: 513



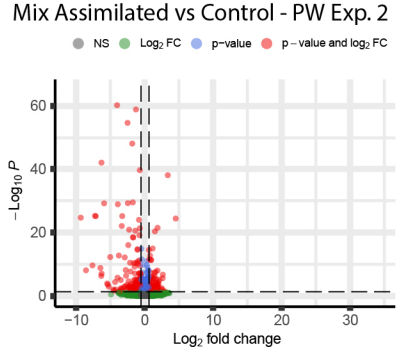
DEGs: 221



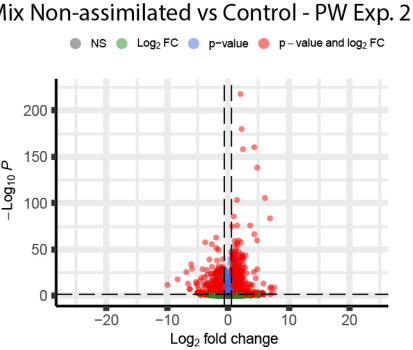
DEGs: 260



DEGs: 262



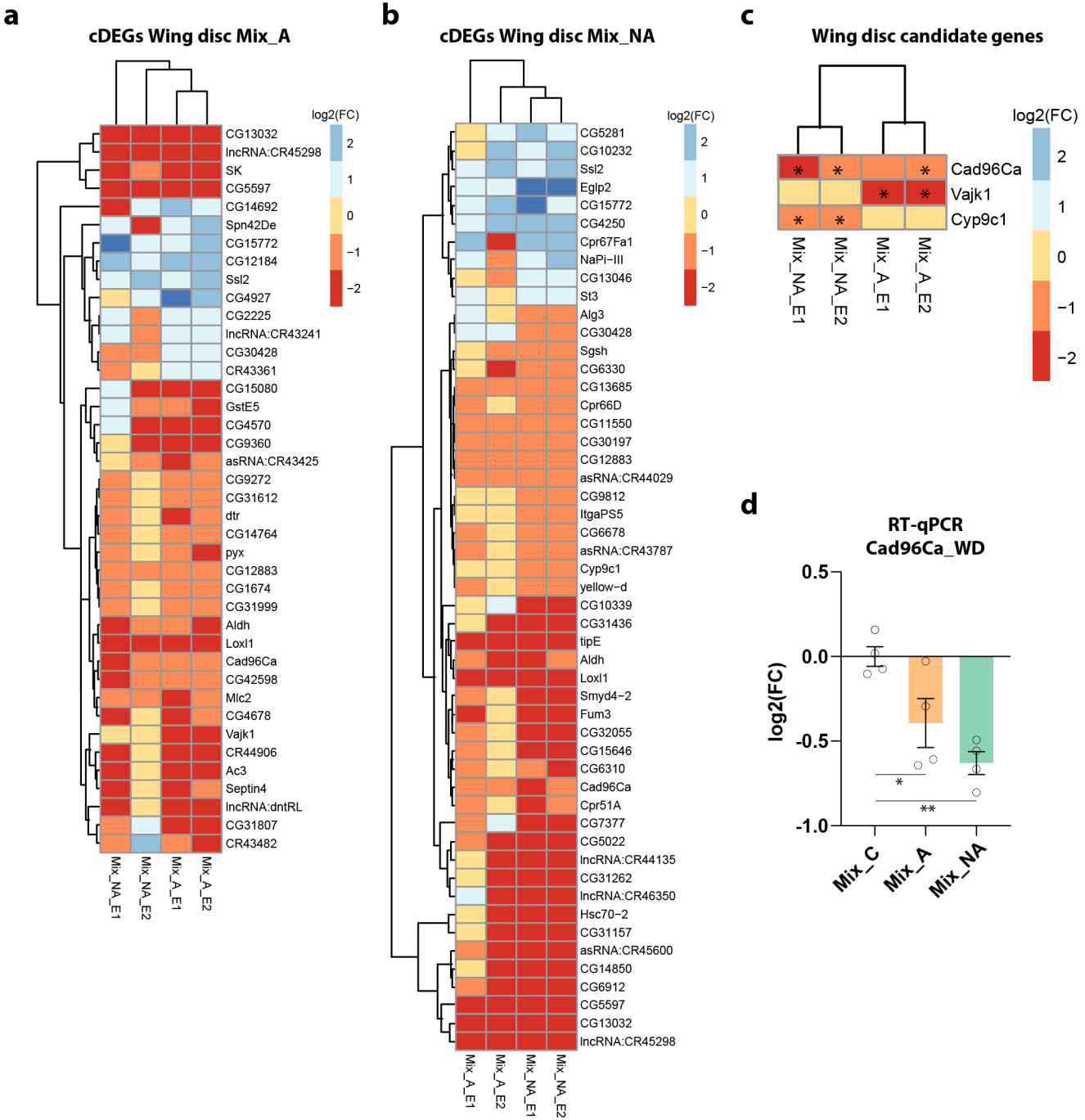
DEGs: 1499

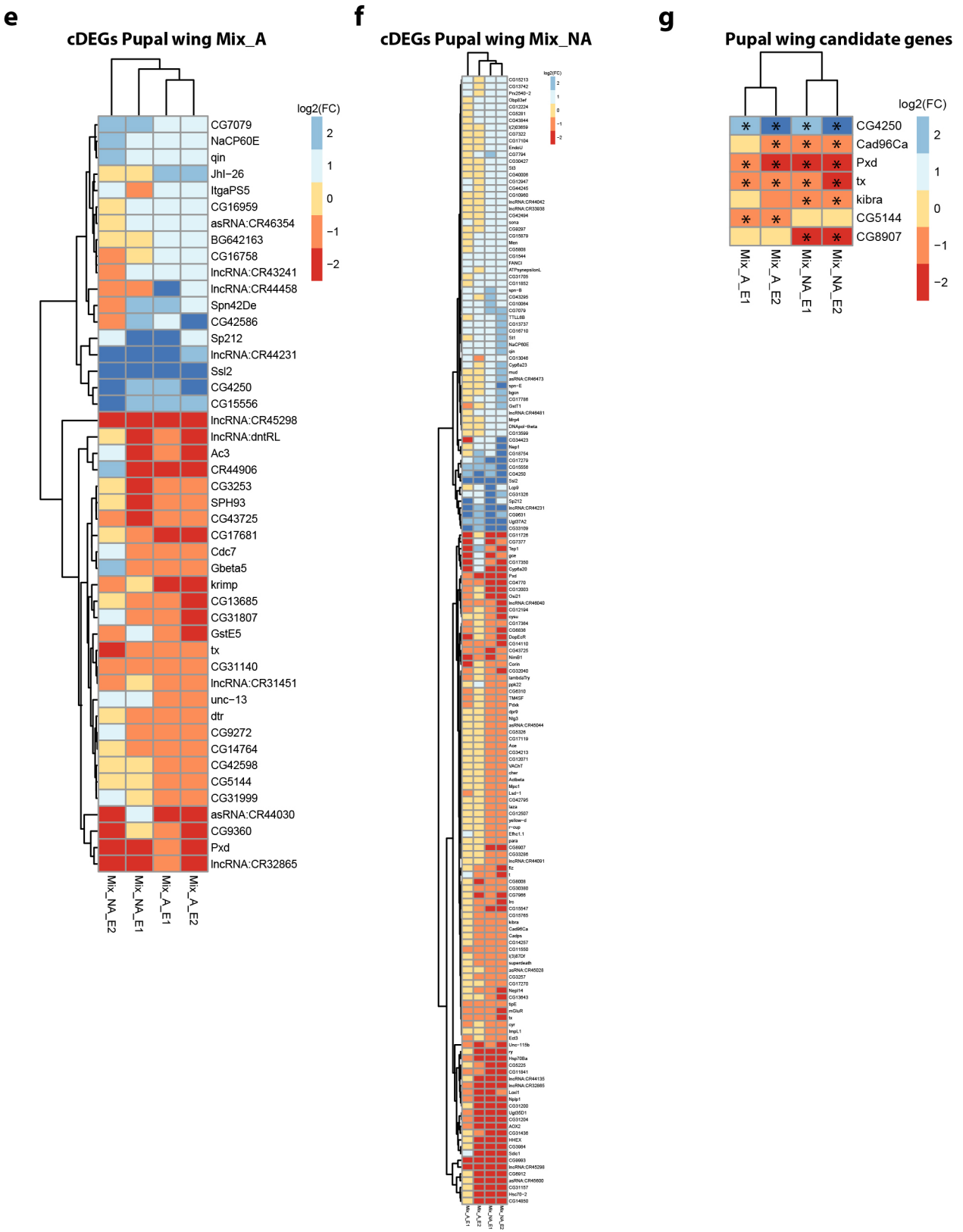


Extended Data Fig. 3: Differentially expressed genes in Mix selection lines in wing disc and pupal wing

Volcano plots showing the number of differentially expressed genes (DEGs) for the Mix assimilated and non-assimilated (for both selection experiments) relative to control lines in wing disc (WD) and pupal wing (PW). DEGs cut-off (DESeq2): adjusted p-value<0.05 and |log2FC|>0.58.

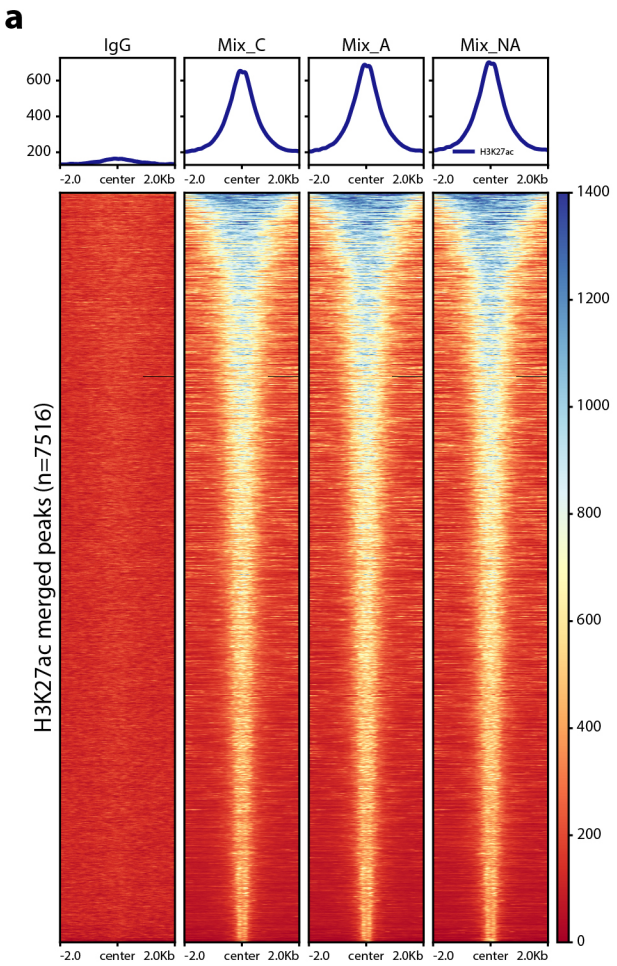
Extended Data Fig. 4

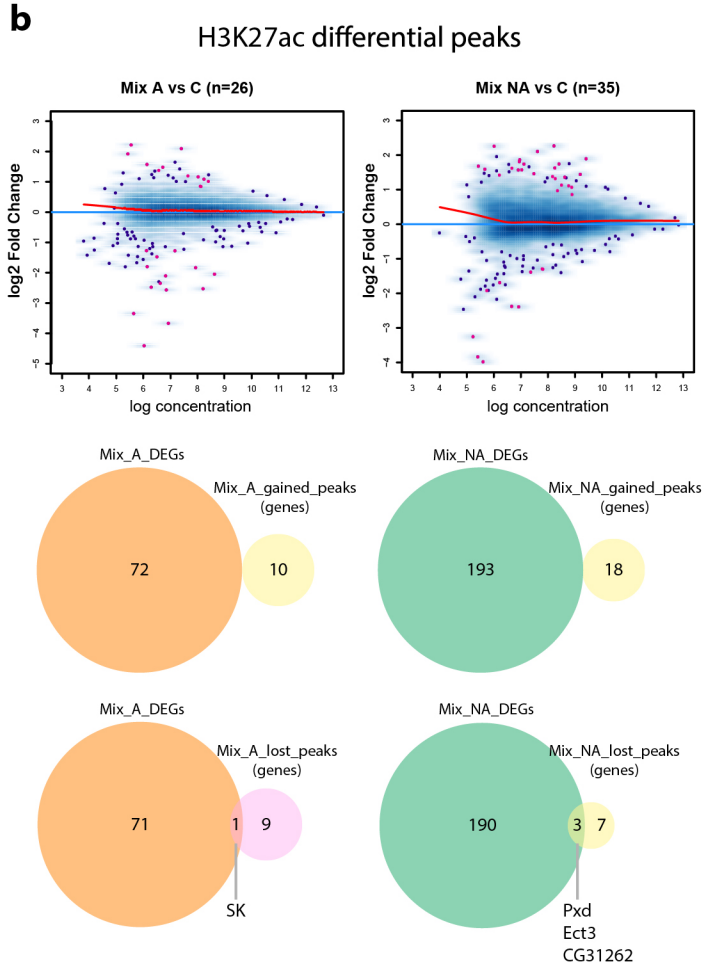


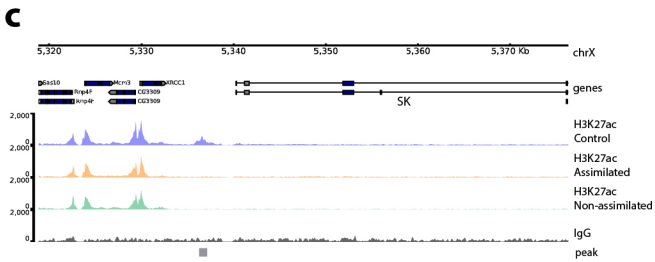

e
cDEGs Pupal wing Mix_A
log2(FC)
2
1
0
−1
−2
CG7079
NaCP60E
qin
Jhl−26
ItgaPS5
CG16959
asRNA:CR46354
BG642163
CG16758
lncRNA:CR43241
lncRNA:CR44458
Spn42De
CG42586
Sp212
lncRNA:CR44231
Ssl2
CG4250
CG15556
lncRNA:CR45298
lncRNA:dntRL
Ac3
CR44906
CG3253
SPH93
CG43725
CG17681
Cdc7
Gbeta5
krimp
CG13685
CG31807
GstE5
tx
CG31140
lncRNA:CR31451
unc−13
dtr
CG9272
CG14764
CG42598
CG5144
CG31999
asRNA:CR44030
CG9360
Pxd
lncRNA:CR32865
Mix_NA_E2
Mix_NA_E1
Mix_A_E1
Mix_A_E2
f
cDEGs Pupal wing Mix_NA
log2(FC)
2
1
0
−1
−2
Mix_A_E1
Mix_A_E2
Mix_NA_E1
Mix_NA_E2
g
Pupal wing candidate genes
log2(FC)
2
1
0
−1
−2
CG4250
Cad96Ca
Pxd
tx
kibra
CG5144
CG8907
Mix_A_E1
Mix_A_E2
Mix_NA_E1
Mix_NA_E2

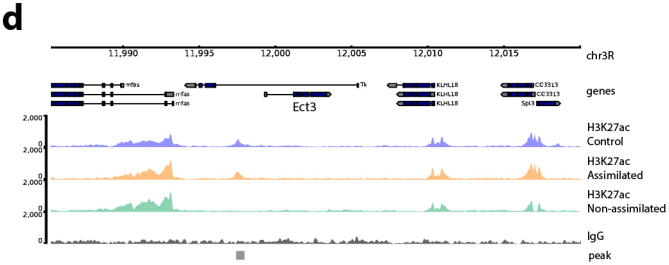
Extended Data Fig. 4: Candidate genes for Mix assimilated and non-assimilated selection lines
a,b- Heatmap with *k*-means clustering analysis of consistently differentially expressed genes (cDEGs) for both selection experiments in wing discs of Mix assimilated (A) (**a-**) and non-assimilated (NA) (**b-**) lines. **c-** Shortlisted candidate genes in wing disc. Asterisks indicate significant differences in expression compared to control (adjusted p-value<0.05 and |log2FC|>0.58). **d-** Expression level of the *Cad96Ca* gene in the wing disc (WD) of Mix control (C), assimilated (A) and non-assimilated (NA) flies analyzed by RT-qPCR. The plot shows the averaged log2 fold change normalized to control flies and error bars represent the standard error of the mean (SEM) from four biological replicates. Significance was calculated using one-way ANOVA with Dunnett's multiple comparisons test relative to control (*p<0.05, **p<0,01). **e,f-** Heatmap with *k*-means clustering analysis of consistently differentially expressed genes (cDEGs) for both selection experiments in pupal wings of Mix assimilated (A) (**e-**) and non-assimilated (NA) (**f-**) lines. **g-** Shortlisted candidate genes in pupal wings. Asterisks indicate significant differences in expression compared to control (adjusted p-value<0.05 and |log2FC|>0.58).

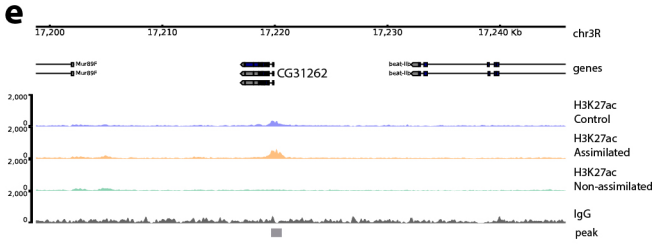
Extended Data Fig. 5







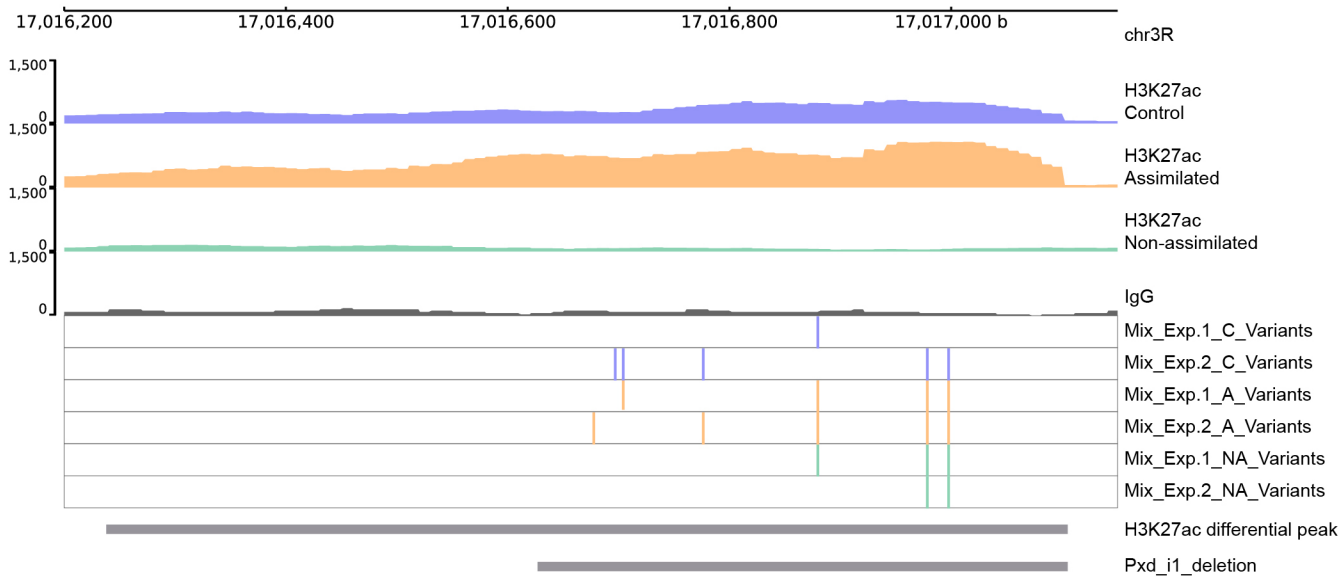




Extended Data Fig. 5: Genome-wide differential enrichment analysis of the H3K27ac mark in the Mix selection lines

a- Heat-map of CUT&RUN merged H3K27ac peaks (n=7516) in the wing disc of Mix control, assimilated and non-assimilated lines. **b-** MA plots showing the number of differentially enriched H3K27ac peaks in Mix assimilated (A) and non-assimilated (NA) relative to control (C) (three replicates per line, FDR<0.05 and |log2FC|>0.58). Overlap of the genes associated with the differential peaks with the consistent differentially expressed genes in each selection line. **c:e-** CUT&RUN profiles for H3K27Ac (merge of three replicates) and IgG (control) in the wing disc for the Mix lines in the *SK* (**c**), *Ect3* (**d**) and *CG31262* (**e**) loci. The grey boxes indicate the H3K27ac differential peaks.

Extended Data Fig. 6

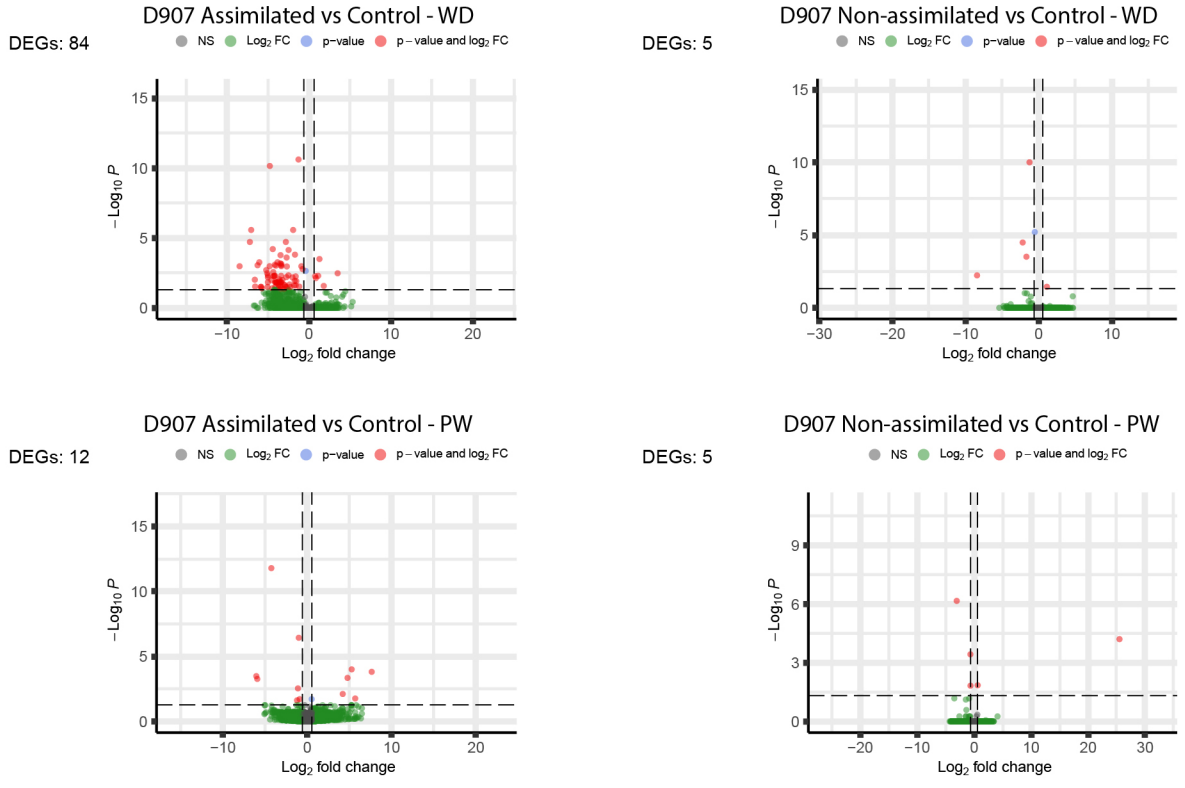


Extended Data Fig. 6: Absence of exclusive cis-mutation in Mix non-assimilated lines associated with the loss of H3K27ac enrichment at the Pxd locus

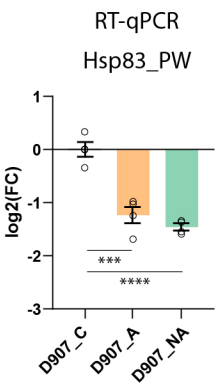
Detail view of the differential H3K27ac peak for the Mix non-assimilated flies at the *Pxd* locus. CUT&RUN profiles for H3K27ac (merge of three replicates) and IgG (control) in the wing disc. The positions of the significant genetic variants (SNPs and InDels) for all Mix lines are shown as vertical lines. The grey boxes indicate the H3K27ac peak and the position of the CRISPR/Cas9 targeted deletion (*Pxd_i1_deletion*).

Extended Data Fig. 7

a



b

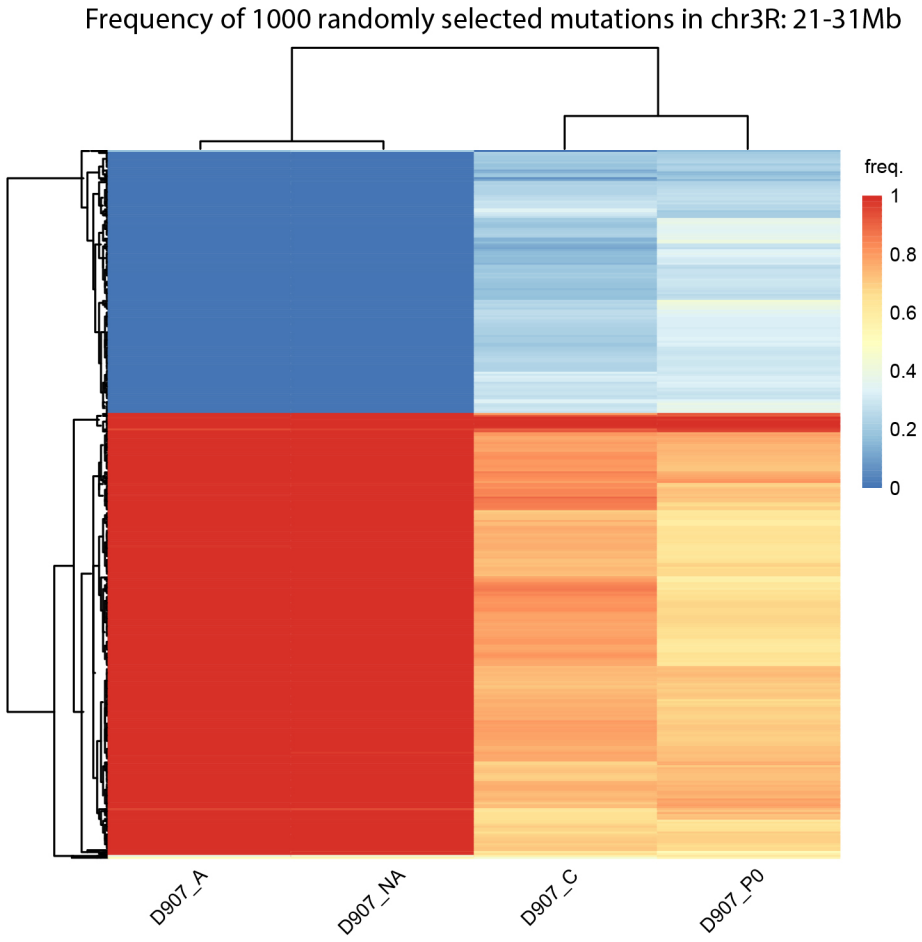


Extended Data Fig. 7: Differentially expressed genes in D907 selection lines in wing disc and pupal wing

a- Volcano plots showing the number of differentially expressed genes (DEGs) for the D907 assimilated and non-assimilated relative to control lines in wing disc (WD) and pupal wing (PW). DEGs cut-off (DESeq2): adjusted p-value<0.05 and |log2FC|>0.58. **b-** Expression level of the *Hsp83* gene in the pupal wing (PW) of D907 control (C), assimilated (A) and non-assimilated (NA) flies analyzed by RT-qPCR. The plot shows the averaged log2 fold change normalized to control flies and error bars represent the standard error of the mean (SEM) from four biological replicates. Significance was calculated using one-way ANOVA with Dunnett's multiple comparisons test relative to control (***p<0.001, ****p<0,0001).

Extended Data Fig. 8

a



b

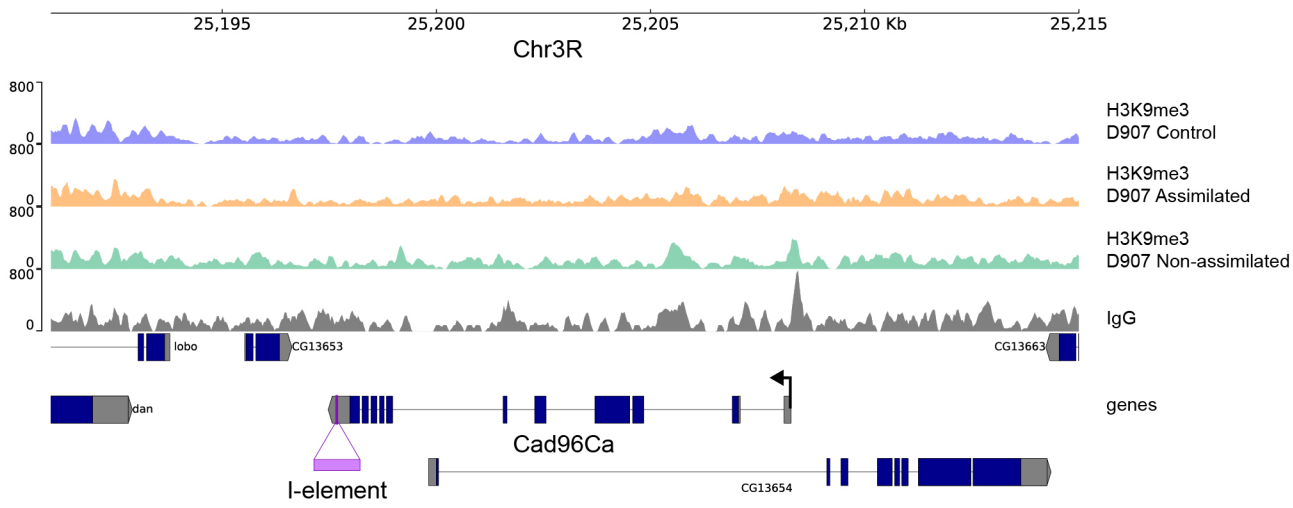
Frequency of mutations in the counter-selected D907_A and D907_NA alleles

	D907_A	D907_C	D907_NA	D907_P0
Mean	0	0.226	0	0.302

Extended Data Fig. 8: Frequency estimation of In(3R)C in the different D907 lines

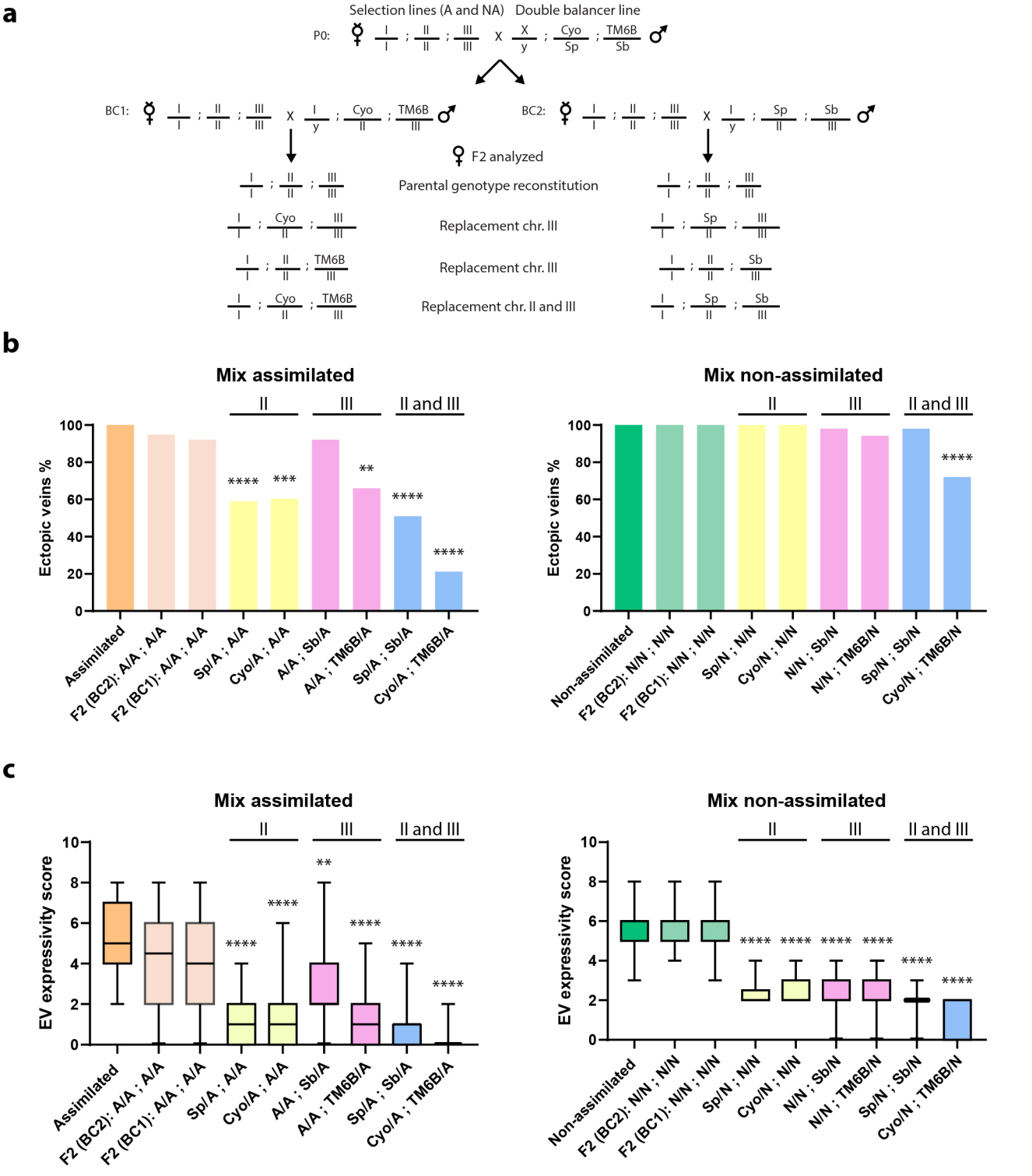
a- Heat map showing the frequency of a thousand randomly selected mutations (SNPs and InDels) in chr3R: 21-31Mb from D907 parental (D907_P0), control (D907_C), assimilated (D907_A) and non-assimilated (D907_NA) fly populations. **b-** Table showing the average of the frequency of the counter-selected mutations (frequency = 0) in D907 A and NA in this chromosome region.

Extended Data Fig. 9



Extended Data Fig. 9: H3K9me3 enrichment at the Cad96Ca locus in the D907 selection lines
CUT&RUN tracks for H3K9me3 (merge of two replicates) and IgG (control) in the wing disc of D907 lines at the *Cad96Ca* locus. The *I-element* insertion mapped to the 3'UTR of the *Cad96Ca* gene is shown.

Extended Data Fig.10



Extended Data Fig. 10: Contributions of the major autosomal chromosomes to the EV evolution in Mix selection lines

a- We analyzed the contribution of the large somatic second and third chromosomes to the EV phenotype using chromosome heterozygous replacements in the selected lines. The crossing scheme represents the genetic setup used to analyze the contribution of single and double heterozygous replacements of the second and third chromosomes to the EV phenotype of the Mix selection lines. b- Penetrance of ectopic veins trait in adult females of the different genotypes analyzed in the F2 progeny. In Mix assimilated, heterozygous replacement of both the second and third chromosomes caused a significant decrease in EV penetrance and expressivity, whereas in Mix non-assimilated, neither of these heterozygous replacements affected EV penetrance. c- expressivity of ectopic veins trait in adult females of the different genotypes analyzed in the F2 progeny. The statistical differences in EV penetrance and expressivity were calculated using a Chi-squared test for pairwise comparisons with the respective control (the F2 with the selection line genotype reconstitution for each backcross). Taken together, these studies suggest that alleles located on both the second and third chromosomes are involved in the evolution of the EV trait in the Mix selection lines, although the contribution of alleles harbored on each chromosome are not equivalent in the Mix assimilated and non-assimilated lines.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ExtendedDataTable1.xlsx
- ExtendedDataTable2.xlsx
- ExtendedDataTable3.xlsx
- ExtendedDataTable4.xlsx
- ExtendedDataTable5.xlsx
- ExtendedDataTable6.xlsx
- ExtendedDataTable7.xlsx