

miR-9a Minimizes the Phenotypic Impact of Genomic Diversity by Buffering a Transcription Factor

Justin J. Cassidy,¹ Aashish R. Jha,² Diana M. Posadas,¹ Ritika Giri,¹ Koen J.T. Venken,^{3,4} Jingran Ji,¹ Hongmei Jiang,⁵ Hugo J. Bellen,³ Kevin P. White,² and Richard W. Carthew^{1,*}

¹Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

²Institute for Genomics and Systems Biology, Departments of Human Genetics and Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

³Department of Molecular and Human Genetics, Howard Hughes Medical Institute, Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, USA

⁴Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX 77030, USA

⁵Department of Statistics, Northwestern University, Evanston, IL 60208, USA

*Correspondence: r-carthew@northwestern.edu

<http://dx.doi.org/10.1016/j.cell.2013.10.057>

SUMMARY

Gene expression has to withstand stochastic, environmental, and genomic perturbations. For example, in the latter case, 0.5%–1% of the human genome is typically variable between any two unrelated individuals. Such diversity might create problematic variability in the activity of gene regulatory networks and, ultimately, in cell behaviors. Using multigenerational selection experiments, we find that for the *Drosophila* proneural network, the effect of genomic diversity is damped by miR-9a-mediated regulation of senseless expression. Reducing miR-9a regulation of the Senseless transcription factor frees the genomic landscape to exert greater phenotypic influence. Whole-genome sequencing identified genomic loci that potentially exert such effects. A larger set of sequence variants, including variants within proneural network genes, exhibits these characteristics when miR-9a concentration is reduced. These findings reveal that microRNA-target interactions may be a key mechanism by which the impact of genomic diversity on cell behavior is damped.

INTRODUCTION

A central goal in biology and medicine is to understand the relationship between an individual's genome and his/her disease susceptibility. An individual typically carries loss-of-function alleles in about 300 genes, and 15%–30% of these are classified as causing inherited disorders (1000 Genomes Project Consortium et al., 2010). Unrelated individuals typically have genomes that differ from one another by approximately 0.5%–1.0% with respect to copy number and sequence, any of which might cause alterations in traits, ranging from physical charac-

teristics to disease susceptibility (Conrad et al., 2010; Frazer et al., 2009). The relationship between an individual's genome and his or her phenotype is highly complex, and the mechanisms that regulate the relationship are poorly understood (Gibson and Wagner, 2000).

Although it has been hypothesized that mechanisms exist to actively repress the impact of genomic variation on phenotypic outcome (de Visser et al., 2003; Gibson and Dworkin, 2004; Hornstein and Shomron, 2006; Masel and Siegal, 2009), these have for the most part remained a mystery. Such mechanisms could have profound effects on the contributions that genome variation makes to disease susceptibility. Moreover, impairment of these mechanisms might increase disease risk by elevating the impact of existing genomic variation, making their identification of high interest and significance.

Because genomic variation likely has its most primary effect on gene expression, we hypothesized that counteracting mechanisms might operate at this level. We have focused our attention on microRNAs (miRNAs) as potential mediators of such mechanisms. They are well suited to potentially dampen the impact of genomic variation because they regulate a majority of protein-coding genes via posttranscriptional repression (Bartel, 2009). They are common components of negative feedback and feed-forward regulatory motifs with their gene targets (Ebert and Sharp, 2012; Tsang et al., 2007). These motifs can confer homeostasis to target protein levels, thereby buffering variation in upstream processes such as chromatin dynamics, transcription, and splicing. miRNAs exert modest repression on their targets, but they act rapidly to change protein synthesis (Carthew and Sontheimer, 2009). This restrained yet rapid action makes them particularly effective for regulating homeostasis. When any significant drift from the desired steady state of a target protein leads to pathological consequences, miRNAs may counteract such outcomes.

The links between human disease and miRNAs are extensive (Mendell and Olson, 2012). However, these links by themselves offer no evidence that any of the connections are due to

desuppression of genomic variation. A controlled experimental approach in a model system must be applied to determine whether such connections are possible. As a proof of principle, we have performed experiments to measure the impact of miRNAs on buffering genomic diversity. We chose to use *Drosophila* because its mechanism of miRNA regulation is highly conserved with humans (Carthew and Sontheimer, 2009). Moreover, many miRNAs are conserved between the two species, including the miRNAs we have studied (Christodoulou et al., 2010). We chose a nonpathological phenotypic outcome that occurs robustly and quantitatively to allow for sensitive detection of variation. To measure the impact of miRNAs on buffering genomic diversity, we performed multigeneration selection experiments. We show how the interaction between the miR-9a miRNA and its target gene called *senseless* (*sens*) can make a cell phenotype more resistant to genomic variation. We used whole-genome sequencing to identify candidate variants that have affected the phenotype. When miR-9a concentration is reduced, a broader group of variants appears to influence the phenotype. Thus, we have established a precedent for miRNAs to act in a mechanism that counteracts variation between different individual genomes.

RESULTS

Sens Regulation by miR-9a Generates Robustness

Determination of sensory cell fate is controlled by a network of regulatory factors in *Drosophila* (Figure 1A). The core circuit comprises three transcription factors required for the fate switch: Achaete (Ac), Scute (Sc), and Sens (Quan and Hassan, 2005). A subset of ectodermal cells is endowed with a combination of signaling inputs and transcription factors such that ac and sc transcription is established. Ac and Sc are bHLH proteins that form heterodimers with Daughterless (Da) and activate transcription of target genes. One of these target genes encodes Sens, which feeds back to regulate transcription of the ac and sc genes (Nolo et al., 2000). The switch to a sensory cell fate depends on amplification of subtle initial differences in these transcription factors between putative precursor cells leading to a binary switch in fate (Acar et al., 2006; Jafar-Nejad et al., 2003). Amplification of these differences is achieved by coupling of positive feedback with lateral inhibition (Quan and Hassan, 2005).

The precision of the fate switch is influenced by miR-9a and its interaction with *sens*. Loss of miR-9a results in more variable formation of sensory organs on the adult scutellum (Li et al., 2006) (Figure 1B; Table S1 available online). This effect is suppressed by reducing the expression of *sens*, suggesting that fate precision requires *sens* regulation (Figure 1C). There are two conserved binding sites for miR-9a in the 3' UTR of the *sens* transcript (Figure 1D), and miR-9a can inhibit *sens* gene expression in cell culture (Li et al., 2006). To more fully probe the interaction between miR-9a and *sens* in vivo, we incorporated a single copy of the *sens* gene into a transgenic landing site on chromosome II. The transgene contained all known endogenous *sens* regulatory regions, and thus, its protein expression pattern was identical to the endogenous *sens* pattern (data not shown). Moreover, the transgene rescued the loss of the endogenous *sens* gene. We also generated *sens* transgenes with one or both miR-9a binding sites mutated within the 3' UTR. We compared the effects of

each transgene on scutellar sensory organ (SSO) fate when the endogenous *sens* gene was absent. The *sens* transgenes with mutated miR-9a binding sites showed increased variability in SSO fate (Figure 1E), similar to what was observed when miR-9a was lost. However, the mutated transgenes did not further increase fate variability when miR-9a was absent (Figure 1F). This result argues that the binding sites exclusively mediate the effects of miR-9a on reducing fate variability. To further test this, we reduced *miR-9a* copy number in individuals who had two copies of the *sens* transgene in addition to the endogenous *sens* gene. This increased SSO number when the *sens* transgene was fully wild-type but had no significant effect on SSO number when both miR-9a binding sites were mutated (Figure 1G). The *sens* transgenes with only one mutated site had greater SSO numbers when *miR-9a* copy number was reduced, suggesting that both sites mediate miR-9a regulation of *sens*. In summary, miR-9a regulation of *sens* generates a robust outcome and suppresses noise in SSO fate determination.

How this occurs must be considered within the broader context of miRNA regulation. Modeling and experimental analysis indicates that a simple miRNA-target interaction can create a threshold in the relationship between target mRNA and protein synthesis (Mukherji et al., 2011). Below the transcription threshold, protein synthesis is silenced by the miRNA; above the threshold, protein output becomes highly sensitive to transcription (Figure 1H). Because miRNA is titrated away by an increasing number of transcripts, repression is weakened until protein output becomes sensitive to transcription. In this context, miR-9a might create a threshold for Sens protein output such that variable *sens* transcription is prevented from inappropriately triggering the positive feedback circuit (Figure 1A).

Robust Response to Temperature Variation Requires

Sens Regulation by miR-9a

Network stability during development could be due to buffering of stochastic events, environmental variation, underlying genomic variation, or some combination of the three (Félix, 2012). We first considered the impact of miR-9a-regulated *sens* on environmental variation. For this, we sought to sensitize the fate switch to temperature because under normal conditions, physiological temperatures between 18°C and 29°C have little effect. In a *sc¹* genetic background, temperature has a quantifiable effect on fate even though *sc¹* does not alter the Sc protein sequence (Campuzano et al., 1986; Sheldon, 1968). The probability of SSO fate is inversely proportional to temperature (Figure 2A). We constructed four *sc¹* strains, each strain replacing the endogenous *sens* gene with a *sens* genomic rescue transgene. One strain had the transgene with intact miR-9a binding sites, two strains had one or the other miR-9a binding site intact, and one strain had neither binding site intact. Progressive loss of miR-9a binding sites caused a progressive enhancement of the temperature effect on SSO fate (Figure 2B; Table S2). We also altered the regulatory circuit in a different way. We reduced the copy number of the *miR-9a* gene from two to one and saw a similar enhancement of the temperature effect on SSO fate (Figure 2C; Table S3). This behavior shows the importance of *sens* regulation by miR-9a on the robustness of the cell fate switch against variation in temperature.

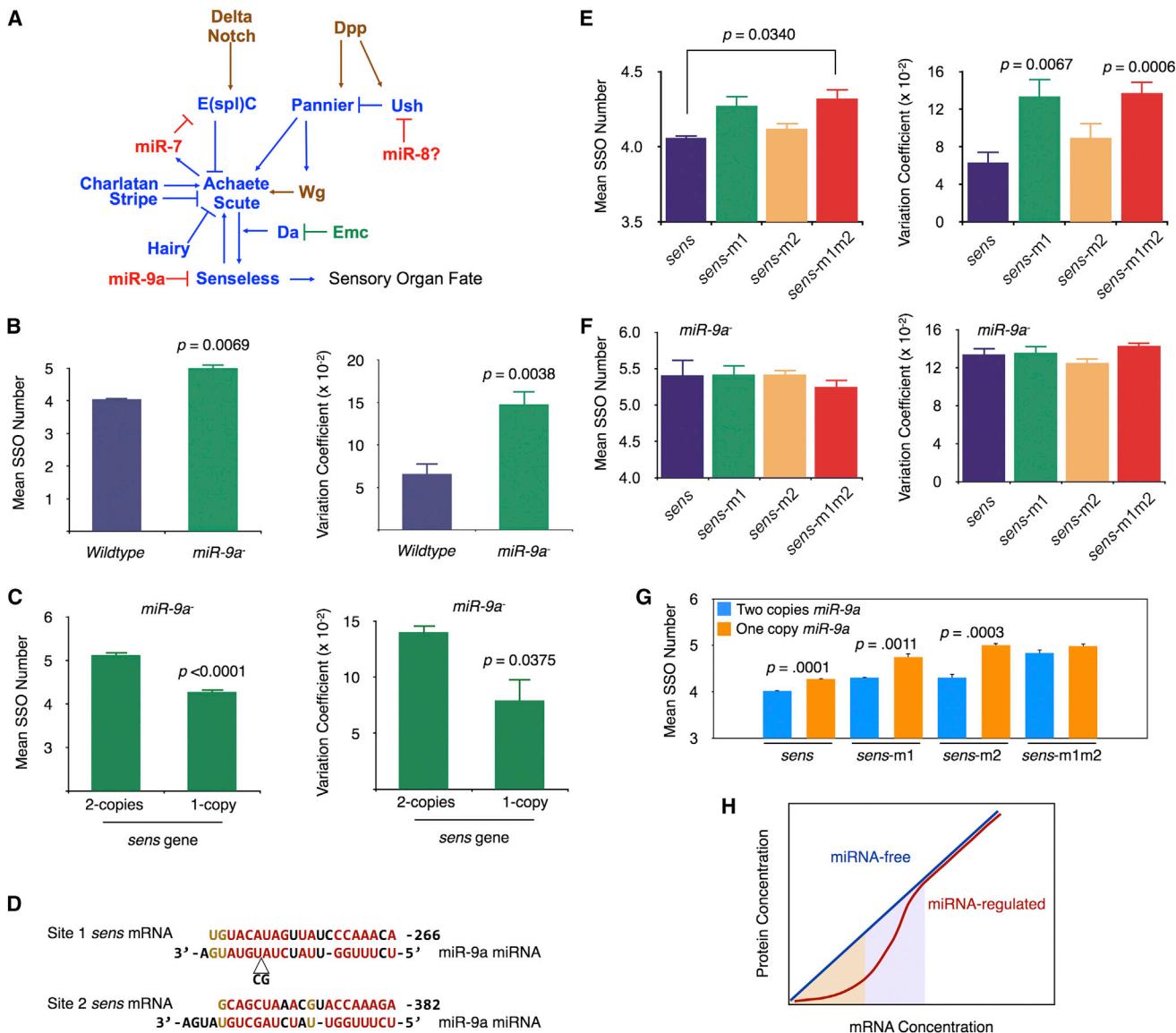


Figure 1. Sens Regulation by miR-9a Affects Precision of the Sensory Cell Fate Switch

- (A) The sensory organ network is shown. Signaling factors are in brown, transcription factors are in blue, protein modifiers are in green, and miRNAs are in red.
- (B) SSO number (left) and variation (right) as a function of *miR-9a* are shown.
- (C) SSO number (left) and variation (right) in a *miR-9a* null mutant that has either one or two copies of the wild-type *sens* gene are shown.
- (D) Predicted binding sites for *miR-9a* in the *sens* 3' UTR based on base pairing and conservation are shown. Red highlights Watson-Crick pairing, and brown highlights wobble pairing.
- (E) SSO number (left) and variation (right) as a function of the *sens* transgene are shown. Each individual had two copies of the *sens* transgene, and its endogenous *sens* gene was nullified by mutation. The *miR-9a* binding sites are mutated at site 1 (m1), site 2 (m2), or both sites (m1m2). *p* values are shown for differences between mutant and wild-type transgenes.
- (F) SSO number (left) and variation (right) in *miR-9a* null mutants that also contained either wild-type or mutant *sens* transgenes, as indicated, are shown.
- (G) SSO number in individuals with one or two copies of the *miR-9a* gene and that also contained either wild-type or mutant *sens* transgenes, as indicated, is shown. Individuals with the *sens-m1* or *sens-m2* transgene showed ectopic SSOs when *miR-9a* gene copy number was reduced to one. Because *sens* could only be regulated by *miR-9a* through site 1 or 2 in these individuals, it argues that *miR-9a* can act through either site. Individuals with the *sens-m1m2* transgene showed no ectopic SSOs when *miR-9a* gene copy number was reduced to one because *miR-9a* could not act when both sites were impaired.
- (H) Hypothetical mechanism for gene regulation by a miRNA is shown. Nonlinear protein output from gene transcripts under the regulation of a miRNA will generate a threshold (purple), below which output is insensitive to fluctuations in transcript concentration (brown). Above the threshold, output is fully responsive to transcript concentration. Such thresholding when coupled to feedback will create switch-like behavior in regulatory networks.
- For all figure plots, SE is shown, and significant *p* values shown are from two-tailed Student's *t* tests. *p* values not shown were not significant (*p* > 0.05). See also Table S1.

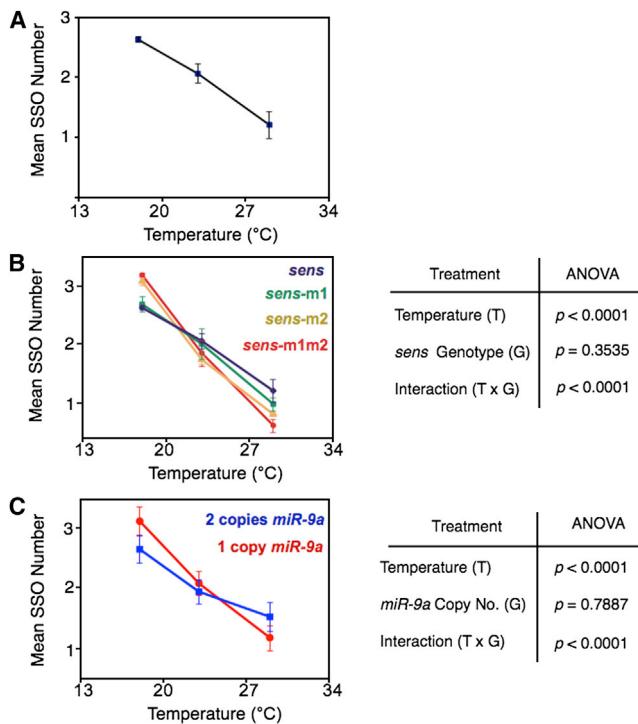


Figure 2. Regulation of *sens* by miR-9a Generates Temperature Robustness to the SSO Fate

(A) Temperature dependence of SSO fate in a *sc⁷* background is shown.

(B) Mutation of site 1 (m1), site 2 (m2), or both sites in the *sens* transgene in a *sc⁷* background enhances the temperature dependence of SSO fate to different extents. Error bars are SEs. To the right is an ANOVA showing temperature, *sens* genotype, and interaction effects.

(C) Reduction of miR-9a copy number enhances the temperature dependence of SSO fate in a *sc⁷* background. Error bars are corrected SEs. To the right is an ANOVA showing temperature, miR-9a genotype, and interaction effects. The miR-9a gene inhibited SSO formation at low temperature and promoted it at high temperature. Possibly, at low temperature, the rate-limiting step to form an SSO directly involves Sens. At high temperature, the rate-limiting step may shift to another node in the network, that inhibits SSO determination. Sens activation of this node would then describe the results.

See also Figure S5, and Tables S1, S2, S3, and S4.

Response to Genomic Diversity Requires Sens Regulation by miR-9a

We next focused on the importance of the miR-9a/Sens regulatory circuit on buffering genomic diversity. Almost 4% of sites in the genome have single-nucleotide variation in outbred *D. melanogaster* populations (Mackay et al., 2012). Similar variation is observed in the genomes of other animal species (1000 Genomes Project Consortium et al., 2010). Typically, multiple unlinked variants act in combination to modify developmental, physiological, and disease outcomes (Frazer et al., 2009). Genome-wide association studies of individuals or selection studies of populations are used to detect variants of this type. The impact of genetic variation on *Drosophila* SSO formation has been under investigation since 1959, making it an ideal model for this sort of analysis (Rendel, 1959). In a sensitized *sc⁷* background, animals subjected to selection for the highest number of SSOs transmit a heritable tendency for their off-

spring to have more SSOs (Rendel, 1959). When selection is repeated generation after generation, the population progressively evolves, and the rate of evolution is a measure of the contribution that genomic diversity has on the variation in SSO fate. A low rate is seen when genomic diversity has little impact on outcome variation, whereas a high rate is seen when diversity strongly affects variation in the cell fate switch.

To probe the relationship between genomic diversity and miR-9 regulation, we used *sc⁷* strains with one copy of the *sens* transgene (Figures S1 and S2; Table S1). The transgene had either intact or disrupted miR-9a binding sites in its 3' UTR. The animals also contained their normal copies of *sens* on chromosome III. We maintained replicate populations of individuals of well-defined number at uniform temperature to minimize the effects of population size and temperature on fate variation (Sheldon, 1968). We selected individuals who had the highest number of SSOs (above the 50th percentile) and allowed these individuals to produce the next generation. This selection was repeated for several generations on the four replicate populations in each strain.

Realized heritability (h^2) measures the proportion of total variation in outcome that is attributable to the effects of the genome (Falconer and Mackay, 1996). A maximal h^2 value of one would occur if genomic diversity perfectly dictated the observed variation in outcome. An h^2 of zero occurs when no genomic component affects the variation in outcome. In our case, h^2 is then a measure of the impact of genomic diversity on variation in SSO fate. We quantified the cumulative response in SSO fate as a function of the cumulative selection pressure over five generations of selection. The response was proportional to the selection pressure, and the slope of the fitted regression line, h^2 , was a constant (Figure 3A). We found that h^2 was dependent on the integrity of miR-9a binding sites in the *sens* transgene. Higher h^2 values were reproducibly observed when miR-9a binding sites were disrupted (Figure 3B). Assuming a uniform genomic landscape between experimental strains at the experiment's beginning, this result implies that *sens* regulation by miR-9a suppresses the impact of genomic diversity on the observed variation in SSO fate.

We could not confirm or disprove the assumption of uniform starting landscapes in our experiment because each transgene was incorporated into a unique founder individual from a common population. Therefore, we sought to measure h^2 in strains with different concentrations of miR-9a. We constructed two new strains: one containing the normal miR-9a gene copy number of two, and the other containing a single gene copy (Figures S1 and S3). Each strain was carefully introgressed with the other to ensure that the genomic landscape of each was well matched except for the region encompassing the miR-9a gene.

Replicate populations for each strain were then independently subjected to selection for individuals with the highest number of SSOs to sire the next generation. Selection was repeated for 15 generations, generating a response in SSO fate that was proportional to selection pressure (Figures 3C and 3D). The strain with a single miR-9a copy had the expected effect, causing an increase in h^2 compared to the situation with two copies of miR-9a (Figure 3E). The ability of two different perturbations (miR-9a

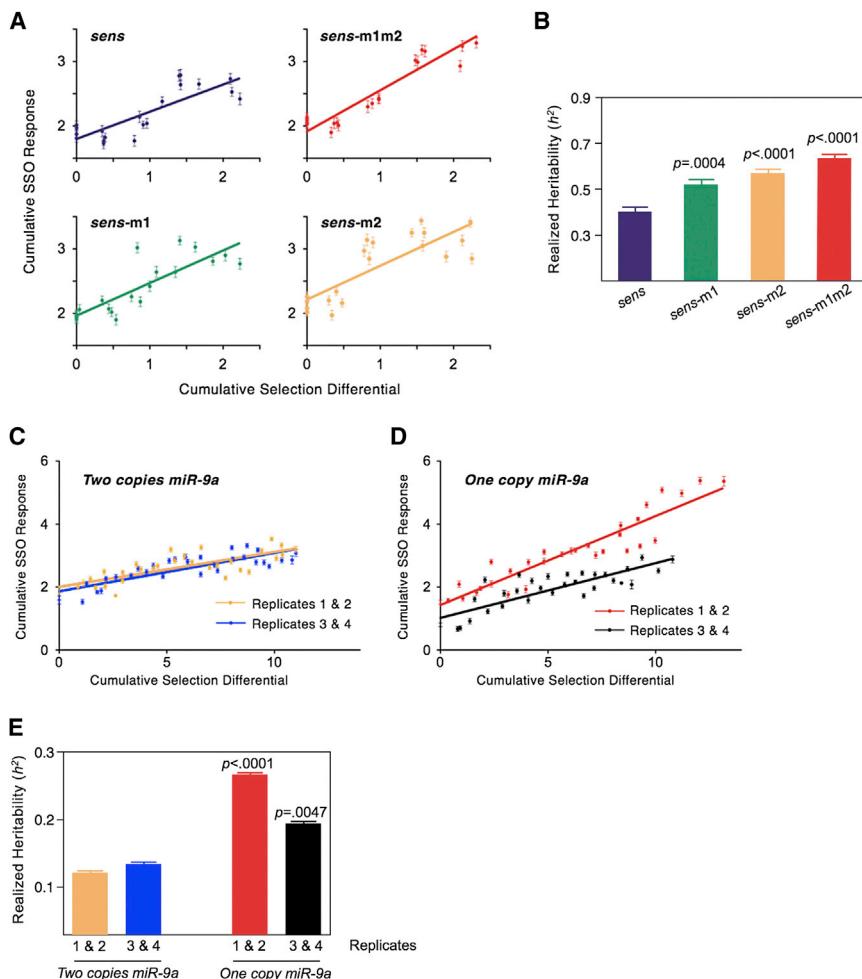


Figure 3. Variation in SSO Fate Caused by Genomic Diversity Is Suppressed by miR-9a

(A) Response of SSO fate to cumulative selection in a *sc¹* background is shown. Strains carried *sens*, *sens-m1*, *sens-m2*, and *sens-m1m2* transgenes, as indicated. Each strain has four replicate populations, which are plotted together. The slopes of the weighted least-squares regression lines are a measure of the h^2 .

(B) h^2 of SSO fate within populations that contain one copy of a *sens* transgene during selection is shown. Significant p values shown are from a Z test. SEs are shown as error bars.

(C) Response of SSO fate to cumulative selection in a *sc¹* background with two copies of the *miR-9a* gene is shown. Replicates 1 and 2 are treated separately from replicates 3 and 4. The slopes of the regression lines are measures of h^2 .

(D) Response of SSO fate to cumulative selection in a *sc¹* background with one copy of the *miR-9a* gene is shown. Replicates 1 and 2 had a stronger response to selection than the other two replicates, as determined by multiple regression analysis, and thus are treated separately.

(E) h^2 of SSO fate within replicate populations undergoing selection that contained either one or two copies of *miR-9a* during selection is shown. p values shown are from Z tests to matched replicates. SEs are shown as error bars. See also Figures S1, S2, S3, and S4 and Table S1.

Diversity at Several Genomic Loci Affects the Cell Fate Switch

A heritable response to selection is predicted to occur when the genomic landscape of a population changes from its initial state during the selection process.

Guided by this prediction, we sought to identify those changes in the genome that occurred during the *miR-9a* selection experiment. We applied shotgun sequencing to pooled genomes sampled before and after the selection experiment (Figure S6A; Table S5). As a further control, we also sequenced genomes from populations that had been randomly selected for 15 generations in parallel with the experimental groups. We achieved a minimum average coverage of 29 reads per genomic site per strain, allowing for variant identification with a maximum false discovery rate of 2% (Table S5). Across all strains and conditions, we identified 730,479 variants (SNPs and indels) in 120.4 Mb of sequence (an average of 6.1 variants per kb of genome).

The genomic landscape of a population is determined by the distribution of variants and the allele frequency for each variant. We measured the allele frequencies for all variants, in a given strain and state of selection. Importantly, the allele frequency distributions between one- and two-copy *miR-9a* strains before selection were highly similar (Figure S6B). Coupled with the observation that the number of variants was comparable between the two strains (Table S5), this observation confirmed that genomic diversity in the one- and two-copy strains was similar prior to selection.

concentration and deregulation of the *miR-9a* target *sens*) to elicit the same effect on h^2 is strong evidence that *miR-9a* regulation of *sens* is important for buffering the impact of genomic diversity on variation in SSO fate.

To test the generality of this miRNA effect on the genome, we collected data for another miRNA called *miR-7*. A set of *miR-7* targets encodes for E(spl) transcription factors that repress *ac* and *sc* transcription (Lai et al., 2005) and, therefore, inhibit the cell fate switch (Figure 1A). We reduced the copy number of the *miR-7* gene in a sensitized *sc¹* background (Figures S1 and S4), which impaired the cell fate switch as predicted (Figure S5A). The strains with one or two copies of *miR-7* were subjected to selection for individuals with the greatest number of SSOs. After 15 generations, the heritable response to selection was quantified as a function of the cumulative selection pressure (Figures S5B and S5C). In these strains, h^2 was not significantly different from each other ($p = 0.101$, Z test). We also grew the two strains at different temperatures, and we found that the effect of temperature on SSO fate was unchanged whether one or two copies of the *miR-7* gene were present (Figure S5D; Table S4). This lack of an effect on temperature- and genome-induced variation highlights the specificity of *miR-9a* on these processes for the cell fate switch.

For each variant, we calculated the difference in allele frequency between the selected and randomly selected groups. We plotted these allele frequency differences for all variants in both replicate populations (Figures 4A and 4B). Most variants had a propensity to modestly change allele frequency in a manner that was not necessarily repeated in replicate populations. These variants were presumably changing because of random genetic drift in each population. However, a small fraction of variants (0.1%–0.2%) showed stronger allele frequency change repeatedly in replicate populations. The intensity and reproducibility of these changes suggest that they possibly occurred because of selection. This type of change was observed in strains with either one or two copies of the *miR-9a* gene (Figures 4A–4C). The reproducible responses may be regarded as a single realization in a large space of possible changes in the genomic landscape.

We focused on the variants that exhibited strong and reproducible changes in allele frequency (Figure 4C). We mapped these variants to the genome and found that they clustered into discrete blocks of varying lengths (Figures 4D and 4E). We expected very few variants on the X chromosome because we had isogenized this chromosome prior to the experiment (Figure S1). As predicted, no variant blocks were observed on the X chromosome. Strains with one or two copies of the *miR-9a* gene showed signs that some blocks overlapped between strains. However, other blocks appeared unique to one strain or the other.

Positive selection usually leads to diversity reduction in those variants that are under selective pressure in a population (Weir, 1996). We estimated diversity for each variant and observed that many of the larger variant blocks had a strong reduction (>5-fold) in diversity after selection (Figure 4F). Some peaks of low diversity were shared between the one- and two-copy *miR-9a* strains, but strain-specific peaks were also observed. Positive selection not only affects the diversity of variants under direct selection but also variants that are linked to them (Smith and Haigh, 1974). We estimated diversity for all identified variants in the genome and found highly localized diversity reduction (Figures 4G, S6C, and S6D). Strikingly, most of these localized regions overlapped with the larger blocks of variants previously identified. Therefore, only a limited number of regions in the genome changed diversity in a manner consistent with a selection response.

Identification of the variants that promote cell fate switching in the selected populations is challenging. Within each block in the genome, there are tens to hundreds of genes. Moreover, only 20%–25% of variants were located in ORFs, and of those, the majority were silent mutations (Table S5). The selective advantage of an allele that promotes heritable change in the experiment might be difficult to reliably detect if assayed in isolation. Therefore, we determined which variants that reduced diversity after selection were located within genes implicated in sensory organ formation (see Extended Experimental Procedures). When considering all 1,233 implicated genes (Figure S6E), only 76 genes contained one or more variants that experienced diversity reduction after selection (Figure 4F). Variants were significantly enriched in candidate genes ($p < 10^{-7}$, binomial test). Variants in 11 candidate genes experienced diversity

reduction in both one- and two-copy *miR-9a* strains. One of these genes provided a potential explanation for the enhanced fate switching after selection. *extra macrochaetae* (*emc*) encodes a repressor of Da and thereby inhibits Ac and Sc (Bhattacharya and Baker, 2011; Van Doren et al., 1991). The two-copy *miR-9a* strain had an additional 24 genes with diversity reduction, and 1 of these genes encodes the Notch ligand Delta. Strikingly, the one-copy strain had 41 additional genes with diversity reduction. Several of these genes encode proneural network factors (Figure 1A). The product of the *u-shaped* (*ush*) gene indirectly represses Ac and Sc (Cubadda et al., 1997), *Hairless* (*H*) antagonizes Notch signal transduction and promotes the cell fate switch (Bang et al., 1995), and *charlatan* (*chn*) directly activates *ac/sc* transcription (Escudero et al., 2005).

When compared to the strain with two copies of *miR-9a*, the one-copy strain had 2.5-fold more variants that experienced diversity reduction after selection (Figure 4F). The one-copy strains also exhibited 2- to 3-fold greater h^2 (Figure 3E). These results indicate that when *miR-9a* concentration was reduced, a much larger fraction of genomic variation was able to influence cell fate, causing greater phenotypic variation.

Genomic Landscape Affects the Fate Switch when *miR-9a* Regulation of *sens* Is Impaired

To further test whether the genomic landscape is buffered by *miR-9a*-regulated *sens*, we examined cell fate switching when *sens* regulation is impaired in different genomic landscapes. We placed *sens* transgenes with mutated *miR-9a* binding sites into the genomic landscape created by 15 generations of selection. This action enhanced the effect of the selected landscape on cell fate when compared to the randomly selected (non-selected) landscape (Figure 5A). In contrast, placing the wild-type *sens* transgene into the selected genomic landscape did not significantly enhance the landscape's effect on cell fate. Our interpretation of these results is that genomic landscape has a greater impact on the cell fate switch if *sens* regulation is crippled.

Having determined that *sens* regulation by *miR-9a* can buffer the genome, we tested whether *miR-9a* concentration plays some role in this buffering. We constructed eight strains: four containing one to two copies of the *miR-9a* gene in an unselected landscape, and four strains containing one to two copies of *miR-9a* in a selected landscape. Introduction of the selected landscape had the expected effect, enhancing the switch in cell fate. Impaired buffering of this genomic landscape was expected to occur when *miR-9a* concentration dropped, and this was precisely what we observed (Figure 5B). Strains with one copy of *miR-9a* were affected by the selected landscape to a greater extent than strains with two copies of *miR-9a*.

This concentration dependence was expected to be mediated by *sens*. Accordingly, we examined *sens* expression at the time of the cell fate switch. The strain with two copies of *miR-9a* had a normal expression pattern of Sens protein, even in a selected landscape (Figures 6A and 6B). The strain with one copy showed ectopic Sens expression that completely depended on the selected landscape (Figures 6C and 6D). Thus, *miR-9a* concentration affected the potential for the genomic landscape to promote Sens expression. Interestingly, the genomic landscape also affected noise in SSO fate determination (Figure S7). The

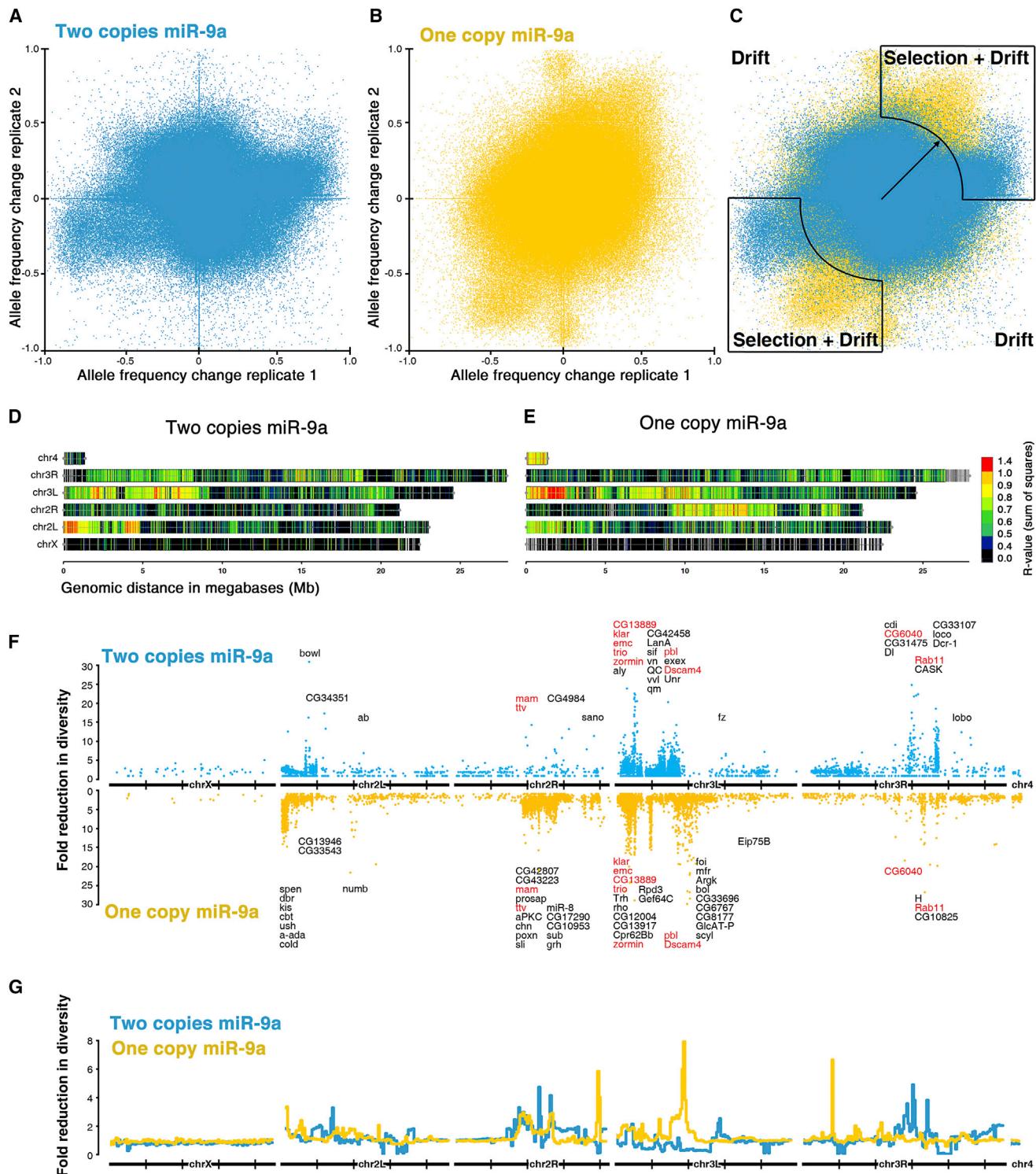


Figure 4. The Genomic Landscape Evolves during Selection in a miR-9a-Dependent Manner

(A–C) Allele frequency difference between selected and randomly selected populations, scatterplotting all genomic variants for two replicate populations, is shown.

(A) Populations carrying two copies of the *miR-9a* gene are presented.

(B) Populations carrying one copy of the *miR-9a* gene are shown.

(C) Merge of the two plots is presented. Outlined are regions that define variants where frequency differences are positively correlated between replicates. A radial threshold (arrow) called *R* was set by the magnitude of frequency differences in both replicates. The derivation of *R* is provided in [Supplemental Information](#).

(legend continued on next page)

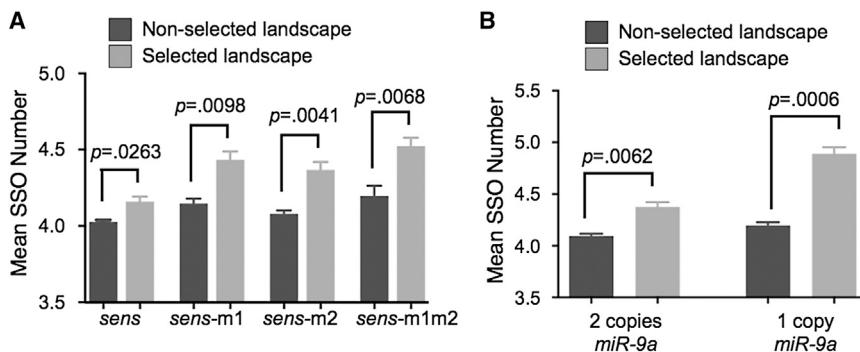


Figure 5. miR-9a Regulation of sens Affects the Impact of Genomic Landscape

(A) One copy of the *sens* transgene was placed into a *sc⁺* genomic landscape in which one copy of the genome had been shaped by 15 generations of selection for more SSOs (Selected). As a control, the transgene was placed into a *sc⁺* landscape where neither genome copy had been selected (Nonselected). Transgenes had either normal or mutated miR-9a binding sites as indicated. Resulting individuals were scored for SSOs. Error bars represent SE. p values are the results of a two-tailed Student's t test comparing selected and nonselected pairs.

(B) Individuals in which both copies of their genomes had been shaped by 15 generations of selection for more SSOs (Selected) or not (Nonselected) are shown. Resulting *sc⁺* individuals had either one or two copies of the *miR-9a* gene. These were scored for SSOs. Error bars represent SE. p values are the results of a two-tailed Student's t test comparing selected and nonselected pairs. See also Figure S7 and Table S1.

selected landscape enhanced SSO variability, suggesting that the genome variants under selection decanalized the fate switch.

Expression of other factors in the network might also be affected by miR-9a, albeit indirectly. The transcription factor Hairy represses *ac/sc* transcription (Ohsako et al., 1994). In one replicate population, the strain with one copy of *miR-9a* altered its Hairy expression after selection (Figures 6E–6G). Expression was reduced in nonneurogenic cells, possibly accounting for some of the enhanced fate determination in this group.

miR-9a Concentration Affects the Variability of Cell Fate between Natural Genomic Landscapes

We had tested two genomic landscapes for their sensitivity to miR-9a. To test many more landscapes for sensitivity, we examined 32 wild strains of *D. melanogaster* collected from 13 locations on five continents (Figure 7A; Table S6). We anticipated that the genomic landscapes of these geographically isolated wild strains would be different from one another, and each strain's genome would affect SSO fate in a different way. *sc¹* and *miR-9a* mutations were crossed into each wild strain, and serial backcrossing was performed until the recombinant strains had 95% of their genomes derived from the parental wild strains but still retained the mutations. *sc¹* individuals with two copies of the intact *miR-9a* gene had an average SSO fate ranging between 0.9 and 3.5, depending on the recombinant strain (Figure 7B). *sc¹* siblings of these individuals carried one copy of the intact *miR-9a* gene, and they had an average SSO fate ranging between 0.3 and 3.4 SSOs, depending on the recombinant strain. For the large majority of recombinant strains, variability in SSO fate within a strain was greater when individuals

had one copy of *miR-9a* (Figure 7C). This result was consistent with the greater variability seen in the *miR-9a* mutant lab strain (Figure 1B).

Each wild recombinant strain had a unique genomic landscape potentially capable of affecting SSO fate. We predicted that variation in SSO fate between strains would be influenced by the diversity of genomic landscapes between strains. If miR-9a suppressed the impact of this genomic diversity on the fate variation between strains, then reduction of *miR-9a* gene copy number would increase this variation. This is precisely what was observed (Figure 7D). There was a 30% increase in the variation in SSO fate among the 32 recombinant strains when *miR-9a* copy number was decreased ($p = 0.0193$, Bennett log likelihood test). Thus, miR-9a suppresses variation in cell fate determination that is induced by the variation between different natural genomic landscapes.

DISCUSSION

An outstanding question is whether cell behaviors are actively made insensitive to the immense genomic diversity that exists between individuals (Gibson and Wagner, 2000; de Visser et al., 2003). Here, we present one mechanism by which this may be achieved, and we experimentally validate this mechanism in the context of a natural biological system. We show that the regulation of the transcription factor Sens by miR-9a renders cell fate less sensitive to the varied genomic landscapes of individuals. We explain this effect by the ability of miR-9a to create a threshold response (Figure 1H) (Cohen et al., 2006). Because miR-9a creates a threshold concentration of *sens* transcript that must be crossed before cell fate is switched, genomic

(D and E) Genome maps displayed by chromosome show the locations of variants that show correlated differences in allele frequency between replicates. Each variant is colored according to the magnitude of frequency difference for both replicates, as defined by its R value.

(D) The strain carrying two copies of the *miR-9a* gene is shown.

(E) The strain carrying one copy of the *miR-9a* gene is shown.

(F) Variants with correlated differences in allele frequencies were analyzed for diversity. Shown is their fold reduction in diversity comparing ancestral to selected genomes. Variants from strains with two copies of *miR-9a* are mirror plotted with those from strains with one copy of *miR-9a*. Genes that regulate sensory organ formation and contain such variants are listed by their names and locations. In red are genes shared by both strains that contain one or two copies of *miR-9a*. In black are genes unique to one strain or the other.

(G) Diversity reduction comparing ancestral to selected genomes for all variants mapped to the genome (200 kb bins) is shown.

See also Figure S6, and Tables S1 and S5.

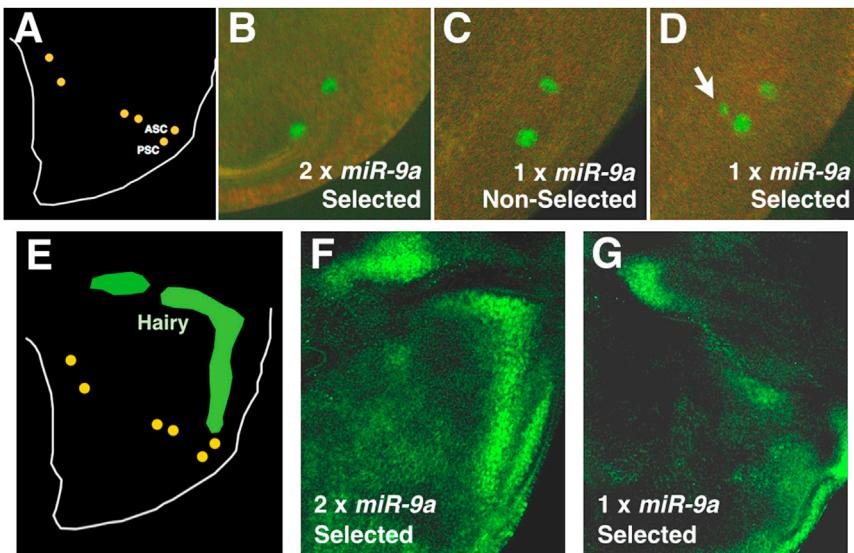


Figure 6. Genomic Landscape Affects Gene Expression in a Manner Dependent on miR-9a

(A) Schematic shows cells that give rise to sensory organs, including the anterior scutellar cells (ASCs) and posterior scutellar cells (PSCs) that form SSOs in an imaginal wing disc.

(B–D) Sens protein expression (green) in ASCs and PSCs is shown.

(B) An individual with a selected genomic landscape and two copies of *miR-9a* is shown. Of 17 discs examined, 0 had ectopic Sens-positive cells.

(C) An individual with a nonselected genomic landscape and one copy of *miR-9a* is shown. Of 24 discs examined, 0 had ectopic Sens-positive cells.

(D) An individual with a selected genomic landscape and one copy of *miR-9a* is shown. Arrow highlights an ectopic Sens-positive cell. Of 20 discs examined, 4 had an ectopic Sens-positive cell ($p < 0.0124$).

(E) Schematic shows Hairy protein expression in a wing imaginal disc. Two stripes of strong expression are set amidst more ubiquitous low-level expression.

(F and G) Hairy protein expression (green) is shown.

(F) An individual from one replicate group after selection with two copies of *miR-9a* is shown.

(G) An individual from one replicate group after selection with one copy of *miR-9a* is shown.

See also Table S1.

variants that perturb *sens* transcription are rendered ineffective if the threshold is not passed. If *miR-9a* is less effective at attenuating *sens*, then these same variants would be more likely to trigger the cell fate switch. Several variants were found to affect genes that regulate *sens* transcription, and these variants correlate with a greater impact on the cell fate switch. The variants that affect these genes are not likely sufficient to perturb the network because a constellation of other loci also showed signs of influence. Each of these might have weak and probabilistic effects on *sens* transcription, which in combination can potentially perturb the network. The impact of such probabilistic fluctuations would be suppressed by the *miR-9a*-engineered threshold.

We believe that this is the first demonstration of a molecular mechanism that buffers genomic diversity via gene expression. Transcription factors have previously been invoked as buffering agents of genome variation (Gibson and Dworkin, 2004; Siegal and Bergman, 2002), although how they achieve this is not known. Their impact on buffering can be significant, as witnessed in classic selection experiments with the *sc¹* mutation (Rendel, 1959, 1963). Impaired Ac/Sc activity generates increased h^2 , a result that we recapitulated here. Genome buffering can also occur by posttranslational means that involve protein chaperones (Jarosz and Lindquist, 2010).

Although miRNAs have features that theoretically make them suitable for buffering, clearly, this function is not generalized to all miRNAs in all cells. *miR-9a* shows this function, whereas *miR-7* does not. Yet, both miRNAs are highly conserved from fly to human, and the tissue specificity of their expression is also highly conserved (Christodoulou et al., 2010). We propose that the difference can be found in each miRNA's target within the fate switch network. *miR-7* indirectly activates ac/sc transcription, and in turn, its transcription is activated by Ac/Sc (Lai

et al., 2005; Li et al., 2009a) (Figure 1A). Thus, genome-induced fluctuation of Ac/Sc activity would be amplified and not damped by *miR-7*. It could be that *miR-7* functions in a nonhomeostatic way in this particular network.

It is rare for genome variants to have deterministic effects on outcome, and yet the personal genome is being heralded as a new instrument for predicting disease risk and prognosis. Genome variation that correlates with certain outcomes could become a powerful predictor for prevention and treatment. However, counteracting this relationship will be buffering mechanisms that resemble the one that we have described. The strength and efficacy of such mechanisms will affect the probability that certain variants are disease causative.

Buffering mechanisms can also affect the evolution of diseases, in particular cancer. Cancers evolve by clonal expansion, genome instability, and clonal selection (Greaves and Maley, 2012). We suggest that tumor heterogeneity is not only manifested by clones of cells with distinct genomes (Navin et al., 2011) but by the variable buffering of these genomes within a tumor. In support, it has been found that individual tumor cells from a common genetic lineage are phenotypically heterogeneous with respect to growth and responsiveness to therapy (Kreso et al., 2013). This phenotypic heterogeneity then enables selection for heritable features that promote cell survival and growth in the evolving environment of the tumor.

We found that subtle change in *miR-9a* copy number had large effects on buffering genomic diversity. Hence, the fluid variation in gene copy number, commonly seen in tumor cells (Hanahan and Weinberg, 2011), might have an impact in ways previously unforeseen. Moreover, epigenetic variation of gene expression in tumor cells could have similar consequences. In this regard, hypermethylation of the human *miR-9* promoter is frequently

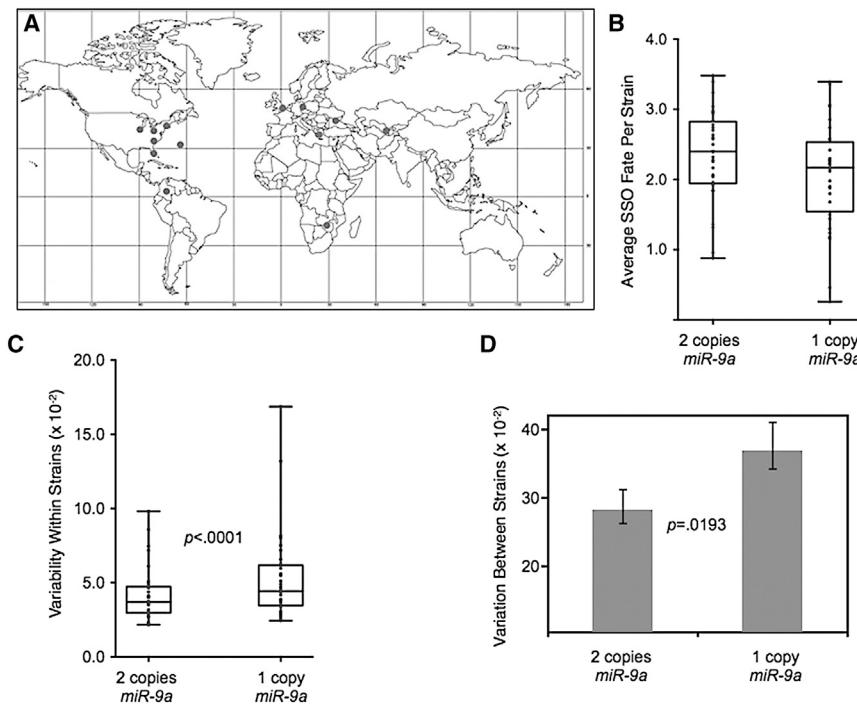


Figure 7. miR-9a Concentration Modulates the Impact of Worldwide Genomic Diversity on SSO Fate

(A) Map shows collection locations for the wild strains used in the experiment.

(B) Boxplot distributions of average SSO fate in 32 recombinant strains that had been made by introgressing wild strains with sc^1 and $miR-9a$ alleles are shown. For each strain, sibs with either one or two copies of the wild-type $miR-9a$ gene were assayed.

(C) Boxplot distributions of SSO fate variation among individuals within the 32 recombinant strains are shown. For each strain, sibs with either one or two copies of the wild-type $miR-9a$ gene were assayed. A Wilcoxon rank test was applied to compare the one- and two-copy groups.

(D) SSO fate variation among the 32 recombinant strains when carrying either one or two copies of the $miR-9a$ gene is shown. Error bars are 68th percentile confidence intervals. A Bennett log likelihood test was applied to compare variation coefficients for the one- and two-copy groups. See also Tables S1 and S6.

observed in various carcinomas; this leads to reduced expression of the miRNA (Bandres et al., 2009; Hildebrandt et al., 2010; Lehmann et al., 2008; Tsai et al., 2011). For renal carcinoma, the epigenetic modification is associated with increased risk for recurrence (Hildebrandt et al., 2010). Our results suggest that reduced miRNA gene expression might affect the ability of cancer cells to evolve under natural and therapeutic conditions.

In conclusion, we have identified a miRNA that inhibits the potential for genomic diversity to express itself at the level of a cell phenotype. Because selection experiments are applicable to many genes and cell types, our integrated approach should aid in understanding how genomic diversity is buffered in other organisms for traits that include disease risk and prognosis.

EXPERIMENTAL PROCEDURES

Genetics

The sc^1 allele is a gypsy transposon insertion within an enhancer of the *achaete-scute* gene complex on the X chromosome (Campuzano et al., 1986). The $miR-9a^{J22}$ and $miR-9a^{F39}$ null alleles are gene replacements of the pre-miR sequence with a mini-white⁺ gene (Li et al., 2006). The $miR-7^{J1}$ allele is a deletion of miR-7 resulting from imprecise excision of the white⁻ marked EP954 P element (Li and Carthew, 2005). The *sens^{E1}* allele is a protein null EMS mutation (Nolo et al., 2000). The genomic *sens* transgene P[acman] construct (18.1 kb gap-repaired fragment *sens-L*) has been described (Venken et al., 2006). It was inserted by phiC31-mediated recombination into the VK37 attP landing site on chromosome II (Venken et al., 2006). This transgene was shown to rescue the *sens^{E1}* null mutation. The miR-9a binding sites predicted by TargetScan and PicTar in the 607 nt *sens* 3' UTR sequence were mutated by deleting the sequences (5'-TAAGTCTGTACATAGTTATCCAAACA-3' and 5'-CAAAATTGGCAGCTAACAGTACCAAAGA-3') by serial recombineering. All versions of the *sens* gene were inserted into the same VK37 attP site to neutralize position effects. A summary of all genotypes used in experiments is provided in Table S1.

SSO Quantitation

For interaction and temperature experiments, SSO numbers were counted on 50 females randomly chosen from a population. This procedure was replicated three to four times with independent populations. Replicate means and coefficients of variation were averaged, and their SEs were corrected for biased underestimation due to the small number of replicates (Sokal and Rohlf, 1995). For selection experiments, SSO numbers were counted on 100 females of the appropriate genotype that were randomly chosen from the population at each generation. The mean and error estimates (variance and SE) for each replicate were kept independent of the other replicates for multivariate regression analysis (see *Supplemental Information*). For genomic landscape experiments, SSO numbers were counted on 150–300 females randomly chosen from a population. The mean, coefficient of variation, and SE were calculated. This was repeated for four different selected/nonselected pairs of landscapes. For the experiments with wild recombinant strains, 50 females were scored for SSOs, and means and uncertainty were calculated for each group.

Selection

Establishment of strains for selection is described in the *Extended Experimental Procedures*. Replicate populations were grown in uncrowded and well-fed conditions at a constant temperature of $23^\circ\text{C} \pm 1^\circ\text{C}$. For each generation, individuals of the appropriate *sens* or $miR-9a$ genotype were identified before SSO scoring. Details of how genotyping was performed are described in the *Extended Experimental Procedures*. To score offspring with the appropriate genotype, a subset of all such individuals in the population was randomly chosen, and 100 females and 100 males had their SSO number counted. After scoring, 50 females and 50 males were selected from the scored groups, and these individuals were used to parent the next generation. Selection was dictated by ranking of the individuals within the group by SSO number. For one group, the 50% of scored individuals with highest SSO number were selected for breeding. For another group, individuals were randomly selected for breeding. This process of mating-scoring-selecting was repeated for a total of 5 generations for the *sens* experiment, and a total of 15 generations for the *miR-7* and *miR-9a* experiments.

Genomic Landscape Tests

After 15 generations of selection (random or directional), exchange of the sc^1 w^{118} X chromosome with a sc^+ X chromosome was performed by crossing

selected males to sc^+ $y\ w$ females that had balanced autosomes. Balanced offspring were self-crossed, and F2 sibs were mated to generate F3 individuals with $sc^+ y\ w$ X chromosomes and autosomes that were derived from selected ancestors. The F3 individuals were self-crossed to make a pure stock. To analyze the effect of *sens* on different genomic landscapes, $sc^+ miR-9a^+$ flies whose ancestors had undergone selection were crossed with flies bearing a *sens* transgene on chromosome II and a *sens*^{E1} mutant allele on the third chromosome. We scored offspring bearing one copy of the transgene and one copy of *sens*^{E1}.

Wild strains isolated from around the world were obtained from Bloomington Stock Center and the *Drosophila* Genetic Resource Panel (DGRP). See Table S6 for a complete description of the strains. $sc^+ w^{1118}; +/CyO; miR-9a/TM6B$ females were crossed to each wild strain, and backcrossing to the wild strain was repeated for five generations; each generation selecting for the $sc^+ w^{1118}; miR-9a/+$ genotype. In the generation in which the final backcrossing was completed, individuals possessed a genome, 95% of which was derived from the wild strain. These individuals were interbred to form a $sc^+ w^{1118}; miR-9a/+$ recombinant stock, which was propagated for 20 generations to blend genomes. Sibs were mated, and their *miR-9a/+* and *+/+* female offspring were scored for SSOs. To estimate the variation among the 32 strains, the variation coefficient was calculated from the pool of 32 SSO averages.

Temperature Challenge

Nonselected sc^+ lines with fixed *miR-7*, *miR-9a*, or *sens* genotypes were subjected to temperature perturbation. Males and females were mated and allowed to egg lay in bottles at 23°C for 3 hr before transfer to fresh bottles. Bottles with offspring were immediately transferred to incubators set at 18°C, 23°C, or 29°C. All offspring were allowed to eclose into adults before a sample of 50 females were scored for SSOs. The experiment was repeated at least three times. Different *miR-9a* mutant alleles (*J22* and *E39*) were tested with similar results.

Immunohistochemistry

Imaginal wing discs from female white prepupae were fixed in 4% formaldehyde in PBS. Discs were washed in PBS plus 0.3% Triton X-100 (PBST) before incubating in 1:1,000 anti-Sens or anti-Hairy for 3 hr at 23°C. Discs were washed with PBST and incubated for 1 hr with 1:250 Alexa Fluor 488-conjugated IgG. Samples were visualized on a Zeiss LSM 510 Meta confocal microscope. Images were taken at 1.0–2.0 μm optical slice widths, and all imaging parameters were kept constant between sample runs.

Genomic DNA Sequencing

A total of 25–30 frozen virgin female flies were pooled, and their genomic DNA was prepared by mechanical lysis in PBS, and 0.3 mg Proteinase K was added to tubes for 6 hr at 55°C. The DNeasy Blood & Tissue Kit (QIAGEN) was used to purify genomic DNA. DNA was fragmented, and an Illumina TruSeq kit was used to prepare libraries for sequencing. Library composition is outlined in Figure S3A. Multiplexed pooled sequencing was performed according to standard Illumina protocols. Sequence analysis is described in detail in the Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2013.10.057>.

ACKNOWLEDGMENTS

We thank the following for reagents: F.B. Gao, J. Reinitz, the Bloomington Stock Center, the Developmental Hybridoma Studies Bank, and NCI Frederick. Thanks to C. LaBonne for comments on the manuscript. We also thank Bionimbus (<http://www.bionimbus.org>). We acknowledge support from the Chicago Biomedical Consortium (to J.J.C.), the Malkin and Rappaport Foundations (to J.J.C.), the Pew Latin American Fellows Program (to D.M.P.), the McNair Medical Institute (to K.J.T.V.), the Howard Hughes Medical Institute

(to K.J.T.V. and H.J.B.), and the NIH for grants T32GM008061 (to J.J.C.), T32 HL094282 (to A.R.J.), P50 GM81892 (to K.P.W.), T32 GM007526 (to K.J.T.V.), and GM068743 and GM077581 (to R.W.C.).

Received: May 7, 2013

Revised: September 16, 2013

Accepted: October 29, 2013

Published: December 19, 2013

REFERENCES

- 1000 Genomes Project Consortium, Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073.
- Acar, M., Jafar-Nejad, H., Giagtzoglou, N., Yallampalli, S., David, G., He, Y., Delidakis, C., and Bellen, H.J. (2006). Senseless physically interacts with proneural proteins and functions as a transcriptional co-activator. *Development* 133, 1979–1989.
- Bandres, E., Aguirre, X., Bitarte, N., Ramirez, N., Zarate, R., Roman-Gomez, J., Prosper, F., and Garcia-Foncillas, J. (2009). Epigenetic regulation of microRNA expression in colorectal cancer. *Int. J. Cancer* 125, 2737–2743.
- Bang, A.G., Bailey, A.M., and Posakony, J.W. (1995). Hairless promotes stable commitment to the sensory organ precursor cell fate by negatively regulating the activity of the Notch signaling pathway. *Dev. Biol.* 172, 479–494.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.
- Bhattacharya, A., and Baker, N.E. (2011). A network of broadly expressed HLH genes regulates tissue-specific cell fates. *Cell* 147, 881–892.
- Campuzano, S., Balcells, L., Villares, R., Carramolino, L., García-Alonso, L., and Modolell, J. (1986). Excess function hairy-wing mutations caused by gypsy and copia insertions within structural genes of the achaete-scute locus of *Drosophila*. *Cell* 44, 303–312.
- Carthew, R.W., and Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136, 642–655.
- Christodoulou, F., Raible, F., Tomer, R., Simakov, O., Trachana, K., Klaus, S., Snyman, H., Hannon, G.J., Bork, P., and Arendt, D. (2010). Ancient animal microRNAs and the evolution of tissue identity. *Nature* 463, 1084–1088.
- Cohen, S.M., Brennecke, J., and Stark, A. (2006). Denoising feedback loops by thresholding—a new role for microRNAs. *Genes Dev.* 20, 2769–2772.
- Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., et al.; Wellcome Trust Case Control Consortium (2010). Origins and functional impact of copy number variation in the human genome. *Nature* 464, 704–712.
- Cubadda, Y., Heitzler, P., Ray, R.P., Bourouis, M., Ramain, P., Gelbart, W., Simpson, P., and Haenlin, M. (1997). u-shaped encodes a zinc finger protein that regulates the proneural genes achaete and scute during the formation of bristles in *Drosophila*. *Genes Dev.* 11, 3083–3095.
- de Visser, J.A., Hermisson, J., Wagner, G.P., Ancel Meyers, L., Bagheri-Chaichian, H., Blanchard, J.L., Chao, L., Cheverud, J.M., Elena, S.F., Fontana, W., et al. (2003). Perspective: evolution and detection of genetic robustness. *Evolution* 57, 1959–1972.
- Ebert, M.S., and Sharp, P.A. (2012). Roles for microRNAs in conferring robustness to biological processes. *Cell* 149, 515–524.
- Escudero, L.M., Caminero, E., Schulze, K.L., Bellen, H.J., and Modolell, J. (2005). Charlatan, a Zn-finger transcription factor, establishes a novel level of regulation of the proneural achaete/scute genes of *Drosophila*. *Development* 132, 1211–1222.
- Falconer, D.S., and Mackay, T.F.C. (1996). Introduction to Quantitative Genetics (Harlow: Longman).
- Félix, M.A. (2012). Evolution in developmental phenotype space. *Curr. Opin. Genet. Dev.* 22, 593–599.

- Frazer, K.A., Murray, S.S., Schork, N.J., and Topol, E.J. (2009). Human genetic variation and its contribution to complex traits. *Nat. Rev. Genet.* 10, 241–251.
- Gibson, G., and Wagner, G. (2000). Canalization in evolutionary genetics: a stabilizing theory? *Bioessays* 22, 372–380.
- Gibson, G., and Dworkin, I. (2004). Uncovering cryptic genetic variation. *Nat. Rev. Genet.* 5, 681–690.
- Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. *Nature* 481, 306–313.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Hildebrandt, M.A., Gu, J., Lin, J., Ye, Y., Tan, W., Tamboli, P., Wood, C.G., and Wu, X. (2010). Hsa-miR-9 methylation status is associated with cancer development and metastatic recurrence in patients with clear cell renal cell carcinoma. *Oncogene* 29, 5724–5728.
- Hornstein, E., and Shomron, N. (2006). Canalization of development by microRNAs. *Nat. Genet. Suppl.* 38, S20–S24.
- Jafar-Nejad, H., Acar, M., Nolo, R., Lacin, H., Pan, H., Parkhurst, S.M., and Bellen, H.J. (2003). Senseless acts as a binary switch during sensory organ precursor selection. *Genes Dev.* 17, 2966–2978.
- Jarosz, D.F., and Lindquist, S. (2010). Hsp90 and environmental stress transform the adaptive value of natural genetic variation. *Science* 330, 1820–1824.
- Kreso, A., O'Brien, C.A., van Galen, P., Gan, O.I., Notta, F., Brown, A.M., Ng, K., Ma, J., Wienholds, E., Dunant, C., et al. (2013). Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. *Science* 339, 543–548.
- Lai, E.C., Tam, B., and Rubin, G.M. (2005). Pervasive regulation of *Drosophila* Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. *Genes Dev.* 19, 1067–1080.
- Lehmann, U., Hasemeier, B., Christgen, M., Müller, M., Römermann, D., Länger, F., and Kreipe, H. (2008). Epigenetic inactivation of microRNA gene hsa-mir-9-1 in human breast cancer. *J. Pathol.* 214, 17–24.
- Li, X., and Carthew, R.W. (2005). A microRNA mediates EGF receptor signaling and promotes photoreceptor differentiation in the *Drosophila* eye. *Cell* 123, 1267–1277.
- Li, Y., Wang, F., Lee, J.A., and Gao, F.B. (2006). MicroRNA-9a ensures the precise specification of sensory organ precursors in *Drosophila*. *Genes Dev.* 20, 2793–2805.
- Li, X., Cassidy, J.J., Reinke, C.A., Fischboeck, S., and Carthew, R.W. (2009a). A microRNA imparts robustness against environmental fluctuation during development. *Cell* 137, 273–282.
- Mackay, T.F., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., et al. (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482, 173–178.
- Masel, J., and Siegal, M.L. (2009). Robustness: mechanisms and consequences. *Trends Genet.* 25, 395–403.
- Mendell, J.T., and Olson, E.N. (2012). MicroRNAs in stress signaling and human disease. *Cell* 148, 1172–1187.
- Mukherji, S., Ebert, M.S., Zheng, G.X., Tsang, J.S., Sharp, P.A., and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. *Nat. Genet.* 43, 854–859.
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., et al. (2011). Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–94.
- Nolo, R., Abbott, L.A., and Bellen, H.J. (2000). Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in *Drosophila*. *Cell* 102, 349–362.
- Ohnsako, S., Hyer, J., Panganiban, G., Oliver, I., and Caudy, M. (1994). Hairy function as a DNA-binding helix-loop-helix repressor of *Drosophila* sensory organ formation. *Genes Dev.* 8, 2743–2755.
- Quan, X.J., and Hassan, B.A. (2005). From skin to nerve: flies, vertebrates and the first helix. *Cell. Mol. Life Sci.* 62, 2036–2049.
- Rendel, J.M. (1959). Canalization of the scute phenotype of *Drosophila*. *Evolution* 13, 425–439.
- Rendel, J.M. (1963). Correlation between the number of scutellar and abdominal bristles in *Drosophila melanogaster*. *Genetics* 48, 391–408.
- Sheldon, B.L. (1968). Studies on the scutellar bristles of *Drosophila melanogaster*. I. Basic variability, some temperature and culture effects, and responses to short-term selection in the Oregon-RC strain. *Aust. J. Biol. Sci.* 21, 721–740.
- Siegal, M.L., and Bergman, A. (2002). Waddington's canalization revisited: developmental stability and evolution. *Proc. Natl. Acad. Sci. USA* 99, 10528–10532.
- Smith, J.M., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet. Res.* 23, 23–35.
- Sokal, R.R., and Rohlf, F.J. (1995). *Biometry: The Principles and Practice of Statistics in Biological Research*, Third Edition (New York: W.H. Freeman and Company).
- Tsai, K.W., Liao, Y.L., Wu, C.W., Hu, L.Y., Li, S.C., Chan, W.C., Ho, M.R., Lai, C.H., Kao, H.W., Fang, W.L., et al. (2011). Aberrant hypermethylation of miR-9 genes in gastric cancer. *Epigenetics* 6, 1189–1197.
- Tsang, J., Zhu, J., and van Oudenaarden, A. (2007). MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol. Cell* 26, 753–767.
- Van Doren, M., Ellis, H.M., and Posakony, J.W. (1991). The *Drosophila* extra-macrochaetae protein antagonizes sequence-specific DNA binding by daughterless/achaete-scute protein complexes. *Development* 113, 245–255.
- Venken, K.J., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* 314, 1747–1751.
- Weir, B. (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data* (Sunderland: Sinauer Associates).

Supplemental Information

EXTENDED EXPERIMENTAL PROCEDURES

Construction of Strains for Selection

Inbred Stocks Used for Selection

sc¹ was recombined with *white* (*w¹¹¹⁸*), and the chromosome was isogenized before introgression with specific autosomes, as shown in Figure S1. Autosomes from three genetic backgrounds were blended before balancing. The blended stock with isogenized X chromosome is referred to as the basal *sc w* introgression stock.

Construction of sens Selection Strains

A single strain with blended chromosomes was created for each *sens* transgene genotype - *sens*, *sens-m1*, *sens-m2*, *sens-m1m2*. To construct these strains, the basal *sc w* introgression stock was crossed to *sens* transgenic stocks, and the F1 offspring were back-crossed to the basal *sc w* stock to create a *sc w* strain with the *white⁺* marked transgenic chromosome II over blended inbred second chromosomes from the basal stock (Figure S2). Chromosome III had its endogenous *sens* gene intact.

Construction of miR-9a Selection Strains

One set of strains labeled A was established which had slightly lower genomic diversity than another set of strains, labeled B (Figure S3). Strain A was constructed such that the X chromosome originated from a single isogenized *sc¹ w¹¹¹⁸* chromosome. Variation on chromosomes III and IV was limited by blending together three independent inbred stocks, as described above. The two-copy *miR-9a A* strain was crossed with a balanced *miR-9a* mutant stock (either mutant allele) to generate a one-copy A strain with equivalent variation on chromosome II, and the *miR-9a* mutant chromosome over the blended wild-type third chromosome. Although this could potentially generate different variation on chromosome III when comparing the one- and two-copy strains, we observed little difference in the number of sequence variants found on chromosome III between A strains. Strain B was constructed such that the X chromosome was blended from two independent isogenized *sc¹ w¹¹¹⁸* chromosomes. Chromosomes III and IV received genetic variation from the three inbred stocks plus a fourth stock - Canton S. These additional introgressions were also performed to construct the one-copy *miR-9a B* strain.

Construction of miR-7 Selection Strains

Mutant *miR-7* and its matched wild-type second chromosome were derived by P-element excision from a common parental EP954 enhancer trap line (Li and Carthew, 2005). Hence, genetic variation on chromosome II was matched between the one- and two-copy strains that we constructed. An A strain with two copies of *miR-7* was prepared by blending autosomes from the same three inbred stocks used to make the *miR-9a A* strains (Figure S4). This strain also had the same isogenized X chromosome as was used to make the *miR-9a A* strain. The B strain with two copies of *miR-7* was prepared by introgressing the same two isogenized X-chromosomes as was used to make the *miR-9a B* strains (Figure S4).

Strain Replication

Strains were amplified once they had been established. For *miR-7* and *mir-9a* selection experiments, A and B strains were each divided into two replicate populations, thus yielding a total of four replicates for the one-copy strains, and four replicates for the two-copy strains. Replicates 1 and 2 derived from the A strains, and replicates 3 and 4 derived from the B strains. For *sens* selection experiments, each strain was divided into four replicate populations. All replicate populations were thereafter kept independent of one another. For the *miR-9a* selection experiment, each replicate population was further subdivided into two sub-populations that were kept independently. One sub-population was subjected to selection for individuals with the most SSOs; the other sub-population was subjected to random selection, which served as a control for drift and background selection.

Selection

During selection, replicate populations were grown in uncrowded and well-fed conditions at a constant temperature of 23°C ± 1°C in bottles with cornmeal-molasses food. Parents were mated and allowed to egg-lay under constant densities of 50 females and 50 males in one bottle for 0.5 - 1 day before they were transferred to a fresh bottle. Egg laying and transfer were serially repeated. Offspring were collected as adults twice-daily to capture virgin females. To minimize bias due to developmental timing, offspring were genotyped and scored for SSO number only after all offspring had eclosed. For populations in which transgenic or heterozygotic parents were mated, the offspring had a 1:2:1 ratio of *sens* or *miR* genotypes. We only scored adult offspring of the appropriate genotype, and this was possible due to the linkage of an eye-color transgene at the relevant loci. For *sens* transgenics and *miR-9a* heterozygotes, these were identified from their homozygous siblings by their intermediate red eye color due to having one copy of a linked *white⁺* gene. For *miR-7* heterozygotes, they were identified from their homozygous siblings as *miR-7¹¹* / EP954. The *miR-7¹¹* mutant allele is a deletion of *miR-7* resulting from imprecise excision of the *white⁺* marked EP954 P element (Li and Carthew, 2005). This EP954 does not affect *miR-7* gene activity and thus is a wild-type allele. Heterozygotes could be identified by the intermediate eye color due to the single *white⁺* copy. Matched *miR-7* two-copy control lines had the EP954 chromosome over a precise excision allele of EP954 called *ΔEP954-101* (Li and Carthew, 2005). This particular genotype was maintained each generation by selecting animals with intermediate eye color due to the single *white⁺* copy.

At each generation, 100 males and 100 females of the appropriate genotype were scored and ranked according to SSO number. The 50 males and 50 females with the highest number of SSOs were selected to propagate the next generation. Males and females were independently ranked for SSOs during selection, since gender influences the *sc¹* phenotype. Selection was repeated 5–15 generations. For the *miR-9a* strains, an additional copy of the replicate populations was subjected to 15 generations of

selection in which 50 males and 50 females were randomly selected from the 100 males and 100 females that were scored each generation.

The *sc¹* mutation is caused by a *gypsy* transposon insertion (Campuzano et al., 1985, 1986). To test whether the transposon had mobilized within populations during selection, we amplified genomic DNA from animals before and after selection by PCR. Primers were used that spanned the insertion site (would detect *gypsy*-free DNA); were positioned to left-flanking *gypsy* and genome DNA (would detect the *gypsy* insertion); were positioned to right-flanking *gypsy* and genome DNA (would detect the *gypsy* insertion). PCR was performed on genomic DNA and amplicons were analyzed. Starting populations showed evidence for the *gypsy* insertion as predicted. All of the lines after selection also showed PCR evidence for the *gypsy* insertion. No evidence of *gypsy*-free DNA was found. Therefore, transposon mobilization was not the cause of phenotypic change during selection.

Selected Genomic Landscape Experiments

After 15 generations of selection, *miR-9a* heterozygous flies were reconstituted homozygous wild-type for *sc* and *miR-9a* as described in the Experimental Procedures, retaining the directionally selected or randomly selected genetic backgrounds from which they descended. These flies were crossed with lines bearing a *sens* transgene on chromosome II and a *sens^{E1}* mutant allele on chromosome III (Nolo et al., 2001). We scored offspring bearing one copy of the transgene and one copy of *sens^{E1}*; this was possible due to the linkage of an eye-color marker (*white⁺*) with the transgene. A total of 200 females and 200 males were scored for each genotype.

In the selected background, the frequency of detecting ectopic Sens expression was lower than the frequency of detecting ectopic SSOs (compare Figures 5 and 6). This could be due to a mechanism that drives ectopic SSOs independent of Sens, or it could be due to the inherently asynchronous onset of Sens expression normally seen with SSO precursor cells. Specification of the SSO precursor cells is not very deterministic with time (Rodríguez et al., 1990). The anterior and posterior SSO SOPs are determined about 10 hr apart, and are poorly synchronized with the developing animal's molting stages. We found the same lack of synchronization, making developmental analysis of SSO SOPs difficult. Therefore, we do not know if the fewer ectopic Sens-positive cells is due to a variable delay in their appearance at the time we collected animals (the time of pupal molting). We could not harvest older animals because the first 10 hr of pupation generate radical morphological changes coupled with massive tissue breakdown.

Genomic Sequence Analysis

Two genetically distinct strains were sequenced: the two-copy *miR-9a* A strain and the one-copy *miR-9a* A strain (Figure S6A). For each strain, 5–6 populations were subjected to library preparation and sequencing, as outlined in Figure S6A. One lane of 50 bp single-end (for quality control), and multiple lanes of 100 bp paired-end libraries were sequenced using an Illumina HiSeq2000. A total of 879 million paired-end reads were generated (Table S5). Data analysis was performed on the Bionimbus cloud (<http://www.bionimbus.org>). Quality control runs were conducted to ensure library preparations were of high fidelity using the FastX toolkit (http://hannonlab.cshl.edu/fastx_toolkit). Reads were aligned to the Berkeley *Drosophila* Genome Project's fifth revision of the *Drosophila melanogaster* reference genome (dm3) using the Burrows-Wheeler Alignment algorithm pipeline (Li and Durbin, 2009). Picard Tools was used to remove PCR duplicates (<http://picard.sourceforge.net>). BEDTools was used to determine sequencing depth (Quinlan and Hall, 2010). Of the reads that had pairs, 96%–97% had their paired read map in close genomic proximity (Table S5). Three tiers of quality control throughout Burrows-Wheeler Alignment ensured high-fidelity alignments. First, the mapping quality score needed to be ≥ 40 . Quality control was also ensured at the level of each individual base pair, where only those with Phred scores ≥ 20 were further considered. The final quality control involved read lengths, where only reads > 14 base pairs were considered further. A minority of reads (20% to 37%) failed to meet these quality control standards. A summation of the analysis is in Table S5. All libraries were sequenced to depth of coverage such that 95% or more of the annotated genome was covered to a minimum depth of 15X (Table S5).

Identification of Sequence Variants

For each library, SAMtools Pileup was used to give every site along the genome a string of variables reflecting the total number of reads and the nucleotides called at that site (Li et al., 2009b). Thus, we generated a list of sites with more than one called nucleotide (variant sites) in each library. We pooled the lists of variant sites for all two-copy *miR-9a* libraries to generate a reference panel of variant sites (VarRef). Summing all library reads at each VarRef site generated deep coverage, with a median coverage of ~265 reads per VarRef variant site (Table S5). A separate VarRef reference panel was generated from the sequenced one-copy *miR-9a* libraries (~213 reads per VarRef variant site). Any VarRef site with total coverage depth < 10 reads was removed from the lists. A total of 6.2 million and 10 million VarRef variant sites remained in the two- and one-copy *miR-9a* panels, respectively. Most of these sites had only one read of the variant nucleotide, and many sites had two different variant nucleotides called. These variants were likely due to sequencing errors. A variant nucleotide call was discarded if its read count was less than 3. Sites that had two or more different variant nucleotides called (tri-allelic) were also discarded. These filters reduced the number of VarRef variant sites to 1.6 million and 1.5 million for the two- and one-copy *miR-9a* panels, respectively.

Exceptionally high read coverage of some VarRef sites was observed, and was likely due to repetitive sequence structure. We considered the likelihood of errors to be greater in such regions. Therefore, we used the 95th percentile of read depth as an

uppermost boundary for site inclusion. In the two-copy panel, this meant we excluded any VarRef site with > 340 reads covering it. The one-copy panel's 95th percentile was such that any VarRef site with > 277 reads covering it was excluded. This boundary was lower due to slightly lower sequencing coverage of the one-copy libraries. We also applied a lowermost boundary for site inclusion. We excluded a VarRef variant site if read depth in any single library of the panel was less than 10X. When combined, these boundaries reduced the number of VarRef variant sites to 1.13 million and 1.03 million for the two- and one-copy *miR-9a* panels, respectively. Thus, all examined variant sites had a reference nucleotide and only one nonreference nucleotide.

We subjected these variant sites to a distinct filtering that considered each site's variation within an individual library, since each library represented the pooled genomes of a population. For each variant site in a given library, we defined the nucleotide with the greater read number to be the major allele and the nucleotide with the smaller read number to be the minor allele. The minor allele frequency was calculated as the number of minor allele reads divided by the total number of reads. The normal approximation of a 99.73% binomial confidence for each minor allele was calculated as:

$$99.73\% \text{ CI} = \widehat{p}_{i,j} \pm z_{1-\alpha/2} \sqrt{\frac{\widehat{p}_{i,j}(1 - \widehat{p}_{i,j})}{n_{i,j}}}$$

where $\widehat{p}_{i,j}$ is the minor allele frequency of the i^{th} site in the j^{th} library, $n_{i,j}$ is the coverage at that site in that library, and $z_{1-\alpha/2}$ is the $1 - \alpha/2$ percentile of a standard normal distribution. For a site to be called variant, at least one of the libraries in the panel was required to have $99.73\% \text{ CI} > 0$. This accounted for coverage differences between libraries at each site.

The rationale for choosing these filters was biologically motivated, and indeed, we used a biological feature to compare, develop, and optimize the filters. The X chromosome was isogenized in all experimental populations that were sequenced, meaning that few variants existed on the X-chromosomes since they all descended from a single X chromosome 10-25 generations prior. We therefore compared variant calls on the X chromosome to calls on the autosomes as a means to develop and test filtering parameters (Table S5). We compared three different filtering parameters: 1) A minimum number of times a minor allele read is called at a site, 2) A minimum percentage of times a minor allele read is called at a site, 3) The confidence interval of a minor allele at a site. Each filtering parameter was varied, and called variant sites were calculated (Table S5). We chose to use the confidence interval method with the parameter set to 99.73% because it made < 2% of variant calls on the X chromosome, expressed as a percentage of all variants called over the genome (Table S5). This represents the maximum false discovery rate (FDR) due to sequencing errors since some of the X chromosome variants could have originated by de novo mutagenesis after the isogenization event. The other methods used for filtering achieved similar FDRs when their parameter values were set to certain levels (Table S5).

We ultimately chose to use the confidence interval method because it allowed the most balanced method to normalize between the one- and two-copy *miR-9a* panels. The depth of sequence coverage is subtly but significantly different between the two panels, and the confidence interval method compensates for any biased variant calling caused by coverage depth. In total, we identified 467,220 and 551,228 variants in libraries from two copy and one copy populations, respectively. We compared these results to the results from the calling method that used a minimum number of variant reads of > 22 (Table S5). More than 88% of variants called by the 99.73% Confidence Interval method were also called by the other method. To further validate our calling method, we used an established algorithm to identify variants from whole genome sequencing to independently analyze our sequence data. GATK Unified Genotyper uses a Bayesian method to calculate the probability of a locus being polymorphic based upon known/estimated genomic heterozygosities (DePristo et al., 2011; McKenna et al., 2010). We compared the variants called by Unified Genotyper with those called by the 99.73% Confidence Interval method (Table S5). The large majority (71 – 76%) of variants called by the 99.73% Confidence Interval method were also called by Unified Genotyper. These results indicate that highly divergent approaches to calling variants were identifying the same sites, and therefore these sites were truly polymorphic.

It was important to confirm that starting genomic variation was similar between one- and two-copy *miR-9a* strains when drawing comparisons between the strains to explain differences in their selection responses. Reassuringly, the number of variants called between one- and two-copy strains was similar (Table S5). Furthermore, one can plot nonreference allele frequency with variant number to analyze the starting genomic diversity in each population. These distributions appear very similar between one- and two-copy strains (Figure S6B).

Allele Frequency Analysis

Each called variant had its reference allele frequency (F) calculated as the ratio of reference read number over the total read number in a library. The F value of a variant in the selected replicate library was subtracted by the F value of the variant in its matched randomly-selected replicate library. This difference was defined as the allele frequency difference ΔF . Matched replicates had their ΔF values plotted and compared in Figures 4A–4C. A filter was applied to extract only variants that showed correlative ΔF values between replicates. The filter had three features: 1) both replicate ΔF values had the same sign, 2) neither ΔF values were zero, 3) the sum-of-squares $R = \sqrt{(\Delta F_{[\text{replicate 1}]}^2 + \Delta F_{[\text{replicate 2}]}^2)}$ was greater than a defined number. This number was varied to make the filter more or less stringent. We found that when $R > 0.6$, the false discovery rate (FDR) was minimal for all libraries. The FDR was calculated

by comparing the number of extracted variants using the filter described above to the number of variants extracted using a filter where replicate ΔF values had the opposite sign.

Gene diversity (D) or expected heterozygosity was calculated as described (Weir, 1996), using F as the reference allele frequency and 1 - F as the nonreference allele frequency. This was calculated for each variant in each sequenced library. The diversity of all variants within 200 kb of one another was binned and averaged, and binned averages were plotted on the genome (Figures S6C and S6D). Diversity reduction for each variant was estimated by taking the ratio of the calculated D from the selected library over the calculated D from the matched ancestral library. The inverse of this ratio was taken as fold-reduction.

Database of SSO Genes

We conducted a literature search for published manuscripts that described genome-wide screens to identify genes involved in sensory organ development. We focused on functional approaches that were able to narrow in on, or near, individual genes. From each work, lists of potential genes were acquired and converted from CG or gene name into FlyBase gene ID numbers using the batch download feature found on [FlyBase.org](#). We were specifically interested in extracting the genomic coordinates of each candidate gene to cross-reference with our lists of variants likely to be under selection (Figure S6E). For this purpose, we converted the database to BED format (file available upon request). We compared lists with each other using ANNOVAR (Wang et al., 2010) and the UCSC genome browser ([genome.ucsc.edu](#)) (Kuhn et al., 2009).

De Novo Mutations

Since the X chromosome had been isogenized to create zero variants before selection, we assumed that the X chromosome variants observed after selection had originated by de novo mutagenesis or were sequencing errors. Importantly, the number of X-linked variants was not greater in the one-copy *miR*-9a replicates than in the two-copy replicates, suggesting that reduced *miR*-9a did not significantly increase the number of variants made by de novo mutagenesis. The ~5,000 variants detected on the X chromosome is close to the expected number of variants based on the rate of natural mutagenesis in *Drosophila* (Haag-Liautard et al., 2007). Extrapolating to the total genome, we estimate that at most, 25,000/730,479 or 3% of all variants originated de novo, and the remainder were pre-existing.

Quantitative Analysis and Statistics

To quantify the heritable response to selection, mean SSO number was calculated by scoring 100 females of the appropriate genotype in a replicate population at each generation of selection. These means were plotted against the selection differential cumulating each generation. The realized heritability was estimated from linear regression of mean SSO number on the cumulated selection differential (Falconer and Mackay, 1996). The slope of the best-fit regression line is an averaged estimate of the realized heritability (Falconer and Mackay, 1996). An SAS procedure was used to perform mixed multivariate linear regression analysis using a model with replicates, genotypes, and gender as discrete variables. For *miR*-7 strains, multivariate analysis supported the null hypothesis that the coefficients/intercepts were the same for all replicates of the same genotype. For *miR*-9a strains, analysis supported the hypothesis that the coefficients were the same for replicates 1 and 2, and they were different from those coefficients for replicates 3 and 4. Therefore when multivariate analysis indicated it was appropriate, observations were pooled from replicates that showed no difference in their regression coefficients. The reason for the different coefficients in *miR*-9a replicates is unclear but the difference correlates with the replicate genetic backgrounds. Replicates 1 and 2 were derived from the A strain, which had a more limited genetic background than the B strain, which gave rise to replicates 3 and 4. We suspect that different combinations of variants were present in one background and not the other. Heteroscedastic error of the SSO mean estimates required the application of a weighted least-squares regression. The slope of the best-fit line was computed using a weighted least-squares method, the weight being the inverse of the estimated variance of each datapoint (Frodesen et al., 1979; Strutz, 2010). Standard error of the slope was also calculated. Z-tests were applied to compare one slope to another. The *miR*-7 and *sens* strains showed significantly different slopes from each other and the *miR*-9a strains. We surmise this effect is due to the different genetic backgrounds between the strains, which has an influence on the SSO phenotypes.

For the temperature perturbation experiments, analysis of variance (ANOVA) was used to partition different sources of the observed SSO variation according to the model:

$$Y = \mu + G + T + B + G \times T + T \times B + G \times B + L(G \times B) + G \times T \times B + \varepsilon$$

where Y is the SSO number, μ is the overall mean, G refers to genotype, T is the temperature, B is the replicate, L (random factor) is the population line nested within the genotype by a replicate interaction term, and ε is the error term. Analysis was done for *sens*, *mir*-7, and *miR*-9a separately using the SAS procedure GLM.

For the experiment involving wild recombinant strains, the coefficient of variation (CV) between the strains was calculated by estimating the mean and variance of SSO means for the 32 strains. To generate a confidence interval for each CV estimate, a noncentral t-distribution was used (Kelley, 2007a). This assumption requires that the set of SSO means falls within a normal distribution. A Shapiro Wilk normality test confirmed that this was the case for both one- and two-copy sets. The confidence intervals were calculated using the MBESS package in R (Kelley, 2007b). The log likelihood ratio test developed by

Bennett (Bennett, 1976) was applied computationally as described (Reh and Scheffler, 1996) to derive the *p* statistic comparing interstrain CVs.

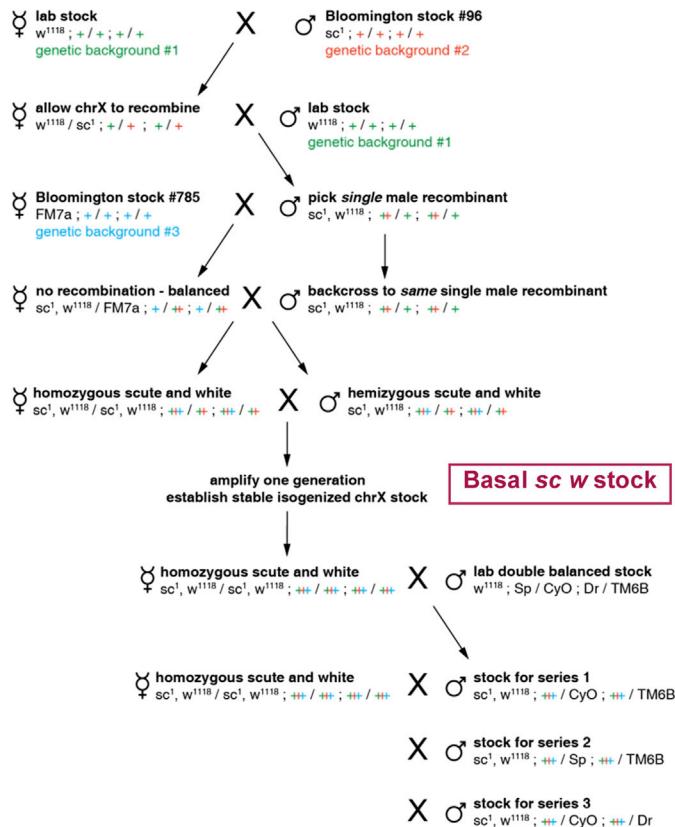
Other Informatics Tools Used

Variants were annotated using the Perl script ANNOVAR (Wang et al., 2010) on our list of all called variants and on our list of likely selected variants (Table S5). ANNOVAR fly genomic curation data (refGene.txt.gz and refLink.txt.gz) was downloaded on August 29th, 2010 from: <ftp://hgdownload.cse.ucsc.edu/goldenPath/dm3>.

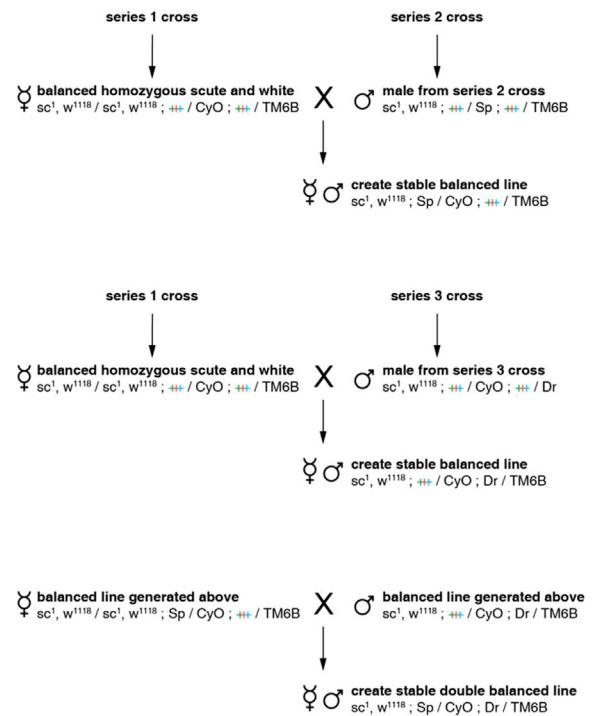
SUPPLEMENTAL REFERENCES

- Bennett, B.M. (1976). On an approximate test for homogeneity of coefficient of variation. In Contribution to Applied Statistics, W.J. Ziegler, ed. (Basel: Birkhäuser Verlag), pp. 169–171.
- Campuzano, S., Carramolino, L., Cabrera, C.V., Ruiz-Gómez, M., Villares, R., Boronat, A., and Modolell, J. (1985). Molecular genetics of the achaete-scute gene complex of *D. melanogaster*. *Cell* 40, 327–338.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
- Frodesen, A.G., Skjeggestad, O., and Tøfte, H. (1979). Probability and Statistics in Particle Physics (Oslo: Universitetsforlaget).
- Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D.L., Houle, D., Charlesworth, B., and Keightley, P.D. (2007). Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445, 82–85.
- Kelley, K. (2007a). Constructing confidence intervals for standardized effect sizes: theory, application, and implementation. *J. Stat. Softw.* 20, 1–24.
- Kelley, K. (2007b). Methods for the behavioral, educational, and social sciences: an R package. *Behav. Res. Methods* 39, 979–984.
- Kuhn, R.M., Karolchik, D., Zweig, A.S., Wang, T., Smith, K.E., Rosenbloom, K.R., Rhead, B., Raney, B.J., Pohl, A., Pheasant, M., et al. (2009). The UCSC genome browser database: update 2009. *Nucleic Acids Res.* 37 (Database issue), D755–D761.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009b). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Nolo, R., Abbott, L.A., and Bellen, H.J. (2001). *Drosophila* Lyra mutations are gain-of-function mutations of senseless. *Genetics* 157, 307–315.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Reh, W., and Scheffler, B. (1996). Significance tests and confidence intervals for coefficients of variation. *Comput. Stat. Data Anal.* 22, 449–452.
- Rodríguez, I., Hernández, R., Modolell, J., and Ruiz-Gómez, M. (1990). Competence to develop sensory organs is temporally and spatially regulated in *Drosophila* epidermal primordia. *EMBO J.* 9, 3583–3592.
- Strutz, T. (2010). Data Fitting and Uncertainty. A Practical Introduction to Weighted Least Squares and Beyond (Germany: Vieweg+Teubner Verlag).
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164.

SELECTION POPULATION SETUP - 1



SELECTION POPULATION SETUP - 2

**Figure S1. Crossing Scheme to Basal Stocks for All Selections, Related to Figure 3**

Left, sc^1 and w^{118} were recombined onto one X chromosome, which was then isogenized to create the basal $sc\text{ w}$ stock. Right, autosomes derived and mixed from three distinct genetic backgrounds were individually or altogether balanced.

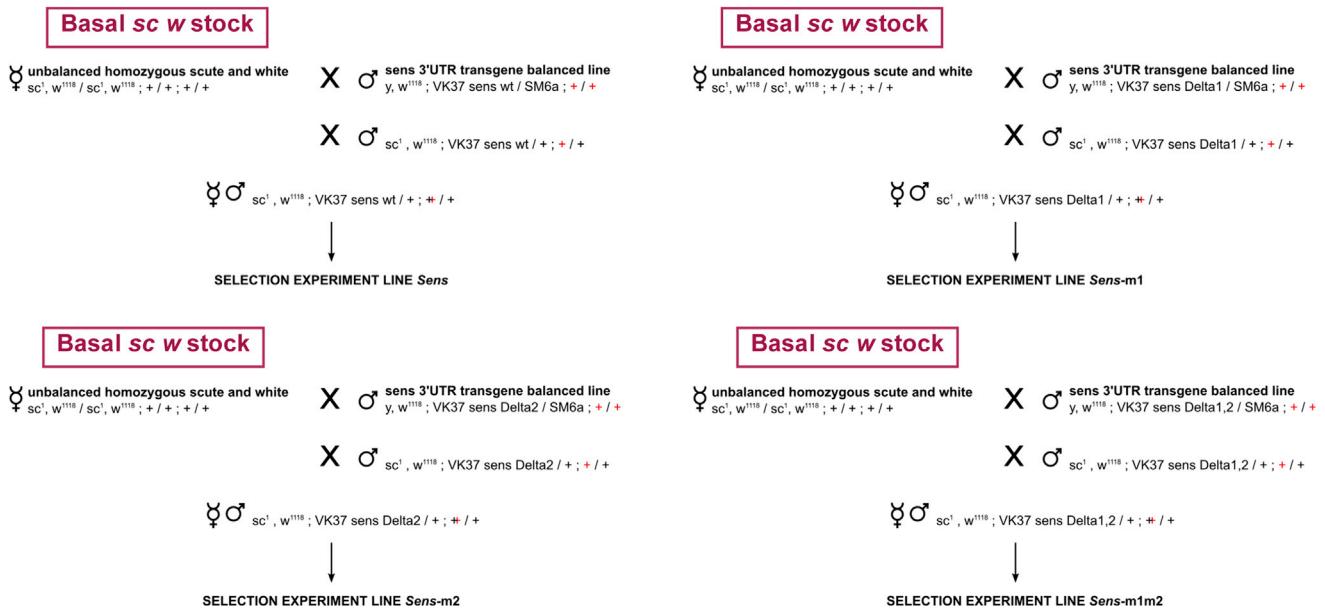


Figure S2. Crossing Schemes to Generate Genomic Backgrounds for *sens* Selection Experiments, Related to Figure 3
The *sens* transgenic stocks were backcrossed to the basal *sc w* stock twice.

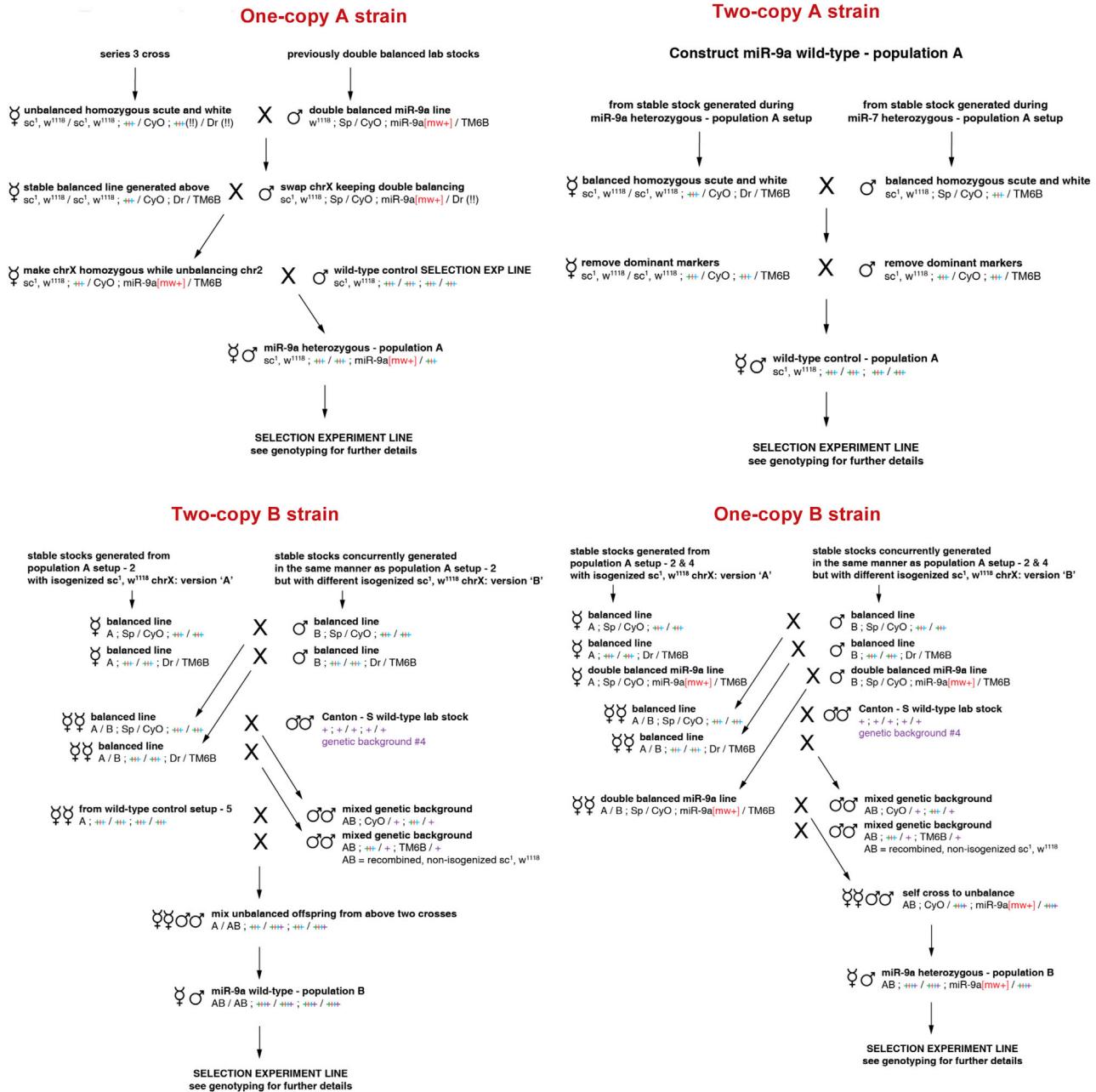


Figure S3. Crossing Schemes to Generate Genomic Backgrounds for *miR-9a* Selection Experiments, Related to Figure 3

Top, the *miR-9a* strains labeled A were constructed with a distinct genetic background from the strains labeled B, shown below. One and two copy refers to the number of wild-type *miR-9a* alleles in the final strains after construction.

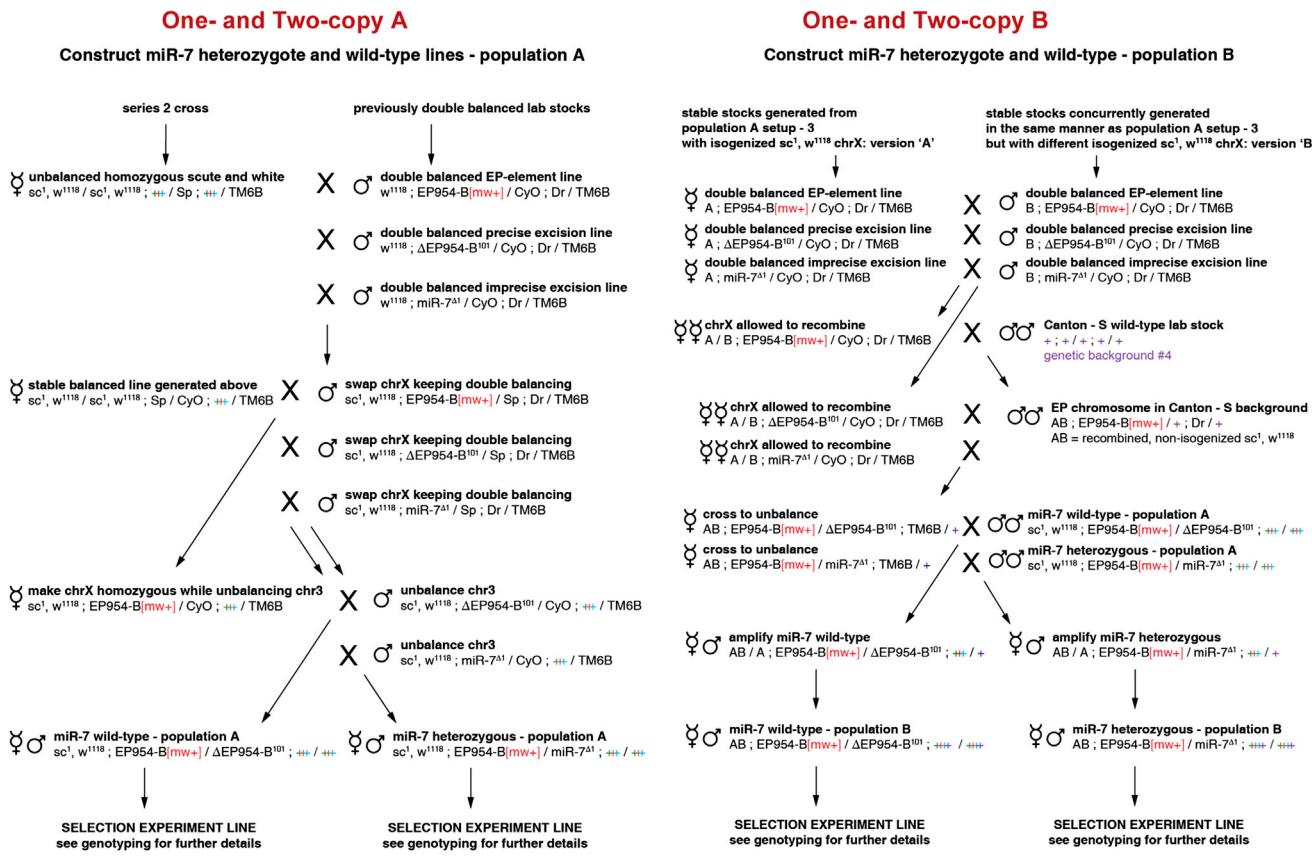


Figure S4. Crossing Schemes to Generate Genomic Backgrounds for *miR-7* Selection Experiments, Related to Figure 3

Left, the *miR-7* strains labeled A were constructed with a distinct genetic background from the strains labeled B, shown right. One and two copy refers to the number of wild-type *miR-7* alleles in the final strains after construction.

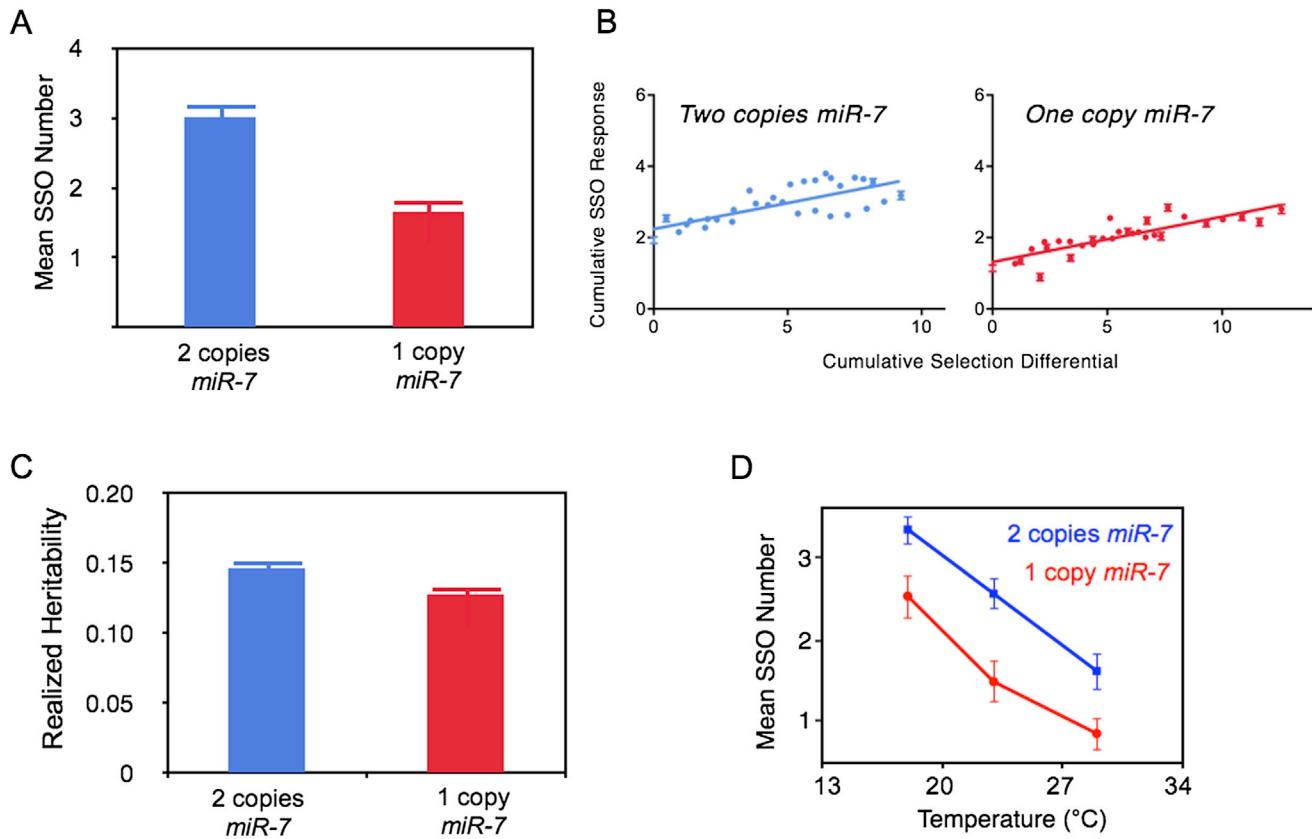


Figure S5. miR-7 Does Not Affect Robustness to Genomic and Temperature Variation, Related to Figure 2

(A) SSO formation is inhibited when miR-7 concentration is reduced. Error bars are standard errors. Means are significantly different ($p < 0.0001$, two-tailed t test).

(B) Response in SSO fate to cumulative selection in strains carrying one or two copies of *miR-7*, as indicated. Each strain had two replicate populations under selection for 15 generations, and their responses are plotted together in each graph. The slopes of the weighted least-squares regression lines are a measure of the realized heritability.

(C) Realized heritability of SSO fate within replicate populations undergoing selection that contained either one or two copies of *miR-7* during selection. Standard errors are shown as error bars. The difference in realized heritability between one- and two-copy strains is not significantly different ($p = 0.101$, z-test).

(D) Change in *miR-7* copy number does not significantly affect the temperature dependence of SSO formation, although copy number has an independent impact on SSO formation. Error bars are standard errors. The ANOVA of this data is further described in Table S4.

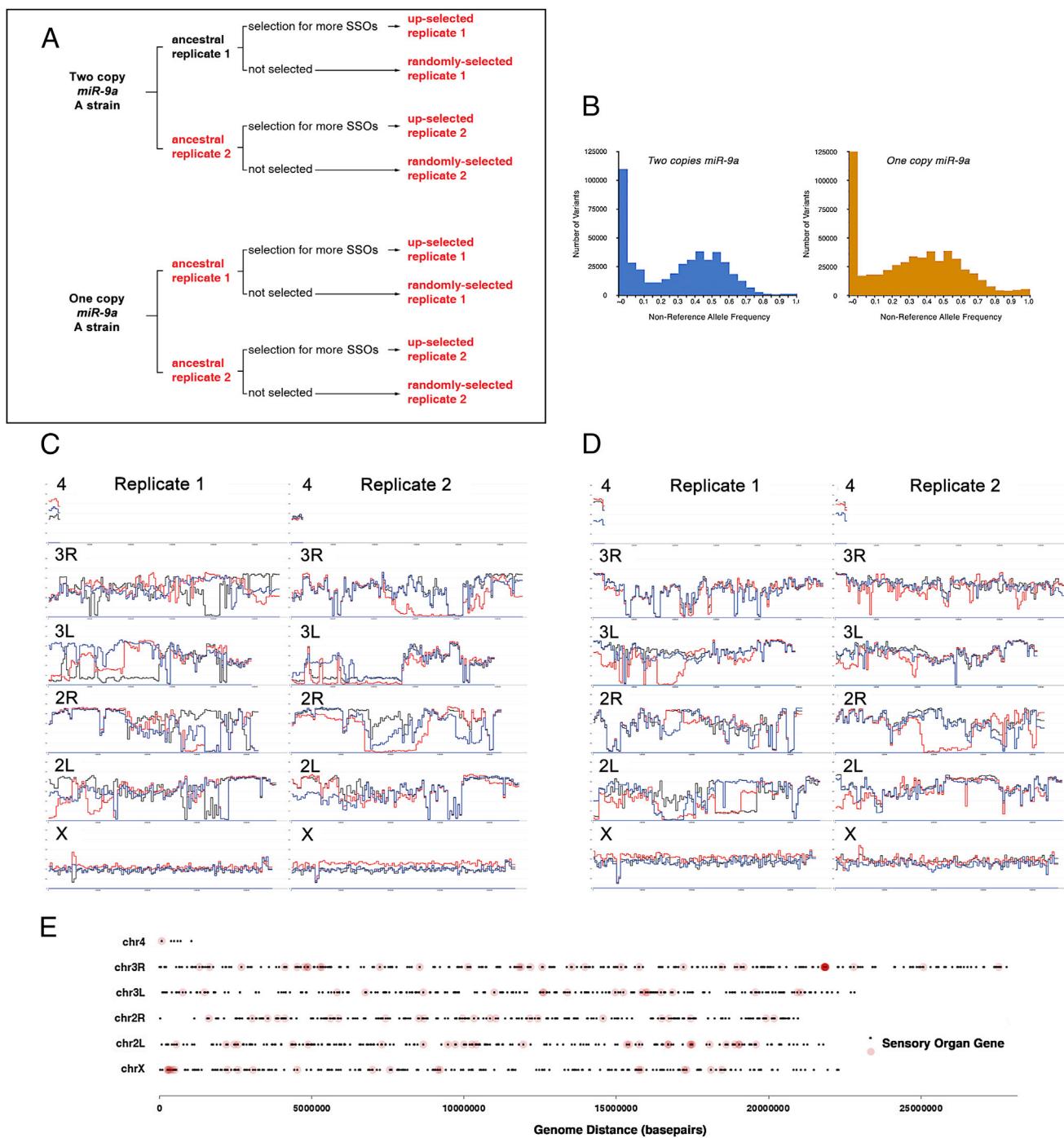


Figure S6. Whole-Genome Sequencing to Identify Genome Variants under Selection, Related to Figure 4

(A) Schematic of populations whose genomes were sequenced. The *miR-9a* A strain was used for sequence analysis. Animals from before selection (ancestral) and after either random or directed selection were pooled and frozen until DNA was extracted for sequence analysis. Populations from which genomes were sequenced are highlighted in red.

(B) Distribution of nonreference allele frequency for variants in the sequenced genomes of ancestral populations before selection. Left, genomes from the ancestral strain with two copies of *miR-9a*. Right, genomes from the ancestral strain with one copy of *miR-9a*. For both charts, variants were binned in frequency intervals of 0.05, and the leftmost columns represent those variants that had no nonreference reads in the sequenced ancestral libraries. Although these variants are denoted ~0, they are not necessarily mono-allelic in the ancestral population. Rather, the nonreference allele could be rare enough that only reference reads were detected from the sequenced library. The total number of variants detected in each strain is slightly different; 467,220 variants in the two-copy strain and 551,228 variants in the one-copy strain. This might account for the slightly elevated distribution of the one-copy histogram. When considering those variants with

(legend continued on next page)

nonreference allele frequencies of < 0.25, then 43.3% of all variants in the two-copy strain fell into that category, and 41.9% of all variants in the one-copy strain fell into that category. We also examined those variants that were shared between the two strains. The nonreference allele frequency distributions for these variants were also highly similar between the sequenced ancestral genomes.

(C) Allele diversity on the y axis for all called variants in *miR-9a* two-copy replicates averaging in 200 kb bins across each chromosome arm, as indicated (x axis). Shown are ancestral (black), nonselected (blue), and upselected (orange) variant diversity. Replicates 1 and 2 are presented.

(D) Allele diversity on the y axis for all called variants in *miR-9a* one-copy replicates averaging in 200 kb bins across each chromosome arm, as indicated (x axis). Shown are ancestral (black), nonselected (blue), and upselected (orange) variant diversity. Replicates 1 and 2 are presented.

(E) Genome distribution of 1,233 genes curated from the published literature as affecting sensory organ formation. Genes described in more than one published study are highlighted in pink. All others are highlighted as black dots.

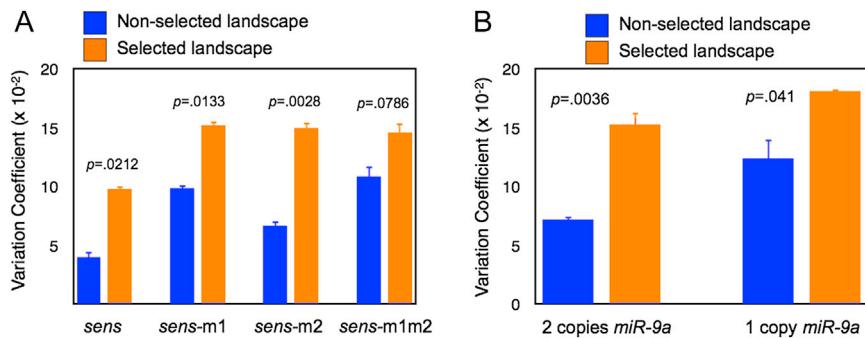


Figure S7. SSO Variation as Measured by Variation Coefficient for Strains in Different Genomic Landscapes, Related to Figure 5

(A) One copy of the *sens* transgene was placed into a genomic landscape in which one copy of the genome had been shaped by 15 generations of selection for more SSOs (Selected). As a control, the transgene was placed into a landscape that had not been selected (nonselected). Transgenes had either normal or mutated miR-9a binding sites as indicated. Resulting individuals were scored for SSOs, and shown is the variation coefficient for each group. Error bars represent standard error. p values are the results of a two-tailed students t test comparing selected and nonselected.

(B) Individuals in which both copies of their genomes had been shaped by 15 generations of selection for more SSOs (Selected) or not (Nonselected). Resulting individuals were scored for SSOs and shown is the variation coefficient for each group. These groups had either one or two copies of the *miR-9a* gene. Error bars represent standard error. p values are the results of a two-tailed students t test comparing selected and nonselected. Note that variation increases in the selected landscape no matter the *miR-9a* or *sens* genotype. However, the variation increases more when either *miR-9a* or *sens* are wild-type. When *miR-9a* or *sens* are impaired, there is less of an increase in variation - in fact it appears as if the variation reaches a "ceiling." Overall, it suggests that the selected genomic variants decanalize SSO number. This occurs in a manner that is not completely additive with the decanalization induced by *miR-9a/sens* mutation. Perhaps their respective effects on canalization are mediated in series rather than in parallel.