Supporting Information

Zhou et al. 10.1073/pnas.1010643108

SI Materials and Methods

Experimental Selection of the Hypoxia-Tolerant *Drosophila* Strain. To generate the hypoxia-tolerant *Drosophila* strain, we pooled 27 wild-type isogenic lines (kindly provided by Andrew Davis) to form the parental population and subjected them to long-term laboratory selection (1). Certain interparental genetic variations of hypoxia tolerance were determined in the parental lines that include a significant variation in eclosion rates under hypoxia and recovery time from anoxic stupor (1). F_1 embryos of this pooled population were collected and cultured at different levels of hypoxia (8%, 6%, or 4% O_2). We found that 6% O_2 dramatically decreased their survival rate, and 4% O_2 was lethal. Under 8% O_2, the majority of the embryos (>80%) completed their development and reached the adult stage. Therefore, we initiated the hypoxia-selection experiment at 8% O_2, and this O_2 concentration was gradually decreased by ~1% every 3–5 generations to maintain selection pressure. After >30 generations of selection, we obtained flies that break through the lethal hypoxia limit and tolerate 4% of O_2 perpetually. To test whether this hypoxia-tolerant trait is heritable, a subset of embryos obtained from the hypoxia-selected flies was collected and cultured under normoxia for several consecutive generations and then reintroduced back into the same hypoxic environment (i.e., 4% O_2); again, the majority (>80%) of these flies completed their development and could be maintained in this extreme condition perpetually. One hundred male and 100 female flies from the 180th generation of hypoxia-selected or control populations were collected and genomic DNA was isolated for sequencing analyses.

Illumina Genome Sequencing Library Preparation and Sequencing. Genomic DNA was isolated from a pool of 100 male and 100 female adult flies collected from hypoxia-selected populations or generation-matched control populations by standard phenol chloroform extraction followed by treatment with DNase-free RNase. DNA quality was assessed by spectrophotometry (260/280 and 260/230) and gel electrophoresis. A total of 3 μg was sheared DNA (Covaris) and was used to construct a library for paired-end sequencing. The DNA fragments were subjected to end repair and phosphorylation by T4 DNA polymerase, Klenow DNA Polymerase, and T4 polynucleotide kinase, respectively, in a single reaction and then ligated to Illumina PE adapters. The adapter-ligated products were purified on Qiaquick spin columns (Qiagen) and then PCR amplified with high-fidelity DNA Polymerase in 12 cycles, using Illumina's PE primer set. PCR products were purified on QiaquickMinElute columns (Qiagen), and the library quality was assessed and quantified using an Agilent DNA 1000 series II assay and a Nanodrop 7500 spectrophotometer (Beckman) and diluted to 10 nM. Cluster generation was performed using the Illumina cluster station and cluster generation kit v2. The 54 + 54 paired-end sequencing was performed using genome analyzer II (Illumina) and sequencing kit v3. The fluorescent images were processed to sequences using the Illumina base-calling pipeline (GA Pipeline-1.4.0). The *Drosophila melanogaster* reference genome, together with the annotation of genes and repeats, was downloaded from the UCSC database (http://genome.ucsc.edu/).

Data Analysis. The next-generation sequencing data for each of the pools were derived from 200 flies descended from 27 parental strains. Neither individual genotypes nor the number of individuals sampled at a region could be determined, precluding use of standard analysis tools to identify differences between control and hypoxia-tolerant populations. We therefore used two complementary analysis methods. One focused on identifying individual loci with high-confidence allelic differences between the control and hypoxia-tolerant populations. The other identified genomic regions characterized by allelic frequencies that differed between control and hypoxia-tolerant flies, representing regions of potential selection. Both analyses used Maq v.0.7.1 (2) under its default parameters to map reads from the four populations (H1, H2, C1, and C2) to the *D. melanogaster* reference genome downloaded from FlyBase (http://www.flybase.org). As in experimental science, the concordance in results obtained by the two methods provides validation not only for the results obtained but also for the methods used.

Test for selection in pooled data. Common tests of selection primarily work through measuring loss of diversity in selected haplotypes (3, 4). In pooled samples, we do not have such data available—for a SNP, the only measure of diversity we can calculate is the SNP's frequency. From the raw mappings to *D. melanogaster* reference release 5.23, we performed a few more processing steps to generate accurate frequencies. We used Maq's rmdup to remove duplicate reads and Maq's cns2snp to identify variant positions. To calculate frequencies, we used a method developed by Holt et al. (5) that measures the frequency of an allele as the fraction of reads covering the locus showing that allele, weighted by their quality scores. We viewed the reference sequence as an outgroup, and thus, any deviation from the reference as a derived allele. To ensure robust frequency estimates, we considered only sites with at least 10× coverage of reads that had a mapping quality of at least 40. This meant that only 93,306,140 loci in H1, 79,726,429 loci in H2, 76,569,897 loci in C1, and 40,708,770 loci in C2 were considered as having enough information to accurately gauge allelic frequencies. We considered only derived alleles with frequencies between 10% and 90% as being polymorphic. This determination resulted in 292,410 SNPs in H1, 288,952 SNPs in H2, 274,200 SNPs in C1, and 97,895 SNPs in C2.

To determine regions under selection, we first developed a metric (the sum of nonfixed derived allele frequencies, F) to estimate the scaled mutation rate $\theta = 4N\mu$. To elaborate, consider a SNP matrix with n individuals and m columns over a region with little to no recombination. Let 1 represent a derived allele and 0 represent an ancestral allele. Derived alleles that are fixed are presumed to occur before the most recent common ancestor and are discarded. Next, consider the coalescent-based genealogy of the sampled population, with the derived alleles marked on the branches. Following Fu (6), we partition time into coalescent epochs with T_k representing the time taken to coalesce from k lineages to $k-1$. Let X_k represent the number of derived alleles that originated during epoch T_k, and let X_{jk} represent the number of derived alleles at site j that originated during epoch T_k. At time T_k, the k lineages have a total of n descendants. By symmetry, the expected number of descendants for any lineage is n/k. On the basis of the coalescent theory, the expected value of the time duration of T_k is $2N/\binom{k}{2}$.

Summing over all branches, the expected number of derived alleles in the SNP matrix can be expressed as follows:

$$E(F\times n) = \sum_{k=2}^{n} E(X_k) = \sum_{k=2}^{n}\sum_{j} E(X_{jk}) = \sum_{k=2}^{n} (k\times\mu\times E(T_K))\binom{n}{k}$$
$$= \sum_{k=2}^{n}\left(k\times\mu\times\frac{2N}{\binom{k}{2}}\right)\binom{n}{k} = 4\times\mu\times N\times n$$

$$E(F) \times n = \theta \times n$$

$$E(F) = \theta.$$

We compute the F statistic for identical regions in flies grown under both normoxia (F_C) and hypoxia (F_H). Under purifying selection for a favorable allele, the effective population size reduces to favor individuals carrying that allele. Thus, a large ratio of F_C to F_H would be indicative of a region under selection. We measure F_C and F_H in identical regions of size 50 kbp across the genome. At this size, the region is large enough to provide robust estimates of θ, yet small enough to not have many recombinations. Assuming that the mutation rate is constant in a region, the log ratio of these two is the log ratio of effective population size:

$$S_f = \log\left(\frac{F_C}{F_H}\right) = \log\left(\frac{\theta_C}{\theta_H}\right) = \log\left(\frac{N_C}{N_H}\right).$$

To correct for genome-wide coverage biases, we first define F_C and F_H as the sum of frequencies across the entire genome. We then add a correction factor corresponding to $\log(F_H / F_C)$ to each window. Even though we expect to see an average of 42 mutations per 50 kbp, the variation in this statistic is fairly high—for several regions, such as those containing repetitive regions, we see very few (and even no) mapped mutations at all. To deal with this, a pseudocount of 0.1 was added to both numerator and denominator of the statistic. The resulting statistic was plotted across the *Drosophila* genome for H1 vs. C1 and H2 vs. C1. As a negative control, C2 vs. C1 was used. A threshold of 4.0 (corresponding to a 1% FDR) was used to characterize positive selection. Fig. S3 shows plots across all major chromosomes for H1 vs. C1, H2 vs. C1, and C2 vs. C1. Genes in regions that were significant in both H1 vs. C1 and H2 vs. C1 were used in subsequent analyses and are available in Table S2.

High-confidence allele calling. An overview of the method is presented in Fig. S1. Using *D. melanogaster* reference genome release 5.16, preliminary analysis established that loci with high-quality base calls (base quality ≥20; best read quality ≥40; and calls limited to A, C, G, or T) showed similar high read consistency for coverages between 20 and 40 (Table S1), whether the base call was the same as or different from the reference base. We therefore examined euchromatin loci that had ≥20× coverage and high-quality base calls in sample pools C1, H1, and H2, which covered ∼45% of euchromatin loci (Table S2) and 70–75% of the exons located in the five major chromosomes (Fig. S2). We disregarded pool C2 because it had much lower 20× coverage than the other three pools (36,594,586 loci in C2 vs. 67,933,396 in C1); furthermore, because neither C1 nor C2 was under selective pressure, only differences due to genetic drift would be anticipated. Scripts were written to parse Maq view output files to identify loci meeting the above constraints in which both the H1 and the H2 base calls differed from both reference and C1 bases. In particular, both the first and the second best Maq calls for H1 and H2 were required to differ from both the first and the second best base calls for C1. In all these cases, the base calls for H1 and H2 were found to agree. Indels were identified using Maq indelpe output files corrected for homopolymer tracts. After eliminating indel positions shared with C1, the set of putative hypoxia tolerance-related indels was further filtered to remove loci with <20× coverage in either the H1 or the H2 pool or in which the percentage of reads mapped with the indel was <50% in either the H1 or the H2 pool. Identified SNPs and indel loci were mapped to genes/gene regions using fasta files downloaded from FlyBase (http://www.flybase.org).

Hypoxia Tolerance and Vulnerability Tests. ***Drosophila stocks.*** The following homozygous stocks were used: *UAS-NICD* (J. Posakony), *Eaat1-GAL4* (Fbst0008849), and *4XSu(H)-lacZ* (kindly provided by J. Posakony). A double-homozygous stock of *UAS-NICD*; *Eaat1-GAL4* (EN-line) was obtained with the following crosses using *Ap/CyO; +/TM3, Sb* (Fbst0002475) (definitions: *n*, *UAS-N*ICD on the third chromosome; *E*, *Eaat1-GAL4* on the second chromosome; *Sb*, TM3 balancer with Stubble marker on the third chromosome; *CyO*, *CyO* balancer on the second chromosome; *Ap*, Apterous, T (2;3) ap [Xa], ap [Xa], on the second chromosome):

Step 1) To generate +/*CyO*; *N*/*N* flies:

P1 ($^{+/+}$; *N*/*N*) × (*Ap*/*CyO*; +/*Sb*)
F_1 select for *only CyO and Sb* flies (+/*CyO*; *N*/*Sb*) × (+/*CyO*; *N*/*Sb*)
F_2 select for *only CyO* flies (+/*CyO*; *N*/*N*).

Step 2) To generate *E*/*E*;+/*Sb* flies:

P1 (*E*/*E*; $^{+/+}$) × (*Ap*/*CyO*; +/*Sb*)
F_1 select for *only CyO and Sb* flies (*E*/*CyO*; +/*Sb*) × (*E*/*CyO*; +/*Sb*)
F_2 select for *only Sb* flies (*E*/*E*; +/*Sb*).

Step 3) Cross F_2s from steps 1 and 2 to each other:

P1 (*E*/*E*; +/*Sb*) × (+/*CyO*; *N*/*N*)
F_1 select for flies that are *only CyO and Sb* (*E*/*CyO*; *N*/*Sb*)
F_2 select the F_1 with the *CyO* and *Sb* markers for double-balanced stock.

To assay for cell type specificity of *N*ICD overexpression homozygous *UAS-N*ICD flies were crossed to *Eaat1-GAL4* flies (Fig. 4*A*). To test for *N*ICD transcriptional up-regulation, double-homozygous *E*/*E*;*N*/*N* males were crossed to *4XSu(H)-lacZ* flies.

Experimental Protocol for Hypoxia Survival of Flies Overexpressing NICD in Glial Cells. Ten virgin female flies homozygous for UAS-NICD were crossed to 5 male flies homozygous for Eaat1-GAL4 and allowed to lay eggs for 48 h in normoxia. The flies were then moved to a control vial (for another 48 h before being discarded) and the vial with the eggs was moved to a 5% oxygen chamber with 12-h dark and 12-h light cycle with a temperature of 22 ± 1 °C. More controls of parental lines were done in parallel. After 4 wk, both sets of flies were assayed for the number of pupal cases that were empty or full. Six vials of each condition were completed in two different experiments for a minimum of 500 pupal cases scored for each condition. To evaluate continued adult survival during hypoxia, all adult flies from the overexpression and two parental lines were moved to the same vial each day for 28 d and a record of flies still alive was recorded.

Immunohistochemical Staining and Fluorescence Microscopy. Details of dissecting, fixing, and staining can be found elsewhere (1, 2). Briefly, brains of wandering third instar larvae were dissected in PBS and fixed in 4% paraformaldehyde in PBS. Cell membranes were permeabilized with 0.3% Triton X-100 in PBS, blocked with 7% goat serum, and put in primary antibody overnight at 4 °C (or for 1 h at room temperature). This procedure was followed by washes in 0.3% Triton X-100 in PBS, incubation with secondary antibody for 90 min, more washes, mounting, and microscopy. Primary antibodies used were mouse anti-NICD (undiluted in-house supernant) [Developmental Studies Hybridoma Bank (DSHB)], rabbit anti-repo (1:500) (kind gift of G. Technau, Institute of Genetics, University of Mainz, Mainz, Germany), rat anti-elav (1:50) (DSHB), and mouse anti-lacZ (1:100) (Pro-

mega). Secondary antibodies used (1:250) were goat anti-mouse; goat anti-rabbit; and goat anti-rat conjugated to Alexa 488, 546, or 647 (Invitrogen). Mounting media were Prolong Gold anti-fade reagent with DAPI (Invitrogen). Confocal microscopy was performed in the University of California at San Diego Neuroscience Microscopy Shared Facility. Imaging was done on a confocal microscope (Olympus FV1000) and the images were processed with Image J.

1. Zhou D, et al. (2007) Experimental selection for Drosophila survival in extremely low O (2) environment. *PLoS ONE* 2:e490.
2. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 18:1851–1858.
3. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709.
4. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
5. Holt KE, et al. (2009) Detecting SNPs and estimating allele frequencies in clonal bacterial populations by sequencing pooled DNA. *Bioinformatics* 25:2074–2075.
6. Fu YX (1997) Coalescent theory for a partially selfing population. *Genetics* 146: 1489–1499.

Other Supporting Information Files

Fig. S1 (PDF)
Fig. S2 (PDF)
Table S1 (DOC)
Table S2 (DOC)
Table S3 (DOC)
Table S4 (DOC)
Table S5 (DOC)