**Web Note A.**

Picking primers for amplifying predicted gene sequences. We followed a sequence of decreasingly stringent tests to pick optimal primer pairs for the PCR fragments on the arrays. For each gene, we looked for (first) a 500b to 2kb sequence at the 3' end and, failing this, (second) two 200b to 500b sequences in:(1) coding sequence, completely within an EST clot, and unduplicated (BLAST$<=10^{-4}$); (2) coding sequence, partially within an EST clot, unduplicated; (3) coding sequence, unduplicated; (4) exon, completely within an EST clot, unduplicated; (5) exon, partially within an EST clot, unduplicated; (6) exon, unduplicated; (7) sequence less than 2kb at the 3' end.

Assigning confidence intervals to the measurements. In order to assign confidence intervals to the developmental changes in expression between stages, we bootstrapped the data set by randomly swapping error terms between measurements and reëstimating the general linear model according to the scheme described by Kerr *et al.*[44]. First we standardized the residuals[47], then estimated a gene specific standard deviation by plotting standard deviation of the standardized residuals for each gene against the estimated gene effect (G_g) for that gene and fitting a loess curve (span = 0.5) to this plot. The standard deviation of the gene was taken to be the value of this curve at that gene's estimated G_g. This avoids the problem of having too few residuals per gene to accurately estimate its standard deviation while allowing for gene specific heteroscedasticity in the data. The standardized residuals were divided by these standard deviations creating a normalized error distribution. New data points were constructed by subtracting the model residual for a particular measurement and replacing it by an error drawn from the standardized residuals multiplied by this estimated standard deviation: $b_{ijkg} = y_{ijkg} - \_{ijkg}+SD_g*e_{i'j'k'g'}$ where b_{ijkg} is the new data point, SD_g is the gene specific standard deviation, and $e_{i'j'k'g'}$ is drawn randomly and with replacement from the normalized error distribution. We repeated this 1000 times, storing the estimated difference $V_{PF}G_g$-$V_{18\ hr\ BPF}G_g$ – the log ratio of the

expression of each gene between the two stages – each time. This produced a distribution for each log ratio, and we designated the 95 percent confidence interval to be between the 2.5 and 97.5 percentiles of this distribution. These confidence intervals span on average 0.73 (average SE 0.13) on a log-2 scale.

A gene's expression changed during development if 0 did not fall within the 95% confidence interval. A gene's expression changed during evolution between two lineages if the developmental changes for the two lineages were significantly different. Since the variance of this evolutionary difference would be the sum of the variances of the developmental change for each population, we extended each confidence interval by a factor of $\sqrt{(1+x)}$ where $x =$ variance(B)/variance(A) for population A and the reciprocal for population B. The variances are the variances of the bootstrap replication distributions for that gene. If the variances are equal, this factor is $\sqrt{2}$. A gene did not change significantly during evolution if the adjusted confidence interval for either population included the estimate of the mean developmental change for the other. Note that for two lineages it is possible for a gene's expression to be developmentally changing in only one of the two and yet not be evolutionarily different between the two.
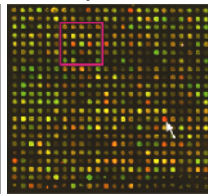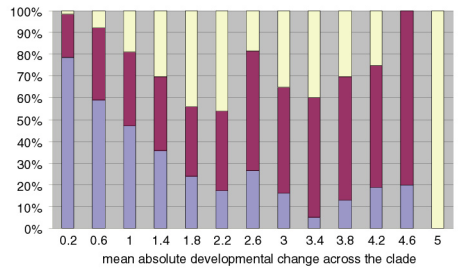
D. mel. CS   *D. simulans*   *D. yakuba*

mean absolute developmental change across the clade
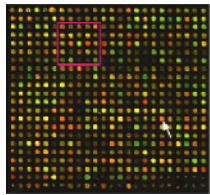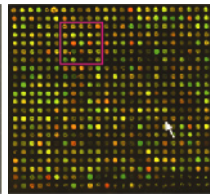
	0.2	0.6	1	1.4	1.8	2.2	2.6	3	3.4	3.8	4.2	4.6	5
neutral	57	121	91	61	55	37	9	13	8	7	4	0	1
lineage-specific selection	836	510	163	69	40	29	27	18	11	13	9	4	0
stabilizing selection	3265	914	227	72	30	14	13	6	1	3	3	1	0

number of genes at each level of developmental change

Web Fig. B

Web Fig. B. Magnitude of developmental change is correlated with evolutionary mode. Each column represents the genes with the average absolute level of developmental change indicated on the x-axis (bin width=0.4 $\log_2$ units). Columns are divided into evolutionary modes percentage-wise. As magnitude of developmental change increases, a greater percentage of genes are evolutionarily changing. The chart below the graph shows the number of genes for each evolutionary mode at each level of developmental change.