Evolution of Gene Expression

1. Molecular basis of expression differences

Expression level differences between individuals or between species could have two genetic causes:

a) *cis* – a change in the regulatory sequence (promoter or enhancer) linked to the target gene.

b) *trans* – a change in a different, unlinked gene (such as a transcription factor) that affects expression of the target gene.

How can we distinguish between these two possibilities?

An experimental approach in Drosophila (Wittkopp *et al.* 2004. Nature 430: 85-88):

The authors made F1 hybrids of *D. melanogaster* and *D. simulans*, then compared the expression level of genes in the two parents plus the hybrid offspring using pyrosequencing This is a very sensitive method that can measure the expression of the two different alleles in a heterozygote. Later studies used the same approach, but employed next generation sequencing technologies.

The expectation:
If expression differences between the parents are caused by *trans*-regulatory divergence, then the two alleles should be expressed equally in hybrids.

If expression differences between the parents are caused by *cis* regulatory divergence, the two alleles should be expressed differently in the hybrid and match the expression difference between the parents. This is also called Allele-Specific Expression (ASE).
Results:
29 genes were tested that differed in expression between the two parental species.
28 (97%) showed a difference in alllelic expression in the hybrids (*cis* changes).

For about 50% of these genes, the expression difference could be explained entirely by *cis* changes. For the other 50%, there appeared to be both *cis* and *trans* changes.

Conclusion:
Gene expression differences between species are caused mainly by *cis* changes, but *trans* changes are also frequently involved.

Similar experiments in yeast indicate that gene expression differences between alleles from *Saccharomyces cerevisiae* and *S. paradoxus* in F1 hybrids are mostly caused by *cis*-regulatory divergence. However, the relative contribution of *cis* vs. *trans* differences to gene expression divergence depends on the environmental conditions (media) used for the experiment.

Reference: Tirosh *et al.* (2009) A yeast hybrid provides insight into the evolution of gene expression regulation. Nature 324: 659-662.

2. DDT resistance in Drosophila

Reference: Daborn *et al.*, 2002. Science 297: 2253-2256.

Insecticide resistance provides one of the best examples of "evolution in action".

Since the 1940's, humans have used insecticides (such as DDT) to control insect pests (such as mosquitos). Typically, the insects rapidly evolve resistance. This occurs not only in the target species, but also in other insect species exposed to the insecticide. For example, some Drosophila flies have also evolved DDT resistance. With the genetic and genomic resources

available for Drosophila, it is possible to map and identify the genes responsible for DDT resistance (DDT-R).

In *D. melanogaster*, DDT-R was mapped to a cytochrome P450 gene, *Cyp6g1*. Cytochrome P450's are a large family of related genes (≈90) involved in the metabolism of many compounds.

In DDT-R flies, there is a transposable element insertion at the 5' end of the *Cyp6g1* gene. DDT-S flies (susceptible) do not have this TE insertion. DDT-R flies have higher expression of the *Cyp6g1* gene than DDT-S flies. This suggests that the TE insertion increases gene expression and leads to DDT resistance.

In this example, the TE insertion appears to be beneficial to the host and drives adaptive evolution.

3. Population Transcriptomics

The comparison of global gene expression levels among individuals from natural populations of a species. This is a combination of population genetics and transcriptomics.

Natural selection indirectly changes the frequencies of genotypes within populations by sorting among the phenotypes they influence. Transcriptomic methods, such as microarrays and RNA-seq, allow the large-scale quantitative measurement a phenotypes (the expression level of a gene). Gene expression is sometimes considered an "intermediate phenotype" because it lies between the genotype and the organismal phenotype that ultimately responds to selection.

Drosophila melanogaster has an ancestral species range in sub-Saharan Africa and has only recently spread to Europe and the rest of the world (within the past 15,000 years).

If the expression divergence of a gene between African and European flies is the result of recent adaptive *cis*-regulatory evolution, then DNA sequence polymorphism in (or near) the gene should be affected in two ways:

a) DNA sequence polymorphism should be reduced in at least one of the populations, producing a signal of a "selective sweep"

b) There should be one or more fixed (or nearly fixed) sequence differences between the populations. That is, polymorphisms at high frequency in one population, but low frequency (or absent) in the other.

4. Example: *MtnA*

The brain transcriptomes have been compared between males and females and between European and African *D. melanogaster*. This revealed that there were more expression differences in the brain between populations than between sexes and most genes that differed in expression between populations showed the same pattern in both sexes. This is different from what is seen in whole flies.

A gene that is expressed differently between populations in the brain is the metallothionein gene *MtnA*, which is involved in heavy metal detoxification and oxidative stress tolerance. The expression difference appears to be caused by the deletion of a negative regulatory element in the *MtnA* 3' UTR, which may be a binding site for a microRNA. The deletion is in high frequency outside of Africa and shows evidence for positive selection in European populations. The deletion shows a clinal pattern on multiple continents, with the frequency increasing with distance from the equator. The deletion (and high *MtnA* expression) are associated with increased oxidative stress tolerance. As with *CG9509*, the deletion remains polymorphic in most cosmopolitan populations, suggesting that there may be a selective trade-off and it may be subject to balancing selection.

5. Example: *CG9509* (aka *fezzik*)
An example is the gene *CG9509*, which consistently shows 2–3 times higher expression in non-African ("cosmopolitan") flies than in African flies. There is an absence of sequence polymorphism in the genomic region just upstream of *CG9509* in the European population. There are also several fixed sequence difference between the European and the African populations in this region. These observations suggest that selection favored increased expression of *CG9509* in the European population and that the difference in expression is caused by sequence divergence in the *CG9509* upstream region.

How can this be tested?

A "reporter gene", in which the *CG9509* upstream region from either an African or a European allele is placed in front of the *E. coli lacZ* gene (encoding beta-galactosidase) can be constructed *in vitro*.

The reporter genes can then be inserted into a precise location of the *D. melanogaster* genome using a method known as PhiC31 site-specific integration. With this approach, the influence of the two different upstream regions (African and European) on gene expression can be compared in the same genetic background. If there is a difference in expression between the two reporter genes, then it must be the result of *cis*-regulatory divergence.

For *CG9509*, the European upstream region drives 2–3 times greater expression than the African upstream region. This indicates that the expression difference seen between the natural populations can be completely explained by *cis*-regulatory sequence divergence.

Interestingly, the SNP that has the greatest effect on *CG9509* expression is at intermediate frequency (40–50%) in cosmopolitan population, but at very low frequency (0–5%) in sub-Saharan African populations. This does not fit the classical model of a selective sweep, but instead suggests that the polymorphism might be maintained by balancing selection in cosmopolitan populations.

The exact function of *CG9509* is unknown, but it is predicted to encode a choline dehydogenease enzyme and to be involved in ecdysteroid metabolism (growth hormone). It is expressed specifically in the Malphigian tubules, which are the analog of "kidneys" in insects. Studies using knock-out mutations or RNA interference have shown that *CG9509* expression influences stress tolerance and larval/adult growth. Because *CG9509* knock-outs/knock-downs are larger than wild-type flies, the gene has been named *fezzik*, after a giant character in *The Princess Bride*.

6. Example: Allele-specific expression (ASE) in Malpighian tubule
By comparing gene expression in Swedish and Zambian *D. melanogaster*, as well as ASE in their offspring, the contribution of *cis*-regulatory variation to gene expression variation in natural populations could be estimated.

Overall, much more variation could be explained by *trans* factors than *cis* factors. There was an enrichment of cis-acting variation in cytochrome P450 genes. Transgenic reporter genes indicated that most *cis*-variants in P450 geneswere located in the region just upstream of the gene (within 2 kb of the start codon).

References:
Catalán, A., S. Hutter, and J. Parsch (2012) Population and sex differences in *Drosophila melanogaster* brain gene expression. BMC Genomics 13: 654.

Catalán, A., A. Glaser-Schmitt, E. Argyridou, P. Duchen, and J. Parsch (2016) An indel polymorphism in the *MtnA* 3' untranslated region is associated with gene expression variation and local adaptation in *Drosophila melanogaster*. PLoS Genet. 12(4):e1005987.

Glaser-Schmitt A., and J. Parsch (2018) Functional characterization of adaptive variation within a *cis*-regulatory element influencing *Drosophila melanogaster* growth. PLoS Biol 16: e2004538.

Ramnarine, T. J. S., A. Glaser-Schmitt, A. Catalán, and J Parsch (2019) Population genetic and functional analysis of a cis-regulatory polymorphism in the *Drosophila melanogaster Metallothionein A* gene. Genes (Basel) 10: 147.

Glaser-Schmitt, A., A. Zečić, and J. Parsch (2018) Gene regulatory variation in *Drosophila melanogaster* renal tissue. Genetics 210: 287-301.