Evolution of Gene Expression

1. Comparative Transcriptomics
Comparison of expression levels between closely-related species. Can be done using microarrays or RNA-seq.

Typically, microarrays are based on the genome sequence of one species. However, they can also be used for closely-related species if the DNA sequence is very similar and can still hybridize to the array. One has to be careful that DNA sequence divergence does not affect hybridization.

Example:
Comparison of *D. melanogaster* and *D. simulans* using cDNA microarrays (≈6,000 genes). Compared both males and females.
Reference: Ranz *et al.* (2003). Science 300: 1742-1745.

Results:
52% of the genes showed a significant expression difference between the two species in at least one sex.

Most of these (83%) showed differences in expression between the sexes in at least one species (that is, they were sex-biased genes).

Genes that were expressed equally in males and females (unbiased genes) were less likely to change in expression level between species.

What types of genes differ in expression between *D. melanogaster* and *D simulans*? *D. simulans* males show higher expression of genes involved in vision, while *D. melanogaster* males show higher expression of genes involved in olfaction. These may be related to differences in mating (both species show strong mate-choice preference for their own species).

2. 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.

3. 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.

4. 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.

The earliest population transcriptomics experiments was performed in yeast
Reference: Cavalieri *et al.*, 2000. Proc Natl Acad Sci USA 97:12369-12374.

A lot is known about yeast genetics and genomics, however most of our knowledge comes from the study of a single laboratory strain (or derivatives of it). Much less is known about "wild" yeast in nature.

Yeast cells are rarely found on intact grapes, but often found growing in damaged grapes. Cavalieri *et al.* collected natural yeast isolates from a vineyard in Montalcino, Tuscany, Italy. They found that there is variation in physical and biochemical traits among isolates. One of these isolates was used for transcriptional analysis.

The strain, designated M28, shows independent 2:2 segregation of spores for:

1. resistance to trifluoroleucine
2. resistance to copper sulfate
3. colony morphology (filigreed vs. smooth). Filigreed (F) = cells do not completely separate following replication, form long strands instead of single cells. Smooth (S) = grow as single cells and form smooth, round colonies when grown on agar plates.

How much variation is there in the transcriptome?
The Approach:
Compared M28 F progeny and S progeny using cDNA microarrays (≈6,200 genes).

The Results:
378 genes had 2-fold or greater difference in expression between F and S (≈6% of genes).
As a control, only 8 genes differed by 2-fold or more when different samples of F and F progeny were compared (≈0.1% of genes)

Which genes differ?
19 of the 26 most over-expressed (3-fold or greater) genes in F are involved in amino acid synthesis (12 of them for methionine). Two others are for sulfate transport. This is probably an important adaptation that leads to resistance to sulfates, which are often used as a fungicide on grapes.

The number of differentially expressed genes detected between two samples depends on the quality of the data, the number of replicates, and the statistical method used to define differential expression. Thus, it can vary substantially from experiment to experiment. In yeast, most studies have found ≈2–8% of genes differ in expression between any two strains.

5. Expression quantitative trait loci (eQTL)

Because the level of expression of a given gene is a phenotype that can be measured on a continuous scale, it can be treated like other quantitative traits and one can look for associations between DNA sequence polymorphisms and gene expression levels. This is typically done on a genome-wide scale using a combination of transcriptomic data (microarray or RNA-seq) and genotyping data ("SNP-chips" or genome resequencing). SNPs that show a significant correlation with gene expression are considered eQTL.

It is possible to classify eQTLs as *cis*- or *trans*-acting. This is done by comparing the location of the SNP and the gene whose expression it affects. If the SNP is on a different chromosome than the gene (or far away on the same chromosome), it can be considered *trans*-acting. If it is on the same chromosome and near the gene, it can be considered *cis*-acting. Note that there is no strict definition for "far away" and "near" when dealing with SNPs and genes on the same chromosome. This will depend on the species and the amount of linkage disequilibrium (LD). For species with high LD, like humans, the cut-off distance may be up to 1 MB. For species with low LD, like Drosophila, it may be 10-20 kb.

An early eQTL study of humans (cell lines) identified many more *cis*-eQTL (1,348) than *trans*-eQTL (180).

Reference: Stranger et al. (2007) Population genomics of human gene expression. Nature Genetics 39: 1217-1224.

However, other studies found more *trans*- than *cis*- eQTL.

It is difficult to compare studies that use different methods, experimental designs and samples. If there is a general pattern emerging from human eQTL studies, it is that there are more *trans*-eQTL detected than *cis*-eQTL. However, *cis*-eQTL tend to have larger effects on gene expression and, thus, are detected with greater statistical confidence (lower *p*-values). This means that the proportion of *cis*- vs. *trans*- eQTL that is detected will depend on the statistical power of the experimental design and the *p*-value threshold used to define significance.