January 2017 (Ruben Rellan-Alvarez)

Renny-Byfield, Rodgers-Melnick & Ross-Ibarra Gene fractionation and function in the ancient subgenomes of maize

Genome duplication leading to different levels of ploidy is common among plants species. Maize is not an exception and around 10 million years ago experienced a whole genome duplication. Since then many paralogous genes have been lost and many returned to single copy status. This fractionation process has not been proportional in the two different subgenomes. Genes on the most fractionated subgenome tend to be lower expressed, probably by differences in TE accumulation and methylation levels. Michael Freeling and collaborators proposed back in 2012 a model whereby higher expression of a gene in a paralog pair leads to higher protein accumulation and a ultimately a higher contrinution to phenotypes.
In this manuscript Renny-Byfield, Rodgers-Melnick & Ross-Ibarra use publicly available gene expression, phenotypic and epigenetic data to test this hypothesis.

First, they use a set of around 3K paralogous genes present in both maize 1 (less fractioned) and maize 2 (most fractioned) subgenomes and a set of 45 phenotypic traits and calculated heritabilities for both pairs of genes. Most of the traits have higher heritabilities when heritabilities were calculated in the less fractioned subgenome. Contribution to heritabilities of singleton (only present in one of the genomes) genes from the maize 1 were also higher. Interestingly, heritabilities of flowering related traits like days to anthesis and days to silking, showed no differences between subgenomes when either paralogous gene pairs or singletons were used.

They then asked if these differences in contributions to phenotype heritabilities between genes from different subgenomes were supported by differences in gene expression. Using the maize gene expression atlas data they found that highly expressed paralogous have a higher contribution to heritabilities than their lower expresed pairs. However, when the same analysis was done with a set of randomly selected pairs the same results were obtained. In summary, highly expressed genes have a higher contribution to phenotypes' heritabilities. Again, in the case of flowering time related traits, there were no differences in the level phenotypic varibility explained by upregulated or dowregulated paralogous genes, but in the random set heritabilities calculated from upregulated genes was signficantly higher than their dowregulated counterparts.

Using epigenetic data (cytosine methylation, histone modifications and nucleosome occupancy) they show that maize 1 has slightly lower methylation upstream the transcription start site when either paralogous pairs or singletons are examined. However, histone modification and nucleosome occupancy data showed no differences between the subgenomes.

Finally they examined the relationship between epigenetic marks and expression data using the same set of 3K paralogous genes and random set with differences in expression. In general methylation was higher in the random set of genes than in the paralogous pair group. Methylation was higher in downregulated genes than in the upregulated ones with the exception of CHH methylated sites on the random set. Highly expressed genes also have more open chromatin than the lowly expressed ones.

Overall the authors show that paralogous genes on the maize 1 subgenome explain more heritability for a variety of phenotypes and than

this effect is more likely explained by higher levels of expression regardless of the subgenome of origin something that is also true in the case of non-paralogous genes. The differences in expression that ultimately drive phenotypic varibility are at least in part explained by differences in epigenetic marks between high and low expresed genes.

It will be interesting in the future to explore if these changes in expression and epigenetic marks between subgenomes are different between different maize inbreds and if they have been the drivers of local adaptation in maize. The publication of new genomes, expression datasets and phenotypic data in the near future should open the possibility for such studies. This paper nicely shows how one car learn relevant aspects of maize biology and evolution from publicly available data like the one available at MaizeGDB and highlights the importance of proper data publication and curation.