Rampant Replacements of Heterochromatin-Binding Proteins in Fly Gonads

August 20, 2012

ME Arnegard

By combining evolutionary tree reconstruction with the analysis of genomic data, phylogenomics aims to infer the functions of new genes as well as the processes by which genomes evolve. Sequencing genomes of several related species within different groups of organisms has led to the identification of numerous protein-coding and RNA-coding genes, as well as noncoding regulatory sequences, residing in the euchromatin of eukaryotes. In comparison to gene-rich euchromatin, eukaryotic heterochromatin is gene-poor yet plays key roles in chromosome segregation, telomere stability and gene regulation. Direct phylogenomic analysis of heterochromatin is hindered by its highly repetitive nature, which makes it very difficult to assemble heterochromatic DNA into long contiguous blocks of meaningful sequence. Fortunately, the laboratory of Dr. Harmit Malik (Basic Sciences Division) has developed an alternative approach to the direct study of heterochromatin, in which researchers instead focus on the euchromatic genes that encode heterochromatin-binding proteins (Vermaak *et al.*, 2009). Using this "surrogate approach," postdoctoral researcher Dr. Mia Levine and co-authors recently discovered 21 new genes in the Heterochromatin Protein 1 (*HP1*) gene family, which have undergone a revolving-door-like pattern of gene replacement across 40 million years of evolution in fruit flies (*Drosophila* species).

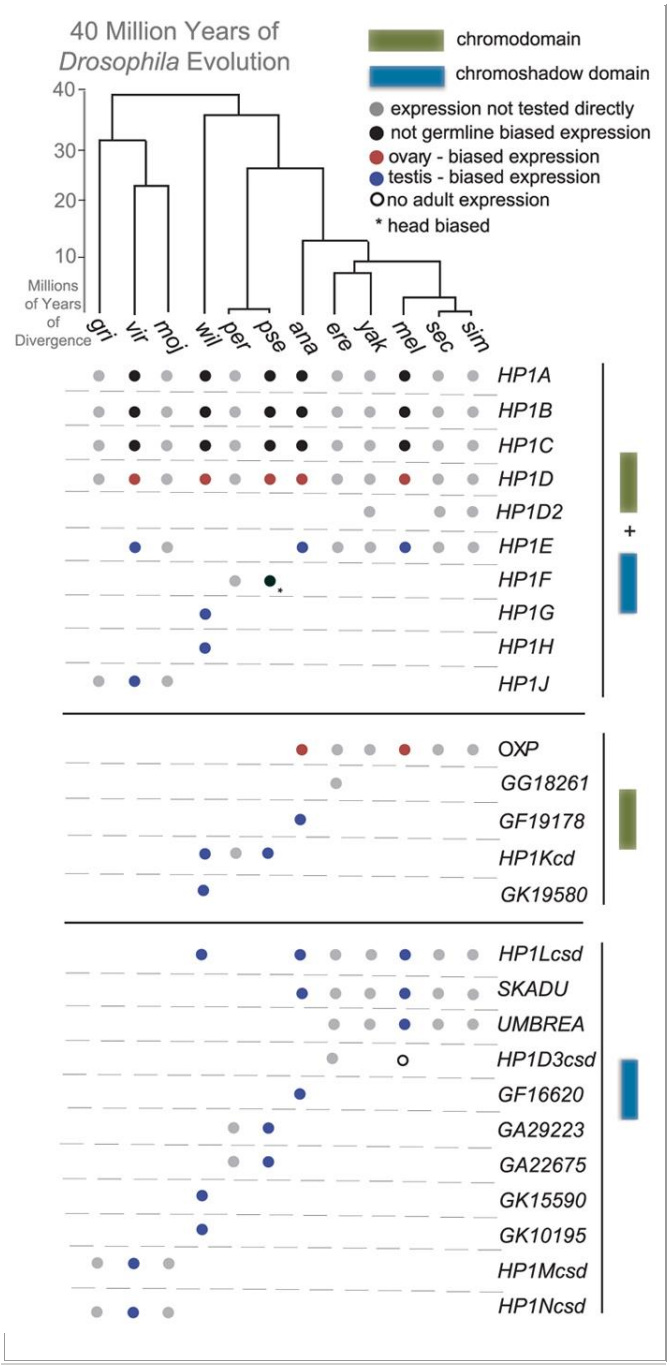
The *HP1* gene family encodes some of the best-known surrogates for understanding heterochromatin function. For instance, mutant versions of the prototypical family member, *HP1A*, helped to illuminate the critical role of heterochromatin in chromosome segregation. *HP1A* and the four other previously known members of the *HP1* gene family code for both a chromatin-binding "chromodomain" (CD) that recognizes histone modifications and a "chromoshadow domain" (CSD) that mediates protein-protein interactions.

Levine *et al.* made use of data from eleven fly genomes that have been sequenced since the original *Drosophila melanogaster* genome was published in March of 2000. The authors sleuthed for *HP1*-like genes in this dataset using the Basic Local Alignment Search Tool, or BLAST algorithm, in combination with a variety of phylogenomic analyses and statistical tests for selection. The newly discovered*HP1* genes exhibited surprisingly high structural variation. Some were full-length, like the archetypal *HP1* genes, containing blocks of sequence encoding both the CD and CSD; however,

most were partial (CD-only or CSD-only) genes. The researchers showed that open reading frames have been preserved in these genes for millions of years, without any mutations degrading them into pseudogenes. Despite this rampant pattern of genetic innovation, Levine *et al.* found a lack of positive selection acting on the new genes, similar to what had been shown for the original family members. Yet, unlike the founding *HP1* genes, which are broadly expressed, the new genes tend to exhibit restricted expression in the gonads of few species, providing a valuable new toolkit for dissecting heterochromatin function in *Drosophila* germline tissue. Most strikingly of all, Levine *et al.* demonstrated a dynamic pattern of gene birth and death, with *HP1* gene number remaining rather constant across species, suggesting that this genetic turnover is associated with some functional biological process happening in the gonads. Finally, the authors noted that loss of testes-restricted *HP1E* from *D. persimilis*and *D. pseudoobscura*, without their replacement by other family members, was associated with the appearance of a new Y chromosome in these species, and a corresponding structural change in male heterochromatin. Based on this intriguing association, the authors propose that bursts of chromosome-localizing protein evolution may follow major chromosomal rearrangements or shifts in heterochromatin-euchromatin boundaries.

Levine MT, McCoy C, Vermaak D, Lee YCG, Hiatt MA, Matsen FA, Malik HS. 2012. Phylogenomic analysis reveals dynamic evolutionary history of the Drosophila Heterochromatin Protein 1 (HP1) gene family. *PLoS Genetics* 8:e1002729.

Also see: Vermaak D, Bayes JJ, Malik HS. 2009. A surrogate approach to study the evolution of noncoding DNA elements that organize eukaryotic genomes. *J. Hered.* 100:624-636.



Modified from the authors' figure in PLoS Genetics, an open access journal

Summary of HP1-like genes identified in the authors' phylogenomic screen. The Drosophila species studied (slanted text consisting of three-letter codes) represent 40 million years of evolution, as shown by the time-calibrated tree (top left). The better-studied species include D. persimils (per), D. pseudoobscura (pse) and D. melanogaster (mel). Gene names are listed at the right: HP1A through HP1E were known prior to this study; the other 21 genes are new to science. Corresponding protein structure (i.e., domain composition) and expression pattern are color coded, as indicated in the key at the top right. Note the consistency of gene number in any given species, combined with the pervasive birth-death dynamics across the broader tree. Much like a "revolving door" of gene replacement, one gene emerges along a lineage as another is lost from a given structural category of HP1-like genes.