

Poster A-11

Identification and Analysis of Retroid Agents in the Insect Genomes: *Drosophila* *melanogaster*, *Anopheles gambiae*, and *Apis mellifera*.



Authors:

Holly A. Basta (*Montana State University, Bozeman, MT*)

Marcella A. McClure (*Montana State University, Bozeman, MT*)

Short Abstract: Retroid agents (RAs) integrate or transpose in a host organism's DNA via reverse transcription (RT) of RNA. The Genome Parsing Suite was developed to identify, characterize, and annotate the RAs that encode potential RTs in any genome. A summary of the RAs found in three insects will be presented.

Long Abstract:

Retroid agents use reverse transcriptase (RT) to transcribe their RNA into DNA, which then integrates or transposes into the DNA of a host organism. Once falsely labeled "junk" DNA, some Retroid agents are implicated in disease, while others are beneficial to their resident organisms. Due to the evolutionary and developmental importance of Retroid agents, the McClure Lab developed the Genome Parsing Suite (GPS) to identify and characterize RT signals in any genome database, and to annotate the Retroid agents that encode these potential RTs. The GPS is a collection of integrated programs that finds RT signals and then proceeds to identify presence, order, and quality of all genes expected for a specific Retroid genome. The results from the GPS program provide comparative analyses between species or organisms, in this case, the differences in Retroid agents found in *Anopheles gambiae*, *Drosophila melanogaster*, and *Apis mellifera*. Stage 1 GPS performs the RT analysis, which includes an output of raw hits, unique hits, and perfect hits. Raw hits are all the hits returned from WUBLAST (Washington University, Basic Local Alignment Search Tool) [1], unique hits are raw hits with corrections for multiple hits on one gene and minus cross-coverage due to multiple queries, and perfect hits are unique hits that contain all six motifs with out any frame-shifts (FS) or stop codons (SC). Stage 2 GPS performs the full agent analysis, with includes documentation of full, FS, one SC, and perfect agents. Full agents contain all genes necessary to a complete Retroid genome defined by the component library. Cellular translational reprogramming allows Retroid agents with one FS or SC to be transcribed and translated. Retroid genomes, therefore, with one FS or SC, as well as perfect genomes are considered potentially active. To date, the GPS is the only software that can systematically identify and classify all Retroid agents in a given genome, while providing the user with various summary tables, as well as all DNA sequences for gene components (long terminal repeats, genes, etc.)[2].

We are currently conducting a trans-organismal study of Retroid agents in all complete and ordered genomes. A relational database for the results from GPS analysis is under construction to facilitate the detailed comparison of all Retroid agents (within and between organisms) and to determine the role these agents have played in the evolution of the Eukaryotic genome.

The Phylum Arthropoda is the most species-rich and morphologically diverse animal group

on the planet. Since their appearance in the early Cambrian and their subsequent radiation, arthropods have come to inhabit and dominate the vast majority of ecological habitats. This study focuses on the Insecta group, with two representatives from the order Diptera (*A. gambiae* and *D. melanogaster*), and one from the order Hymenoptera (*A. mellifera*). To help achieve a global view of the role Retroviral agents play in Eukaryotic organisms, a detailed analysis of the Retroviral agents present in three insect species: the mosquito (*A. gambiae*), the fruit fly (*D. melanogaster*), and the honeybee (*A. mellifera*) will be presented. These three insects are each of particular biological interest; the fruit fly as the model of genetics, the mosquito as the vector of malaria, and the honeybee as the world's primary honey producer. While the fruit fly is a favorite experimental model system, the mosquito genome has been sequenced with a more defensive objective; *A. gambiae* is the primary carrier of the malaria parasite, *Plasmodium falciparum*. Malaria is responsible for up to 2.7 million deaths each year, and the number of cases is expected to double in the next twenty years if new methods of control are not devised and effectively executed [3]. *A. mellifera* was chosen as the first species of the order Hymenoptera to be sequenced because of its potential impact on sociogenomics. To date few studies have analyzed the Retroviral agents present in this genome.

A detailed comparison of the Retroviral agents in the three insect genomes allows classification of the identified Retroviral agents as one of three types of insertion: (1) insertion before the divergence of the Dipteran and Hymenopteran groups, (2) insertion after the divergence of Diptera and Hymenoptera, or (3) insertion after the split of the fly and mosquito lineages. The mitochondrial genomes of the three insects will also be analyzed, as well as the genome of *P. falciparum* in order to assess the presence of Retroviral agents in the endosymbionts and parasites of these insects.

While international research has focused on the Retroviral agents of *D. melanogaster* and *A. gambiae*, very little is known about Retroviral agents in the honeybee. By comparing the results of the *A. mellifera* analysis with those of *A. gambiae* and *D. melanogaster*, the Retroviral agents found in these genomes can be classified as either conserved through evolutionary time, or species specific.

[1] W.R. Gish, <http://blast.wustl.edu>, 1996-2002

[2] McClure, M.A., Richardson, H.S., Clinton, R.A., Hepp, C.M., Crowther, B.A., Donaldson, E.F., (2005) "Automated characterization of potentially active Retroviral agents in the human genome". *Genomics* 85(2005) 512-523.

[3] Gardner, M. J. et. al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. 419, 498-511(2002).