# Investigating Horizontal Transfer of Transposable Elements in Drosophilidae Genomes

Supervisor
Aurélie Hua-Van, EGCE, Building 680, IDEEV, 12 route 128, 91190 Gif-sur-Yvette
aurelie.hua-van@universite-paris-saclay.fr,

## SCIENTIFIC CONTEXT:

Transposable elements (TEs) are mobile genetic elements capable of moving and replicating within host genomes, often independently of the host's fitness. Due to their selfish and invasive nature, TEs are frequently considered genomic parasites (Hua-Van et al., 2011). All eukaryotic genomes harbor TEs, which can constitute over 80% of their total DNA content. Within a single species, several hundred to several thousand distinct TE families may coexist. These families can differ markedly in copy number, age, genomic distribution, and transpositional activity.

While some TE insertions can be co-opted by the host and contribute to evolutionary innovation, the majority are neutral or deleterious. To mitigate their impact, host organisms have evolved regulatory mechanisms—such as the piRNA pathway in metazoans—that suppress TE activity and limit transposition (Brennecke et al., 2007). Furthermore, since TEs encode no essential functions for the host cell, individual insertions can accumulate mutations that are not purged by natural selection. Over time, the control mechanisms, combined with the accumulation of mutations, lead to the inactivation and eventual extinction of TE families.

Despite these constraints, active TE families persist in all eukaryotic genomes, in part due to the recurrent horizontal transfer of TEs (HTT) between species. Unlike conventional genes, TEs can move between reproductively isolated lineages, and comparative genomic studies have revealed extensive HTT events, even between phylogenetically distant taxa (Peccoud et al., 2016). While the mechanisms of HTT remain poorly understood, vectors such as viruses, bacteria, or parasitic interactions are suspected. Ecological and geographic proximity are also thought to facilitate HT events, through shared environments, hosts, or symbionts.

## OBJECTIVES:

The goal of this internship is to assess how geographic, ecological, and phylogenetic proximity influence the rate and success of TE horizontal transfers. Using a comparative genomics and bioinformatics approach, the intern will analyze a unique and exceptionally rich dataset comprising genome assemblies from over 300 Drosophilidae species (Suvorov et al., 2022). The Drosophilidae family, which diverged approximately 50–60 million years ago, includes more than 4,000 described species and exhibits remarkable ecological and morphological diversity. This dataset offers an unparalleled resource in terms of phylogenetic breadth, ecological variation, and genomic quality, providing a powerful framework for investigating TE dynamics on an evolutionary scale.

This project will address key questions such as: Are horizontal transfers more frequent between species that coexist geographically? Do ecological traits influence the likelihood of TE transmission between species?

## METHODOLOGIES:

The intern will adapt a pipeline combining TE annotation, comparative genomics, and statistical modeling to investigate horizontal transfer (HT) patterns across the Drosophilidae phylogeny. The main steps include:

1. TE Annotation
   TEs will be identified using both de novo and homology-based tools (e.g., RepeatModeler, RepeatMasker), with custom libraries built as needed for better accuracy.
2. Family Classification
   TE sequences will be clustered into families based on sequence similarity using tools like mafft or vsearch.
3. HT Detection
   Horizontal transfer events will be inferred by comparing TE divergence with host species divergence, using methods adapted from Wallau et al. (2016) and related approaches.
4. Correlation Analyses
   The intern will test for associations between HT frequency and factors such as phylogenetic distance, geographic overlap, and ecological similarity, using generalized linear and phylogenetically informed models.
5. Robustness Testing
   Permutation tests will be used to assess the reliability of detected associations by comparing observed results to randomized expectations.

We are looking for a motivated student with:

- A background in evolutionary biology, genomics, or bioinformatics.
- Basic proficiency in scripting languages (e.g., Python, R, or Bash).
- Familiarity with sequence alignment and phylogenetic tools is a plus.
- Interest in genome evolution, mobile elements, and species interactions.

## REFERENCES:

Brennecke, J.; Aravin, A. A.; Stark, A.; Dus, M.; Kellis, M.; Sachidanandam, R.; Hannon, G. J. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell 2007, 128 (6), 1089–1103. https://doi.org/10.1016/j.cell.2007.01.043.

Hua-Van, A.; Le Rouzic, A.; Boutin, T. S.; Filée, J.; Capy, P. The Struggle for Life of the Genome's Selfish Architects. Biol Direct 2011, 6, 19. https://doi.org/10.1186/1745-6150-6-19.

Peccoud, J.; Loiseau, V.; Cordaux, R.; Gilbert, C. Massive Horizontal Transfer of Transposable Elements in Insects. Proc Natl Acad Sci U S A 2017, 114 (18), 4721–4726. https://doi.org/10.1073/pnas.1621178114.

Suvorov, A.; Kim, B. Y.; Wang, J.; Armstrong, E. E.; Peede, D.; D'Agostino, E. R. R.; Price, D. K.; Waddell, P. J.; Lang, M.; Courtier-Orgogozo, V.; David, J. R.; Petrov, D.; Matute, D. R.; Schrider, D. R.; Comeault, A. A.

Widespread Introgression across a Phylogeny of 155 Drosophila Genomes. Current Biology 2022, 32 (1), 111-123.e5. https://doi.org/10.1016/j.cub.2021.10.052.

Wallau, G. L.; Capy, P.; Loreto, E.; Le Rouzic, A.; Hua-Van, A. VHICA, a New Method to Discriminate between Vertical and Horizontal Transposon Transfer: Application to the Mariner Family within Drosophila. Molecular biology and evolution 2016, 33 (4), 1094–1109.