



Research module project (Forschungsmodul)

Improve a *de novo* TE library

Background: *De novo* genes are genes that emerge from scratch from noncoding regions of a genome. To do this they first need to gain both transcription and an open reading frame. Recent studies have shown that transposable elements (i.e. mobile genetic elements that can move and amplify in the genome) could play a role in *de novo* gene emergence by providing regulatory motifs that could enable novel transcription events. To study TEs, it is essential to have a good TE annotation. While manual curation of a TE library ensures the best quality, it is time consuming and requires a good knowledge of TE biology. There are some recent programs that aim to automatise this process. Here we want to test, how much these tools can improve TE annotation.

Objectives: The student will get *de novo* TE libraries created for genomes from six species of *Drosophila*. They will then use tools (such as TETrimmer or MCHelper) that aim to improve these TE libraries and re-annotate TEs in the respective genomes using the curated library. Finally they will compare the non-curated with the curated annotation to see if/how much it improved.

Requirements:

- Experience (or interest) in using the command line
- Experience in working with R and/or python

Methods:

- TE annotation
- Data analysis and plotting with python and/or R

Supervision: Marie Lebherz, Room 100.20, m.lebherz@uni-muenster.de, Molecular Evolution and Bioinformatics Group (<https://bornberglab.org/people/lebherz/>).

Selected Literature:

- 1) Lebherz, M. K., Fouks, B., Schmidt, J., Bornberg-Bauer, E., & Grandchamp, A. (2024). DNA Transposons Favor De Novo Transcript Emergence Through Enrichment of Transcription Factor Binding Motifs. *Genome Biology and Evolution*, 16(7), evae134.
- 2) Qian, J., Xue, H., Ou, S., Storer, J. M., Fuertauer, L., Wildermuth, M. C., ... & Panstruga, R. (2024). TETrimmer: a novel tool to automate the manual curation of transposable elements. *bioRxiv*, 2024-06.
- 3) Orozco-Arias, S., Sierra, P., Durbin, R., & González Pérez, J. (2023). MCHelper automatically curates transposable element libraries across species.