

Research module project (Forschungsmodul)

Comparison of two methods to detect *de novo* genes

Background: While it is well established, that new genes can emerge from existing genes, more recent research showed another way in which new genes can emerge *de novo* from previously non-coding DNA. There are different ways to detect such *de novo* emerged genes. Generally two criteria are usually required: (I) The candidate *de novo* gene should show no homology to any other gene in outgroup genomes. (II) There should be a noncoding homologs in synteny in more closely related outgroups confirming the *de novo* status of the gene. Still, so far there is no standardised way of detecting *de novo* genes and detection pipelines as well as the definition of what can be called a *de novo* gene varies between studies. Here we aim to use two strategies to detect *de novo* genes in a dataset of *Drosophila* genomes. Following this the results from both strategies will be compared.

Objectives: In this project we aim to compare different methods to detect *de novo* genes. To do this, the student will use a dataset consisting of 37 genomes deriving from different geographical locations from in total six *Drosophila* species. These genomes were already annotated using reference-based methods but so far they were not annotated including *ab initio* methods, making use of the available RNAseq data. The student will first re-annotate these genomes (or if time is limited: a subset of these genomes). To do this they will use an annotation pipeline that includes *ab initio* gene prediction. Following this the student will detect *de novo* genes with two different methods (one based on annotations and one based on transcriptomes) and compare the results. Finally, the results can be visualised in plots using python or R.

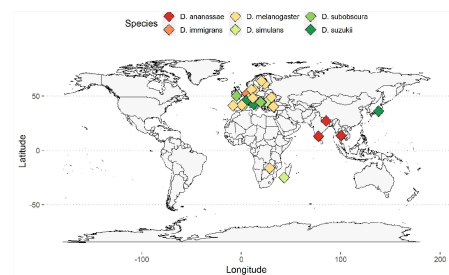


Figure 1: The geographical origins of the *Drosophila* strains that will be used.

Requirements:

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

Methods:

- Genome annotation
- *De novo* gene detection
- R/Python for plotting

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

Selected Literature:

- 1) Roginski, P., Grandchamp, A., Quignot, C., & Lopes, A. (2024). De N ovo Emerged Gene Search in Eukaryotes with DENSE. *Genome Biology and Evolution*, 16(8), evae159.
- 2) Grandchamp, A., Kühn, L., Lebherz, M., Brüggemann, K., Parsch, J., & Bornberg-Bauer, E. (2023). Population genomics reveals mechanisms and dynamics of *de novo* expressed open reading frame emergence in *Drosophila melanogaster*. *Genome Research*, 33(6), 872-890.