Bornberglab

Molecular Evolution and Bioinformatics

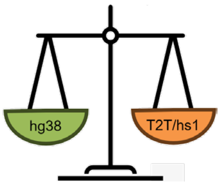
Bachelor's Thesis/MSc FOM/MSc Thesis

De Novo Gene Annotation

Assessing the Impact of Assembly Quality on De Novo *Gene Annotation*

Background: With the advancement of new sequencing technologies, more complete genomes are available for us to ask and answer biological questions. Because of this enhanced completeness, we expect this would allow greater power to detect de novo genes compared to older assemblies. This has not yet been formally evaluated.

BSc/FOM Objectives**:** In this project the student will aim to formally test the hypothesis that more contiguous assemblies identify more *de novo* genes. The student will use two published human genomes assemblies (hg38 and hs1), extract only data related to chromosome 19, and compare the number of *de novo* genes identified using DENSE via the command line. Lastly, the student will use various methods to characterize which *de novo* genes were differentially found between the two assemblies. If time remains, the student can expand this objective by investigating other chromosomes or with additional assemblies from either the 1000 genome project or other species.



MSc thesis Objectives: In this project the student will aim to formally test the hypothesis that more contiguous assemblies identify more *de novo* genes. The student will use two published human genomes assemblies (hg38 and hs1) and compare the number of *de novo* genes identified using DENSE via the command line across the whole genome. Then the student will use various methods to characterize which *de novo* genes were differentially found between the two assemblies. This analysis will be repeated using a second *de novo* gene tool DESwoMAN. If time remains, the student can expand this further by repeating this analysis using old vs T2T genomes of other primates and benchmarking expression levels used in DESwoMAN.

Requirements:

- Interest in genomics and bioinformatics
- Experience or willingness to learn coding

Methods: The student will learn some of the comparative genomic techniques including file manipulation, use of phylostratigraphy for *de novo* gene identification using DENSE, and if time permitting, methods to characterize the new *de novo* genes.

Supervision: Sarah Lucas, Room 100.19, s.lucas@uni-muenster.de, Molecular Evolution and Bioinformatics Group http://bornberglab.org/people/Lucas).

Selected Literature:

1. Zhao L, Svetec N, Begun DJ. De Novo Genes. Annu Rev Genet. 2024 Nov;58(1):211-232. doi: 10.1146/annurev-genet-111523-102413. Epub 2024 Nov 14. PMID: 39088850; PMCID: PMC12051474.
2. Roginski P, et al. De Novo Emerged Gene Search in Eukaryotes with DENSE. Genome Biol Evol. 2024 Aug 5;16(8):evae159. doi: 10.1093/gbe/evae159. PMID: 39212967; PMCID: PMC11363675.
3. Nurk S, Koren S, et al. The complete sequence of a human genome. Science. 2022 Apr;376(6588):44-53. doi: 10.1126/science.abj6987. Epub 2022 Mar 31. PMID: 35357919; PMCID: PMC9186530.
4. Yoo Det al. Complete sequencing of ape genomes. Nature. 2025 May;641(8062):401-418. doi: 10.1038/s41586-025-08816-3. Epub 2025 Apr 9. PMID: 40205052; PMCID: PMC12058530.