Bornberglab

Molecular Evolution and Bioinformatics

***Bachelor's Thesis/MSc FOM/MSc Thesis***

# *De Novo* Gene Maturation

## *Investigating when multi-exon* de novo *genes gain their second exon*

**Background:** Genomics changes can turn ancestrally non-genic sequence into a *de novo* gene. Much of the literature so far has focused on describing the mechanism through which *de novo* genes are formed. While there are still many open questions regarding *de novo* gene formation, very little research has been done to investigate how a *de novo* gene could mature into a canonical protein.

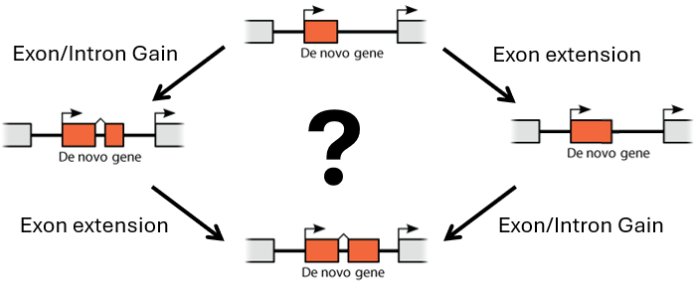
BSc/FOM Objectives**:** In this project the student will aim to formally test the hypothesis if multi-exon *de novo* genes gain their additional exon after speciation. For this work, students will examine published human genomes from the 1000 genome consortium. For chromosome 19, the students will compare the nucleotide diversity ($\pi$) of single and multi-exon genes to determine if there is an elevated nucleotide diversity for multi-exon compared to single exon *de novo* genes. If time remains, the student can expand this work by investigating other chromosomes or human populations.

MSc thesis Objectives: In this project the student will aim to formally test the hypothesis if multi-exon *de novo* genes gain their additional exon for multiple time points. In addition to comparing $\pi$ between multi-exon and single exon *de novo* genes across all chromosomes, the student may study if multi-exon *de novo* genes gained their second exon at the time of speciation with ancestral reconstruction or if they gained their second exon prior to speciation by comparing average phyloP scores.

**Requirements:**

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

**Methods:** The student will learn some of the population genomics techniques including file manipulation, calculation of $\pi$, and statistical comparison techniques.




**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. Schmitz JF, Bornberg-Bauer E. Fact or fiction: updates on how protein-coding genes might emerge de novo from previously non-coding DNA. F1000Res. 2017 Jan 19;6:57. doi: 10.12688/f1000research.10079.1. PMID: 28163910; PMCID: PMC5247788.
3. Sorek R. The birth of new exons: mechanisms and evolutionary consequences. RNA. 2007 Oct;13(10):1603-8. doi: 10.1261/rna.682507. Epub 2007 Aug 20. PMID: 17709368; PMCID: PMC1986822.
4. Ali F. Patterns of Change in Nucleotide Diversity Over Gene Length. Genome Biol Evol. 2024 Apr 2;16(4):evae078. doi: 10.1093/gbe/evae078. PMID: 38608148; PMCID: PMC11040516.