



Understanding *de novo* gene emergence using a stochastic mathematical model of mutagenesis, selection and population growth

Background: New protein coding genes can arise from non-genic genomic regions, through a phenomenon called *de novo* gene emergence. To evolve into a protein coding gene, a DNA region must gain sequence features that allow transcription as well as translation. These features include a promoter, a polyadenylation signal, and an open reading frame. The protein coding sequence can further evolve such that its composition and biochemical properties change. If the expressed protein is beneficial to the host organism, the gene will be fixed in the populations via positive selection. Conversely, if the protein is toxic the gene would be lost in the populations via negative selection. Random mutations that allow gain of transcription and translation features during *de novo* gene emergence, can also cause the loss of newly emerged genes. It is difficult to generate experimental data to understand these phenomena because of the large time scale of the evolutionary processes. Mathematical models are great tools to understand phenomena for which data availability is limited. They can also provide a strong lead for future experimental studies.

Objective: The aim of the project is to understand how mutation and selection plays a role in *de novo* gene emergence. The student will develop a mathematical model of *de novo* emergence and perform simulations to understand evolutionary outcomes. Specifically, they will develop stochastic models of mutagenesis, selection, and population growth, that use probability of different events to more accurately represent evolutionary processes (for example random mutations). The student will use Julia programming language to develop the model and simulate it. This project will allow them to not only understand evolutionary processes but also modeling and simulation techniques, that are widely applied in various areas ranging from biology to economics.

Methods: Mathematical modeling, programming, stochastic simulations

Requirements:

- A good basic knowledge and/or a strong interest in probability theory, and mathematical biology
- Basic programming experience
- Ability to work independently

Supervision: Dr. Bharat Ravi, b.ravi@uni-muenster.de, Molecular Evolution and Bioinformatics Group (<https://bornberglab.org/>)

Relevant literature:

- Van Oss SB, Carvunis AR. *De novo gene birth*. [PLoS Genet. \(2019\)](#)
- Iyengar BR, Bornberg-Bauer E. *Neutral models of de novo gene emergence suggest that gene evolution has a preferred trajectory*. [bioRxiv \(2023\)](#).
- Wilkinson, D. Stochastic modelling for quantitative description of heterogeneous biological systems. [Nat Rev Genet \(2009\)](#).