

Presentation title: Exploring Polygenic Adaptation Patterns in Chironomids: A Combined Bioinformatic Approach on Real-Life Data

Corresponding Author Name: Cosima Caliendo

Affiliation: Johannes-Gutenberg Universität Mainz

Email ID's: caliendo@uni-mainz.de

WhatsApp No.: +0049 176 99660025

Other Authors if any: Prof. Dr Susanne Gerber, Prof. Dr. Markus Pfenninger

Presentation type: Poster presentation



Abstract:

Rapid polygenic adaptation is a critical evolutionary process that allows populations to rapidly adjust to changing environments through subtle shifts in many genes. Pinpointing the genetic basis of this process is therefore a challenge. This study addresses the exploration of rapid adaptation by analyzing gene frequency changes in a controlled multigenerational laboratory experiment of the midge *Chironomus riparius* through an unsupervised Machine Learning Classifier algorithm (One-Class Support Vector Machines).

We established four replicate populations and selected them for reduced emergence time over seven generations. Allele frequency changes were analyzed through cost-effective pool sequencing (time series data).

Anomalous allele frequency changes potentially linked to adaptation were identified by combining a boundary-based classification algorithm, One-Class Support Vector Machines and a hypothesis testing method, Fisher's Exact Test. This proposed novel pipeline for anomaly detection leverages the strengths of two complementary methods. One-Class Support Vector Machine with a radial kernel effectively learns the inherent data structure, enabling the identification of anomalies that deviate from the expected pattern. Furthermore, Fisher's exact test statistically validates these anomalies, ensuring they represent significant deviations from the reference population. This combined approach enhances the robustness and interpretability of anomaly detection compared to traditional methods. We identified candidate loci that exhibited significant allele frequency changes likely associated with the observed rapid adaptation in emergence time.

This study demonstrates the utility of a combined bioinformatic approach for analyzing real-life, time-series data from a controlled selection experiment to obtain insights into the genetic basis of rapid adaptation in *C. riparius*.

Biography:

With a background in biotechnology and expertise in applied bioinformatics, I developed a strong interest into statistical analysis of biological data. My recent work involved a real-life population genetics experiment, where the interplay between statistics and machine learning became a central focus of the data analysis process. Analyzing the data with both statistical and machine learning techniques revealed the potential for a combined approach, harnessing the complementary strengths of each method. As part of the Emergent Algorithmic Intelligence Project at the University of Mainz, I'm interested in the computational exploration of biological data, recognizing its crucial role in analysis success. Looking ahead, I would like to expand my understanding of time series data prediction, enhancing my capability to discern meaningful patterns within biological systems. Alongside my PhD, I've gained expertise in NGS data analysis, mass spectrometry analysis, laboratory work, teaching, conducting tutorials, and project coordination, as well as experience in programming languages of Python, R and Bash scripting, augmenting my skill set for comprehensive scientific inquiry. During my PhD journey, I've gained expertise in NGS data – and mass spectrometry analysis, lab work, teaching, and project coordination. Alongside these, I am experienced in the programming languages Python, R, and Bash scripting.