

Selective sweep localization through object detection

Positive selection leaves traces in the genomes, called selective sweeps. A selective sweep appears as a distinctive pattern of diversity that differs from the rest of the chromosome. Detecting selective sweeps finds practical applications in designing more effective drug treatments by understanding how viruses evolve. Existing methods for selective sweep detection, however, are confounded by demographic changes. To overcome this, a currently active research direction in population genetics explores convolutional neural networks (CNNs) to scan chromosomes for selective sweeps. All approaches so far attempt to approach this as an object classification problem, i.e., distinguishing between neutral genomic regions and selective sweeps. This however does not allow to localize a sweep along a chromosome. The proposed project will build upon the idea of using CNNs but for object detection instead of classification. Images will represent entire population instead of subgenomic regions and the goal of the project will be to explore whether state-of-the-art object detection frameworks can be used for selective sweep detection.

The expected workload is 8 weeks, as follows:

- a) Study our CNN-based sweep-detection framework (1 week)
- b) Literature review on methods and frameworks for object detection and choose one (2 weeks)
- c) Extend our framework to generate data suitable for the object detection framework (1 week)
- d) Prepare run scripts and perform tests to assess performance – simulated data with known ground truth will be provided (2 weeks)
- e) Produce a conference paper on the method and the results (2 weeks)

Contact:

Supervisor (CAES group): Nikolaos Alachiotis, n.alachiotis@utwente.nl, room ZI 5035