

A deep learning approach in population genetics: inferring selection

M2 internship (or long M1) 3 to 6 months

Laboratoire de Recherche en Informatique (Paris-Saclay)

Flora Jay flora.jay@lri.fr and Jean Cury

Our lab is designing and implementing deep learning approaches tailored to population genetics. In particular we are interested in inferring the demographic and adaptive history of populations using genomic data of human or bacterial samples.

This internship aims at testing and contributing to a deep learning method for inferring selection.

Selective pressures can act on a population for many generations in the past and leave patterns in genetic data of present-day individuals. Many methods have been developed to identify selection from these patterns. Recently, deep neural networks have been used to automatically detect selection from a “matrix” / “image” of genetic markers sequenced in multiple individuals [1] One of them is currently in development in the lab [2], and this internship aims at testing its performances under various conditions and contributing to its improvement.

Main Goals

- Compare the method(s) to state-of-the-art approaches based on machine learning and expert features, such as SWIF(r) [3].
- Simulate **ancient samples** or **low quality** modern samples and test the robustness of the method(s) for inferring selection from those data rather than high quality modern DNA.
- Contribute to the design of network architectures better calibrated to real datasets.

This work will be funded by the HFSP international collaboration with Emilia Huerta Sanchez (U Brown, USA) and Maria Avila Arcos (UNAM, Mexico).

Profil

M2 (ou M1) student, machine learning, biostatistics, bioinformatic, math/info, ...

Required skills

Programmation Python, (R)

Machine Learning, Statistical analyses

Knowledgeable or eager to learn about biology and population genetics

Starting date

Preferably February. Salary: gratification

Biblio

Flagel, Lex, Yaniv Brandvain, and Daniel R. Schrider. "The unreasonable effectiveness of convolutional neural networks in population genetic inference." *Molecular biology and evolution* 36.2 (2018): 220-238.

Cury et al. (2019) Back to the future of bacterial population genomics. Oral presentation at ESEB 2019 <https://app.oxfordabstracts.com/events/653/program-app/submission/123100>

Poster at DS3 2018: <http://2018.ds3-datascience-polytechnique.fr/wp-content/uploads/2018/06/DS3-529.pdf>

Sugden, Lauren Alpert, et al. "Localization of adaptive variants in human genomes using averaged one-dependence estimation." *Nature communications* 9.1 (2018): 703.