Validation of machine learning approach for direct mutation rate estimation

Katarzyna Burda[1] and Mateusz Konczal[2]

[1]Adam Mickiewicz University
[2]Affiliation not available

February 27, 2023

Abstract

Mutations are the primary source of all genetic variation. Knowledge about their rates is critical for any evolutionary genetic analyses, but for a long time, that knowledge has remained elusive and indirectly inferred. In recent years, parent-offspring comparisons have yielded the first direct mutation rate estimates. The analyses are, however, challenging due to high rate of false positives and no consensus regarding standardized filtering of candidate de novo mutations. Here, we validate the application of a machine learning approach for such a task and estimate the mutation rate for the guppy (Poecilia reticulata), a model species in eco-evolutionary studies. We sequenced 4 parents and 20 offspring, followed by screening their genomes for de novo mutations. The initial large number of candidate de novo mutations was hard-filtered to remove false-positive results. These results were compared with mutation rate estimated with a supervised machine learning approach. Both approaches were followed by molecular validation of all candidate de novo mutations and yielded similar results. The ML method uniquely identified 3 mutations, but overall required more work and had higher rates of false positives and false negatives. We, thus, recommend its application if most of the mutations are expected to be identified or in case of experiment-specific biases. Both methods concordantly showed that guppy mutation rate is among the lowest directly estimated mutation rates in vertebrates. Similarly, low estimates were obtained for two other teleost fishes. We discuss potential explanations for such a pattern, as well as future utility and limitations of machine-learning approaches.

Hosted file

`Burda_MLmutaitons_MER_FINAL.docx` available at https://authorea.com/users/498740/articles/626645-validation-of-machine-learning-approach-for-direct-mutation-rate-estimation