Signals of selection in primates due to viruses

Master's thesis in bioinformatics

Carlota Myhre de Gouveia

Supervisors:
Thomas Bataillon
Marjolaine Rousselle

Bioinformatics Research Center

Aarhus University

June, 2021

Contents

Abstract

Viruses interact with thousands of their host proteins in mammals, which are known as vírus interacting proteins (VIP). It has been shown that even though VIPs are under a strong selective pressure to conserve their function, they account for a big part of the protein adaptation in mammals. The aim of this project was to analyse the intragenomic determinants for signals of protein adaptive evolution in VIP genes in primates. Genes were divided based on their recombination rate to explore correlation with statistics related to the efficiency of positive and purifying selection, and ensure that the signal of strong adaptation in VIP is robust to differences of this possible confounding factor. Using a hierarchical probabilistic method based on the McDonald-Kreitman test (MK), the distribution of fitness effects (DFE) was estimated allowing to estimate the proportion of adaptive substitutions (α) together with the rate of adaptive and non-adaptive substitutions (w_a and w_{na}) and the proportion of *de novo* advantageous and slightly deleterious mutations, in order to quantify the selective forces acting upon the analysed genes. This was accomplished by generating non-synonymous and synonymous site frequency spectra (SFS) from polymorphisms data in human, chimpanzee, gorilla, orangutan and macaques and combining it with non-synonymous and synonymous divergence between those species.

The results of this study revealed that adaptation observed in VIP genes increases with recombination which indicates that recombination enables a more efficient positive selection and aids in the evolutionary arms race against viruses. It was also shown that VIP genes have a smaller proportion of *de novo* non-adaptive mutations compared to nonVIP genes and that it is not dependent on recombination which confirms the strong selective pressure that they are under to conserve their function. Signals for the proportion of adaptive *de novo* mutations weren't conclusive and reflect the sensitivity of the method to the number of sampled polymorphisms.