Supplementary Material - AmazonForest: In-silico meta-prediction of pathogenic variants

Helber Palheta [1], Wanderson Gonçalves e Gonçalves [1,2], Leonardo Brito [1], Arthur Ribeiro dos Santos [1], Marlon dos Reis Matsumoto [1], Ândrea Ribeiro-dos-Santos [1,2] and Gilderlanio Santana Araújo [1,*]

[1] Laboratory of Human and Medical Genetics, Institute of Biological Sciences, Graduate Program of Genetics and Molecular Biology, Federal University of Pará, Belém 66075-110, Brazil; gilderlanio@gmail.com
[2] Research Center on Oncology, Graduate Program of Oncology and Medical Science, Federal University of Pará, Belém 66073-000, Brazil; akelyufpa@gmail.com
* Correspondence: gilderlanio@gmail.com

Version November 18, 2020 submitted to Journal Not Specified

Abstract: Here we detailed the business process model and notation, use cases and usage examples of AmazonForest.

1. AmazonForest: use cases, web architecture and business process model and notation

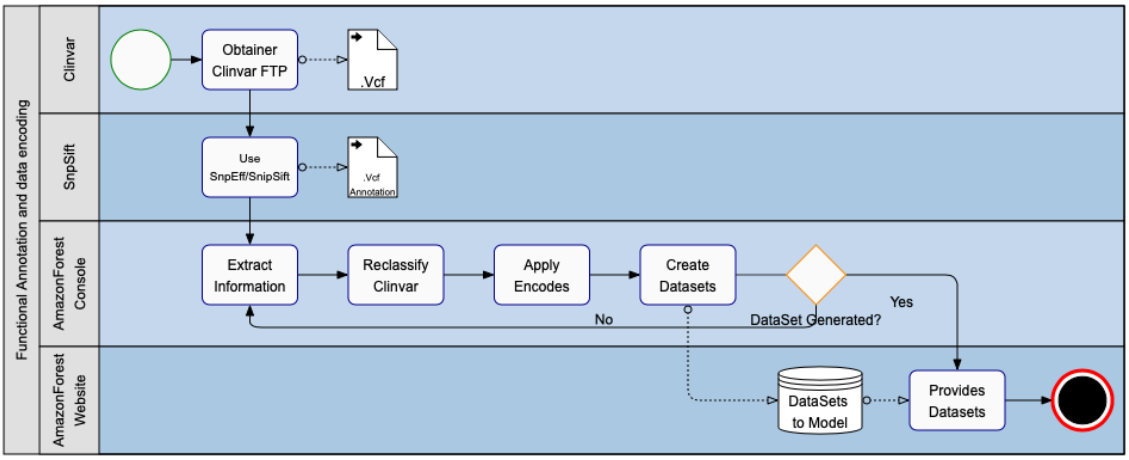

Figure 1. AmazonForest data flow starts by extracting ClinVar *.vcf* file available via File Transfer Protocol (FTP). SnpSift/SnpEff is used for functional variant annotation of ClinVar *.vcf* file for eight in silico predictors. AmazonForest extracts information from *.vcf* annotated file and performs relabel of ClinVar pathogenic classification. Data transformation is performed by encoding strategies. Finally, the datasets are available on the web system for further training and reclassification analysis.

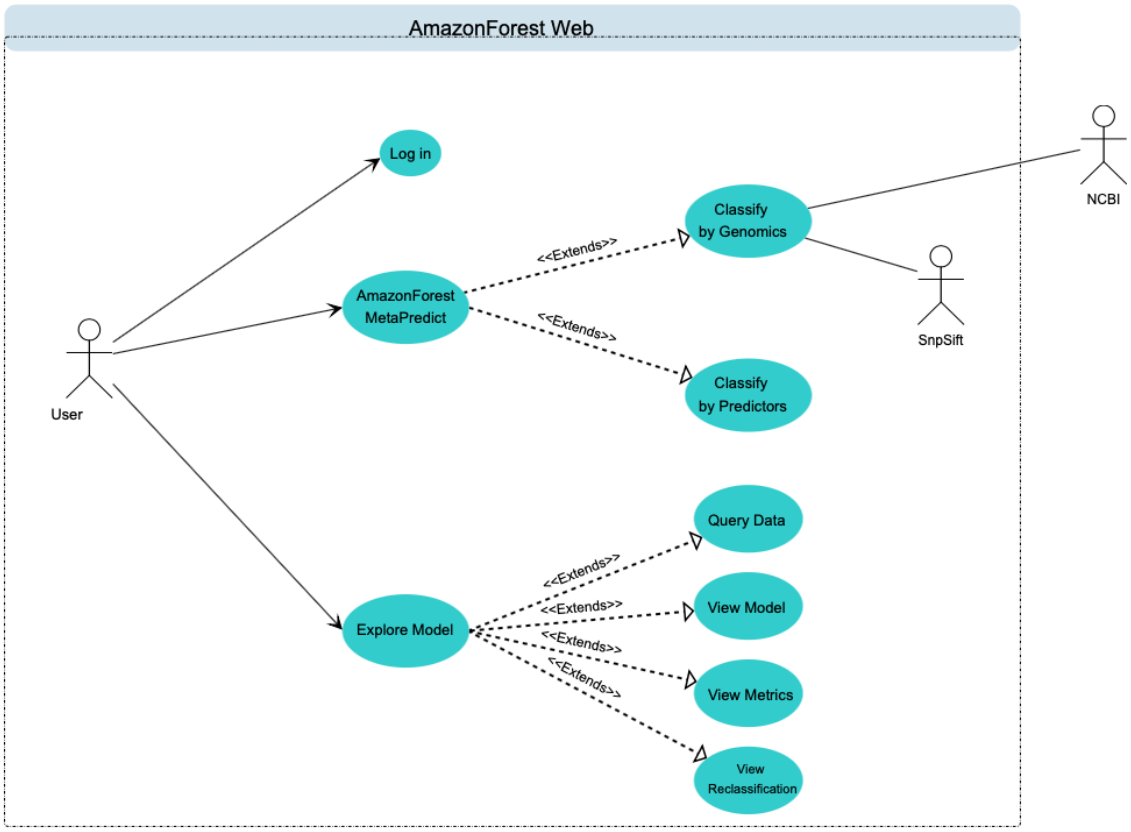

Figure 2. Use Case Diagram of AmazonForest.

2. Querying AmazonForest

The web interface allows querying variant pathogenicity based on chromosome, position, or variant identifier from dbSNP. For example, the user can set the following data: chromosome 1, position 1020355, in reference allele textbox the user can set 'G', and in alternative allele text box set 'C' to perform prediction (see figure 3). Also, AmazonForest allows prediction by inputting rsID (see Figure 4) or a combination of in silico predictors (see Figure 5).

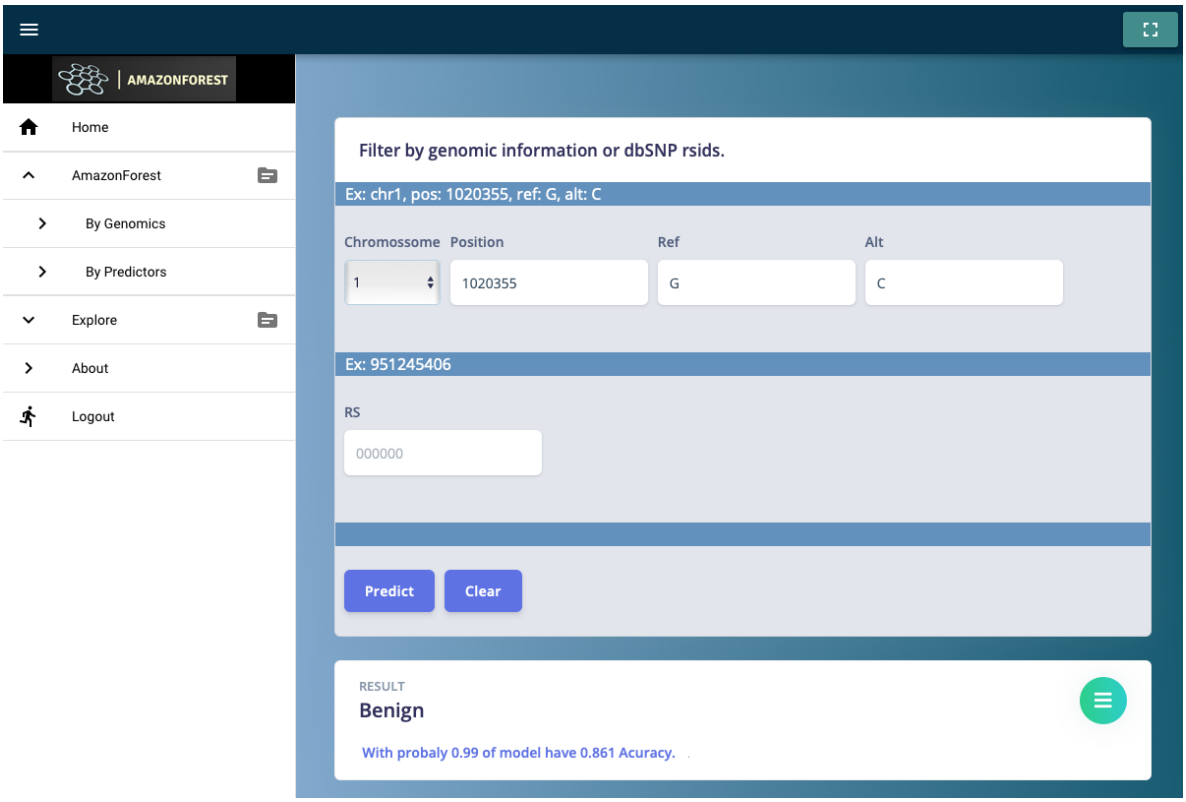

Figure 3. Example of query view on AmazonForest using genomic information.

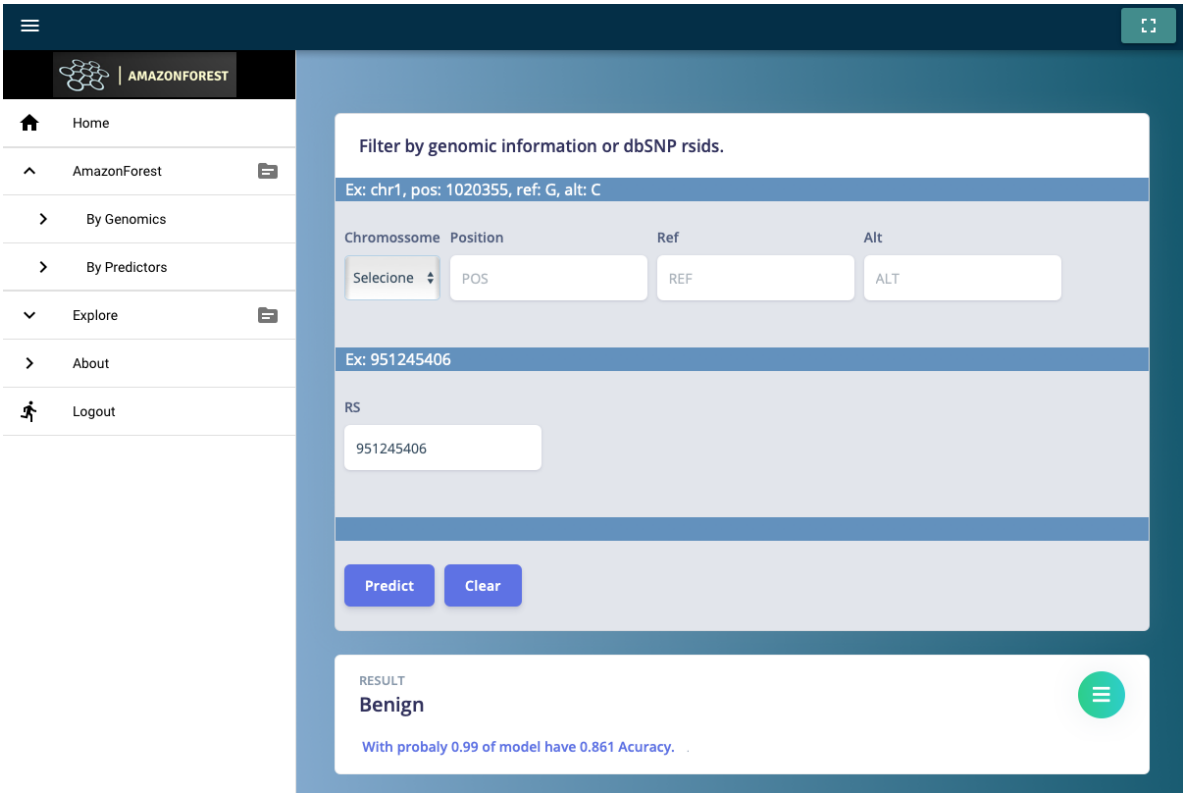

Figure 4. Example of query view on AmazonForest using rsID.

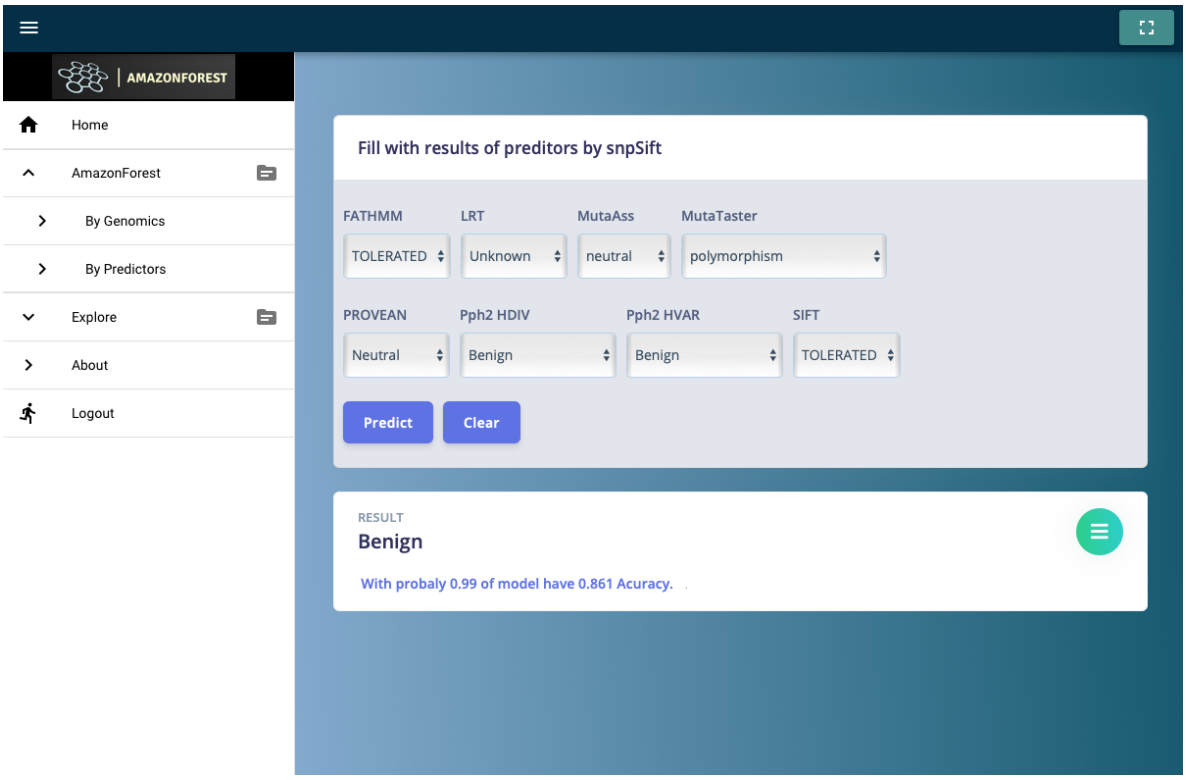

Figure 5. Example of query view on AmazonForest using in silico categorical predictors.