

Population Genetics Data Analysis in R

Muhammad Faisal^{1,*}, Andreas Futschik¹, Claus Vogl² Muhammad Asif Wazir³

1. Department of Statistics and Operations Research (ISOR), University of Vienna, Austria.
 2. Institute of Genetics and Animal Breeding, University of Veterinary Medicine Vienna, Austria.
 3. International Institute for Applied System Analysis (IIASA), Laxenburg, Austria.
- *Contact author: mfaisal_16@yahoo.com

Keywords: Population Genetics, Exact Likelihood, Hidden Markov model, Mutations.

Statistics is essential to understand the genetic variation in population genetics. Estimation of unknown parameters such as, mutation, recombination, migration, etc., from genetic data is always a challenge in population genetics. I implemented the dynamic programming approach (Faisal et al. 2011) in R that calculated the exact likelihood under infinite sites model by summing over all genealogical trees for estimating mutation rate. I will present R package **PGDA** that will estimate mutation rates either time-dependent or it is independent of time. It generates samples from *ms* program (Hudson, 2002) and converted *ms* format into *FASTA* format that would be input for the program estimating mutation rates. The presentation has been motivated from our work of exact likelihood computation method for estimation of mutations rates.

References

1. Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, 18, 337–338.
2. Faisal, M., Futschik, A., and Vogl, C. (2011). Exact Likelihood Computation of Mutation rates under Infinite Sites Model. (submitted)