**MCLink: an R package for genomic analysis using Monte Carlo techniques**

Kawan Santana, Fernanda de Oliveira Busato, Rodrigo Malavazi Corder

Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil

Funding: Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil (FAPESP; 2025/17188-0) and Conselho Nacional de Desenvolvimento Científico e Tecnológico of Brazil (CNPq; 125233/2024-8 and 143825/2025-9)

Linkage disequilibrium (LD) reflects non-random associations between genetic loci and is a key property of population genomic data. In malaria parasites such as *Plasmodium vivax*, LD patterns can provide important insights into recombination, transmission dynamics, and population structure, which are relevant for understanding parasite evolution and guiding control strategies. However, most commonly used LD approaches rely on point estimates that may not fully capture the variability present in genomic data. This project aims to develop and validate an R package for the inference of LD in haploid populations using Monte Carlo and permutation-based approaches, and to apply this framework to genomic data from *P. vivax* populations. The package reads genomic variants from VCF files and converts genotype data into a sparse binary matrix. Monte Carlo sampling is then used to repeatedly select subsets of loci, from which pairwise relative genetic distances between individuals are calculated using normalized Hamming distances. A null model of panmixia is generated by permuting genotypes within loci, preserving allele frequencies while removing non-random associations. The resulting empirical distributions of genetic distances under observed data and under the null model are compared to assess LD patterns. Analyses can be performed genome-wide or at the chromosome level. Preliminary analyses using genomic data from *P. vivax* populations in the Brazilian Amazon reproduced patterns previously reported in the literature. Initial results indicate strong agreement between distributions obtained with the developed package and those from earlier analyses. Ongoing tests are evaluating the effect of imposing a minimum physical distance between SNPs to reduce LD driven by physical proximity. This tool has potential applications in population genomic studies of malaria parasites, contributing to a deeper understanding of parasite population dynamics and evolutionary processes.

Keywords: malaria; *Plasmodium vivax*; linkage disequilibrium; Monte Carlo; R package.