



Estudo de associação genômica ampla (GWAS)

Helio de Souza Junior,*PG:55679

2025-06-01

Índice

| | |
|---|---|
| Pacote - mrMLM | 2 |
| Artigos | 2 |
| Pré-requisitos | 2 |
| Configurando seu ambiente de trabalho | 2 |
| Instalar o pacote mrMLM e suas dependências | 2 |
| Preparando Dados | 3 |
| Devem ser preparados quatro dados de entrada diferentes: | 3 |
| Análise de associação em todo o genoma usando a função mrMLM do pacote mrMLM. | 3 |
| Trait 1 | 3 |
| Trait 2 | 4 |
| Serão produzidos quatro resultados diferentes: | 4 |

*Universidade Estadual de Maringá, pg55679@uem.br

Disciplina: Genômica Aplicada ao Melhoramento de Plantas, ministrada pela Profa. Dra. Maria Celeste e Dra. Stephanie Alves.

Elaborado por [Helio de Souza Junior](#), com [RStudio](#) usando [Rmarkdown LaTeX](#) e [Quarto](#) .
Código disponível no meu [GitHub](#).

Pacote - mrMLM

Multi-Locus Random-SNP-Effect Mixed Linear Model Tools for GWAS. [Manual PDF](#).

Artigos

Methodological implementation of mixed linear models in multi-locus genome-wide association studies. <https://doi.org/10.1093/bib/bbw145>.

A compressed variance component mixed model for detecting QTNs and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies. <https://doi.org/10.1016/j.molp.2022.02.012>.

Pré-requisitos

Antes de executar os scripts de análise, certifique-se de estar usando o **R versão 4.4.3** ou **superior**. Isso é essencial para a compatibilidade com o pacote **mrMLM** e suas dependências.

Configurando seu ambiente de trabalho

Para começar, defina seu diretório de trabalho no R como o local onde você armazenou os scripts de análise e os arquivos de dados. Você pode fazer isso modificando o seguinte comando com o caminho do seu diretório específico:

Instalar o pacote mrMLM e suas dependências

Você pode usar o **mrMLM.GUI** ou a instalação baseada em comandos.

```
options(repos = c(CRAN = "https://cran.r-project.org"))  
install.packages("mrMLM.GUI", dependencies = TRUE)
```

Preparando Dados

Os dados genotípicos e fenotípicos devem ser preparados no formato correto, conforme disponível na pasta de dados de exemplo do pacote [mrMLM no CRAN](#).

Devem ser preparados quatro dados de entrada diferentes:

1. Dados genotípicos
2. Dados fenotípicos
3. Matriz de estrutura populacional (Q)
4. Matriz de parentesco (K) por enquanto não fazer

OBS: Modificar os caminhos dos arquivos de dados

```
GenotypicData <- read.csv("C:/Mapeamento por associação/Genotype_Hmp.txt")
PhenotypicData <- read.csv("C:/Mapeamento por associação/Phenotype.txt")
PopStr <- read.csv("C:/Mapeamento por associação/PopStr.txt")
```

Análise de associação em todo o genoma usando a função mrMLM do pacote mrMLM.

Trait 1

```
# Cria a pasta se ela ainda não existir
dir.create("C:/Mapeamento por associação/Resultados_GWAS", recursive = TRUE)
```

```
# Executa a análise
mrMLM(
  fileGen = "C:/Mapeamento por associação/Genotype_Hmp.txt",
  filePhe = "C:/Mapeamento por associação/Phenotype.txt",
  filePS = "C:/Mapeamento por associação/PopStr.txt",
  Genformat = "Hmp",
  method = c("ISIS EM-BLASSO", "mrMLM", "FASTmrMLM", "FASTmrEMMA", "pLARmEB", "pKWmEB"),
  Likelihood = "REML",
  trait = 1,
  SearchRadius = 20,
  CriLOD = 3,
  Bootstrap = TRUE,
```

```

DrawPlot = TRUE,
Plotformat = "jpeg",
dir = "C:/Mapeamento por associação/Resultados_GWAS",
PopStrType = "Q"
)

```

Trait 2

```

# Cria a pasta se ela ainda não existir
dir.create("C:/Mapeamento por associação/Resultados_GWAS", recursive = TRUE)

```

```

# Executa a análise
mrMLM(
  fileGen = "C:/Mapeamento por associação/Genotype_Hmp.txt",
  filePhe = "C:/Mapeamento por associação/Phenotype.txt",
  filePS = "C:/Mapeamento por associação/PopStr.txt",
  Genformat = "Hmp",
  method = c("ISIS EM-BLASSO", "mrMLM", "FASTmrMLM", "FASTmrEMMA", "pLARmEB", "pKWmEB"),
  Likelihood = "REML",
  trait = 2,
  SearchRadius = 20,
  CriLOD = 3,
  Bootstrap = TRUE,
  DrawPlot = TRUE,
  Plotformat = "jpeg",
  dir = "C:/Mapeamento por associação/Resultados_GWAS",
  PopStrType = "Q"
)

```

Serão produzidos quatro resultados diferentes:

1. Resultados finais contendo associações significativas entre marcadores e características com valores de $-\log_{10}(p)$, valores de R^2 e outros detalhes necessários.
2. Resultados intermediários.
3. O gráfico de Manhattan.
4. O gráfico QQ.