



Associação Genômica Ampla Mapeamento por Associação - GWAS

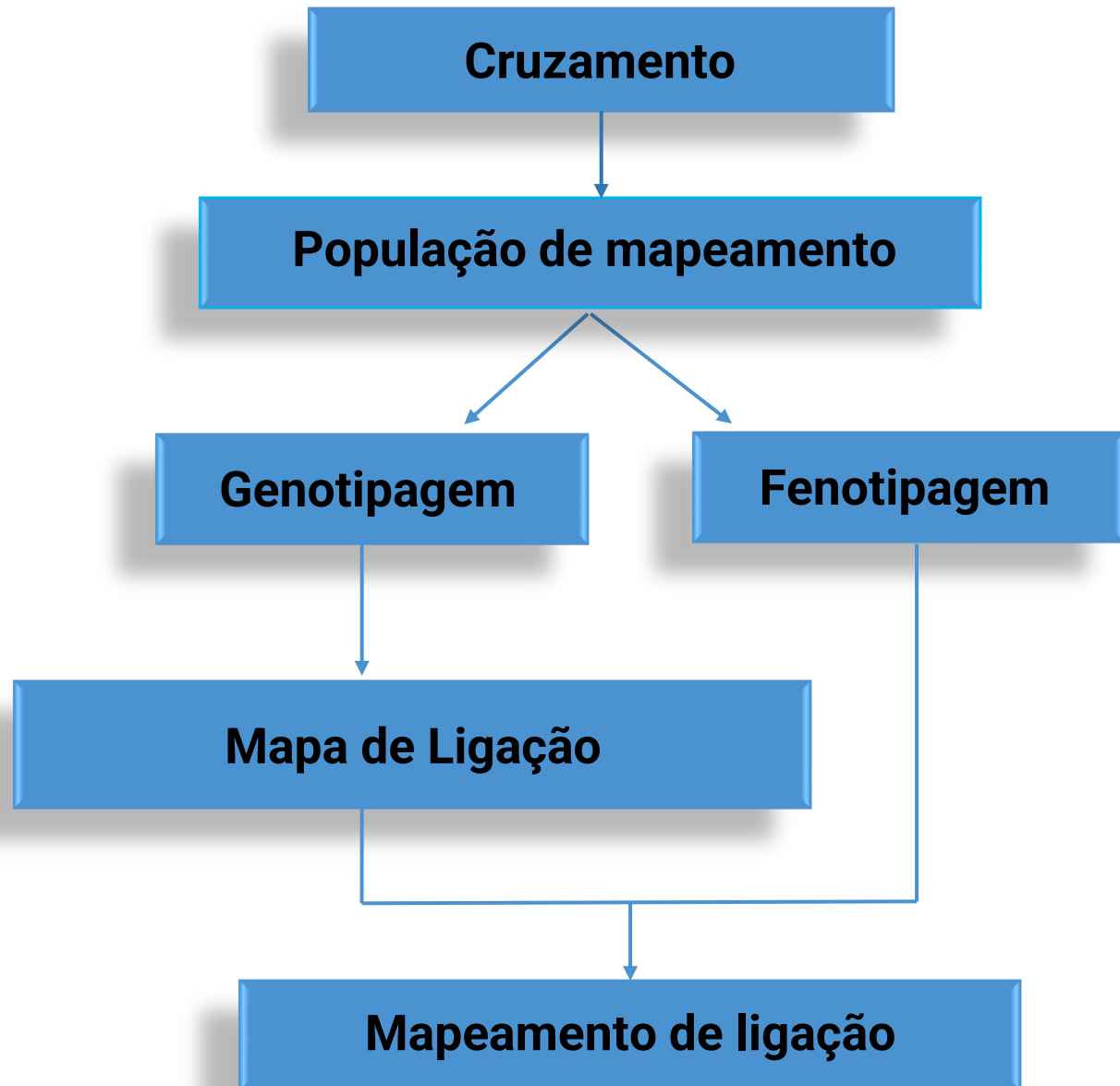
Stephanie Mariel Alves

Maringá - PR, 02/06/2025

**Como mapear genes controladores
de características quantitativas?**



Mapeamento de regiões genômicas controladoras de características quantitativas



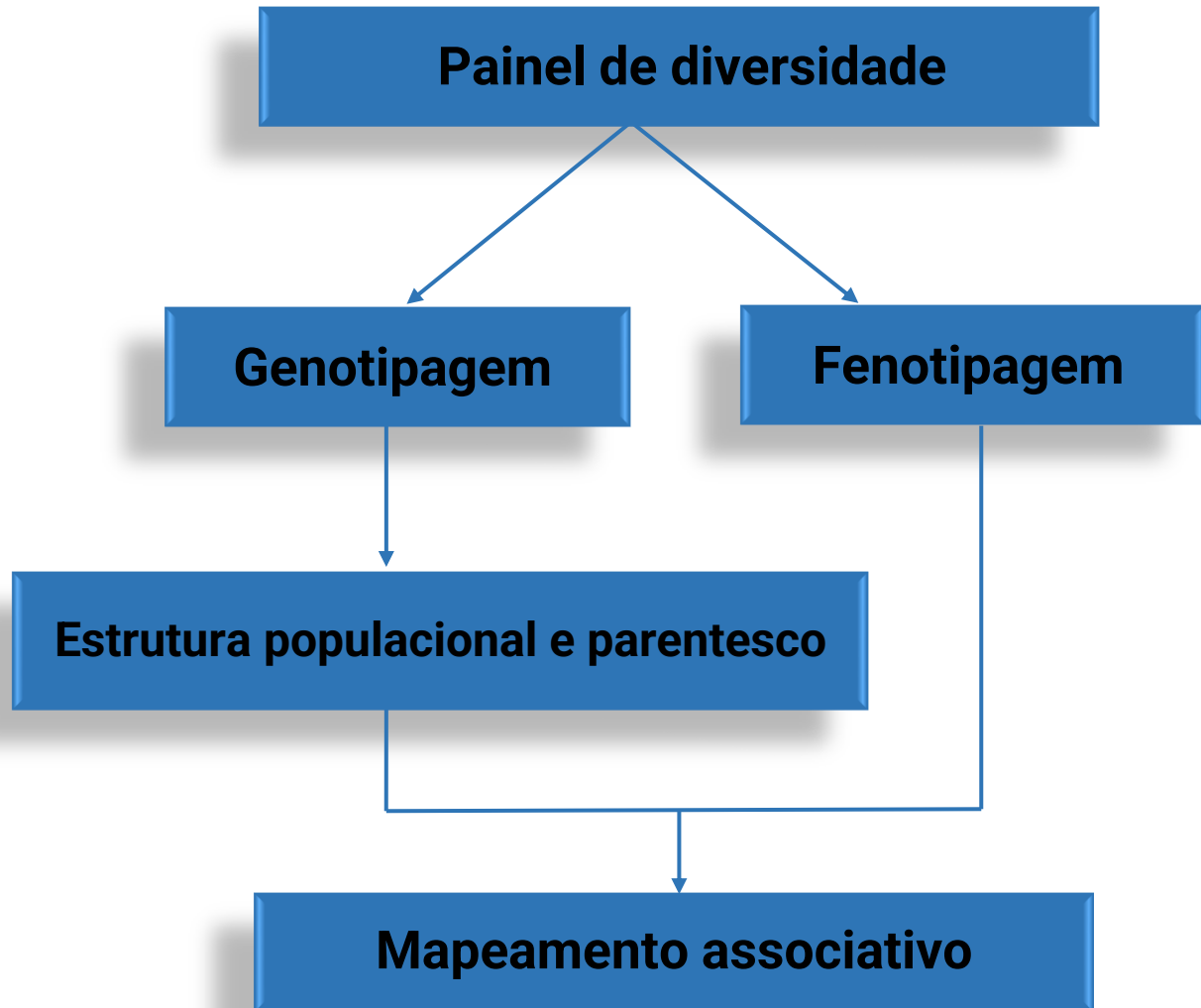
Mapeamento de Ligação

Eventos de recombinação limitados ao número de geração a qual a população foi submetida

Baseia-se na ligação fatorial

Populações controladas

Mapeamento de regiões genômicas controladoras de características quantitativas



Mapeamento Associativo

Exploração recombinações criadas ao longo da história evolutiva da espécie

Baseia-se no desequilíbrio de ligação

Utiliza-se de painéis de diversidade

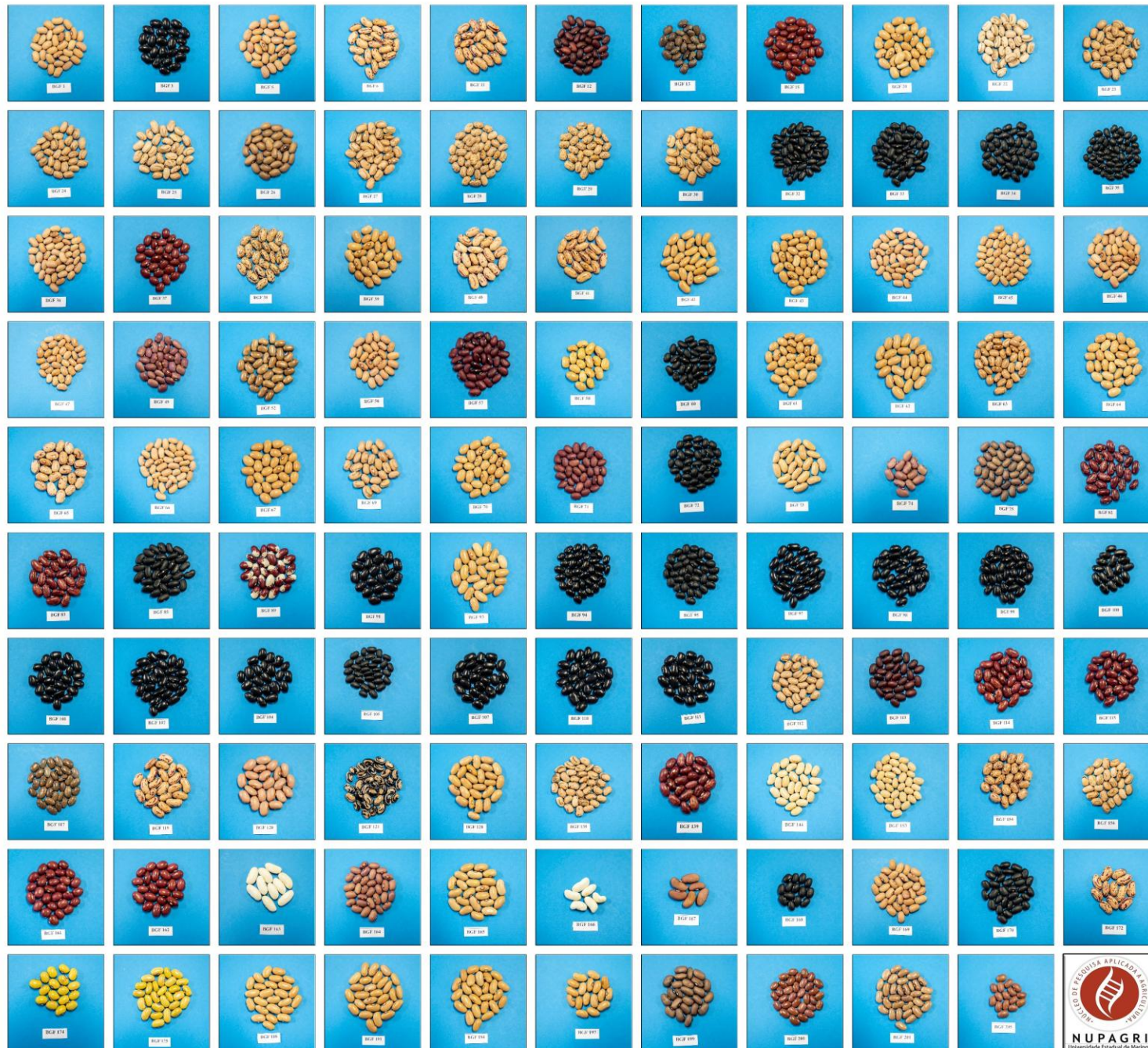
Populações de Mapeamento

Mapeamento de Ligação	Mapeamento Associativo
Populações derivadas de cruzamentos controlados	Painéis de diversidade
Base genética limitada	Base genética mais ampla
Frequências alélicas conhecidas	Frequências alélicas desconhecidas
Ausência de estrutura de população	Pode haver estrutura de população
Parentesco conhecido	Parentesco desconhecido



1

Painel de diversidade



Material Vegetal

- 109 acessos de feijão comum Mesoamericanos e Andinos - Banco de Germoplasma do Nupagri/UEM;
- 55 acessos Mesoamericanos;
- 54 acessos Andinos;
- Estados de coleta: GO, MS, MG, PR e SC.

2

Genotipagem



População de interesse



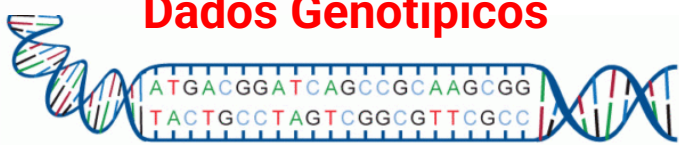
Amostra foliar para extração de DNA



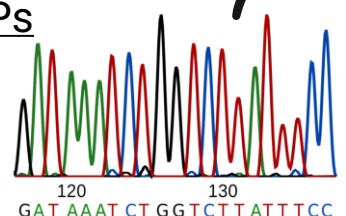
Extração de DNA



Dados Genotípicos



Sequenciamento DNA/ BeadChip contendo os marcadores SNPs



3

Fenotipagem

Dados Fenotípicos

- Produtividade kg/ha;
- Número de vagens por planta;
- Número de sementes por vagem;
- Massa de 100 grãos...



data.txt



Modelos Estatísticos

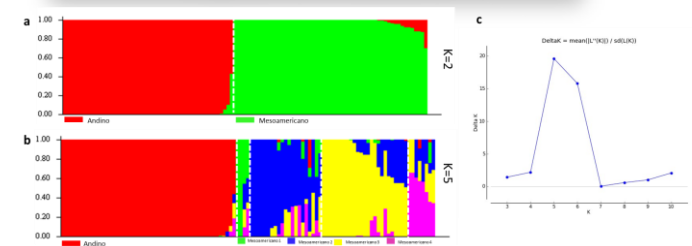
$$Y_{ijk}^* = \mu + \text{SNP}_i + \sum_{k=1} \text{PC}_{jk} + \varepsilon_{ijk}$$

5

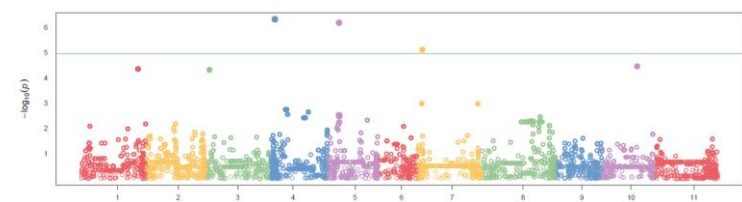
Mapeamento associativo

4

Estrutura populacional e parentesco



Matriz da estrutura populacional e parentesco incluída nas análises.



Análises Estatísticas

Fenotipagem

- Média ajustada dos anos e análise conjunta: Software SELEGEN REML/BLUP, de acordo com o modelo 52:

$$y = Xr + Zg + Wb + Ti + e$$

- A significância dos efeitos do modelo foi estimada utilizando o teste da razão de verossimilhança (LRT):

$$\text{Likelihood-Ratio Test (LRT)} = (-2\text{LogL})_{p-1} - (-2\text{LogL})_p$$

Análises Estatísticas

Genotipagem

- Análise Multilocus (mrMLM v4.0.2) ML-GWAS com 109 acessos e 4.632 SNPs;
- mrMLM, FASTmrMLM, FASTmrEMMA, ISIS EM-BLASSO e pLARmEB para identificação SNPs no software R;
- Os valores significativos para associações foram $\text{LOD} \geq 3$ para todos os métodos;
- Estrutura populacional $K = 2$, obtida por meio do software Structure v2.3.4;
- Matriz de parentesco obtida por meio do software mrMLM v4.0.2;
- Anotação funcional: Genoma de referência *Phaseolus vulgaris* (G19833 versão 1.0).



APPLICATION NOTE

mrMLM v4.0.2: An R Platform for Multi-locus Genome-wide Association Studies



Ya-Wen Zhang^{1, #}, Cox Lwaka Tamba^{2, #}, Yang-Jun Wen³, Pei Li¹
Wen-Long Ren⁴, Yuan-Li Ni³, Jun Gao⁵, Yuan-Ming Zhang^{1, *}



National Library of Medicine
National Center for Biotechnology Information

Log in

www.ncbi.nlm.nih.gov



Phytozome 13
THE PLANT GENOMICS RESOURCE

<https://phytozome-next.jgi.doe.gov/>

(Pritchard et al., 2000; Schmutz et al., 2014; Song et al., 2015; R Core Team, 2014; Wang et al., 2016; Tamba e Zhang, 2018; Wen et al., 2018; Tamba et al., 2017; Zhang et al., 2017; Zhang et al., 2020a)

Instalação R

<https://posit.co/download/rstudio-desktop/>



1: Install R

RStudio requires R 3.6.0+. Choose a version of R that matches your computer's operating system.

R is not a Posit product. By clicking on the link below to download and install R, you are leaving the Posit website. Posit disclaims any obligations and all liability with respect to R and the R website.

DOWNLOAD AND INSTALL R

2: Install RStudio

DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Size: 281.27 MB | [SHA-256: 9E6F68CA](#) | Version: 2025.05.0+496 |
Released: 2025-05-05



CRAN
[Mirrors](#)
[What's new?](#)
[Search](#)
[CRAN Team](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Task Views](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)

Donations
[Donate](#)

*Escolher sistema operacional! →

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#) ([Debian](#), [Fedora/Redhat](#), [Ubuntu](#))
- [Download R for macOS](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2025-04-11, How About a Twenty-Six) [R-4.5.0.tar.gz](#), read [what's new](#) in the latest version.
- The CRAN directory [src/base-prerelease](#) contains R alpha, beta, and rc releases as daily snapshots in time periods before a planned release.
- Between releases, the same directory [src/base-prerelease](#) contains snapshots of current patched and development versions. Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Alternatively, daily snapshots are [available here](#).
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#).

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

Supporting CRAN



R-4.5.0 for Windows

[Download R-4.5.0 for Windows](#) (86 megabytes, 64 bit)

[README on the windows binary distribution](#)
[New features in this version](#)

This build requires UCRT, which is part of Windows since Windows 10 and Windows Server 2016. On older systems, UCRT has to be installed manually from [here](#).

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) of server.

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

Note to webmasters: A stable link which will redirect to the current Windows binary release is <CRAN MIRROR>/bin/windows/base/release.html .

Last change: 2025-04-11

**FAZER DOWNLOAD
E INSTALAR!**

Instalação RStudio

<https://posit.co/download/rstudio-desktop/>

1: Install R

RStudio requires R 3.6.0+. Choose a version of R that matches your computer's operating system.

R is not a Posit product. By clicking on the link below to download and install R, you are leaving the Posit website. Posit disclaims any obligations and all liability with respect to R and the R website.

DOWNLOAD AND INSTALL R

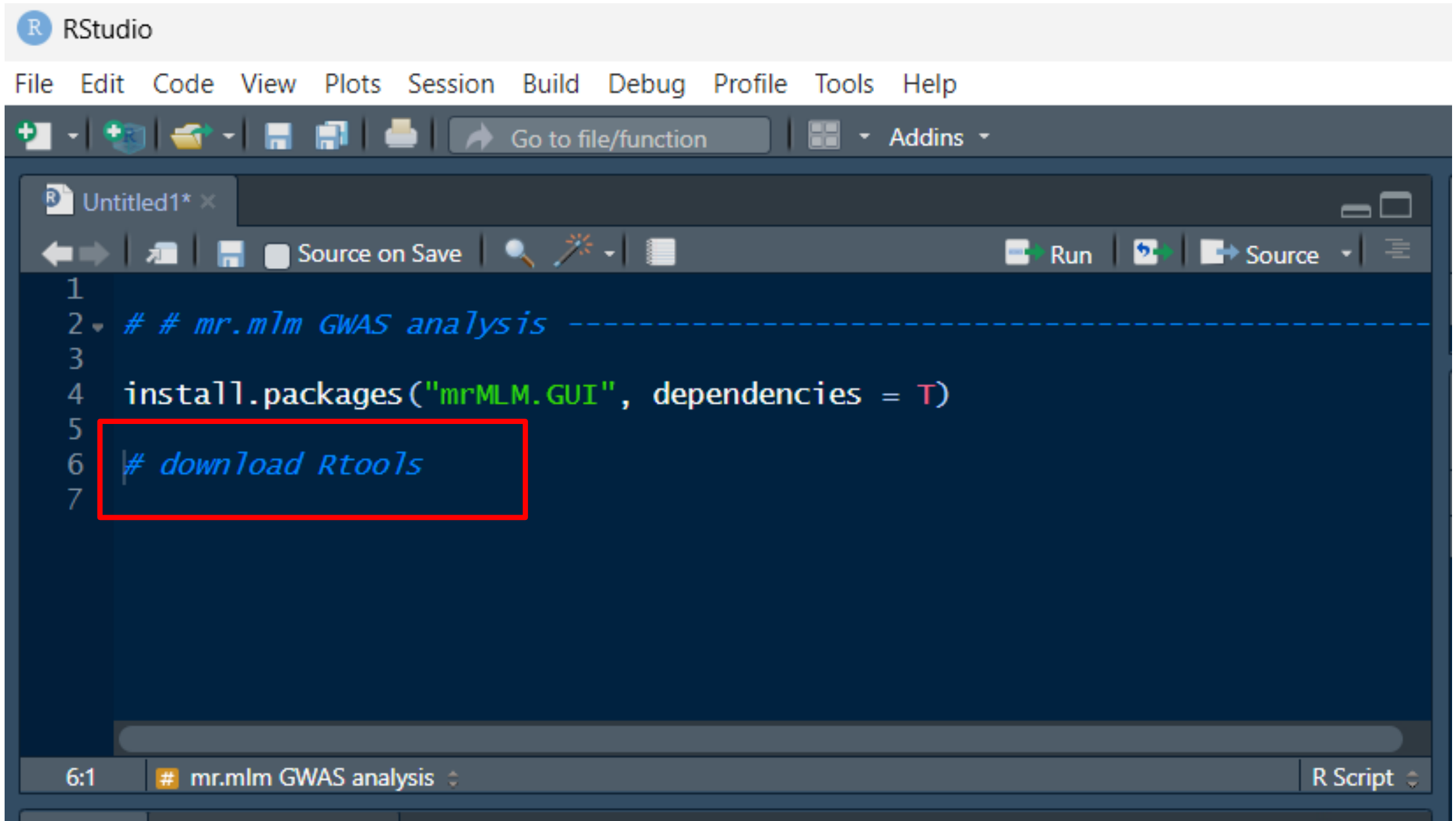


2: Install RStudio

DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Size: 281.27 MB | [SHA-256: 9E6F68CA](#) | Version: 2025.05.0+496 |
Released: 2025-05-05

**FAZER DOWNLOAD
E INSTALAR!**



The image shows the RStudio interface with a menu bar (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help) and a toolbar. The main editor window, titled 'Untitled1*', contains the following R code:

```
1  
2 # # mr.mlm GWAS analysis -----  
3  
4 install.packages("mrMLM.GUI", dependencies = T)  
5  
6 # download Rtools  
7
```

A red rectangular box highlights the comment on line 6: `# download Rtools`. The status bar at the bottom indicates the cursor is at line 6, column 1, and the file is named 'mr.mlm GWAS analysis'.

Download Disponível: <https://cran.r-project.org/bin/windows/Rtools/>

RTools: Toolchains for building R and R packages from source on Windows

Choose your version of Rtools:

RTools 4.5	for R versions from 4.5.0 (R-prerelease and R-devel)
RTools 4.4	for R versions 4.4.x (R-release)
RTools 4.3	for R versions 4.3.x (R-oldrelease)
RTools 4.2	for R versions 4.2.x
RTools 4.0	for R from version 4.0.0 to 4.1.3
old versions of RTools	for R versions prior to 4.0.0

Rtools45 for Windows

Rtools is a toolchain bundle used for building R packages from source (those that need compilation of C/C++ or Fortran code) and for building R itself. Rtools45 is currently used for R 4.5 and R-devel, become R 4.6.0.

Rtools45 consists of Msys2 build tools, GCC 14/MinGW-w64 compiler toolchain, libraries built using the toolchain, and QPDF. Rtools45 supports 64-bit Windows and UCRT as the C runtime.

Compared to Rtools44, Rtools45 for 64-bit Intel machines has newer versions of two core components: GCC and binutils. It is recommended to re-compile all code with the new toolchain to avoid problems.

Rtools45 is also available for 64-bit ARM machines (aarch64): it includes Msys2 build tools (64-bit Intel builds running via emulation) and aarch64 builds of LLVM 19/MinGW-w64 compiler toolchain, and again QPDF. The 64-bit ARM version of Rtools45 is experimental: a number of CRAN packages don't work with it and the Fortran compiler (flang-new) is not yet able to compile Fortran. A number of CRAN packages don't work because they require not-yet-available 64-bit ARM versions of external software.

Installing Rtools45

Rtools is only needed for installation of R packages from source (those that need compilation of C/C++ or Fortran code) or building R from source. R can be installed from the R binary installer and by other versions of CRAN packages, which does not require Rtools45.

Moreover, online build services are available to check and build R packages for Windows, for which again one does not need to install Rtools45 locally. The [Winbuilder](#) check service uses identical setup, performs package checks and has already all CRAN and Bioconductor packages pre-installed.

Rtools45 may be installed from the [Rtools45 installer](#) or [64-bit ARM Rtools45 installer](#). It is recommended to use the defaults, including the default installation location of `C:\rtools45`.

When using R installed by the installer, no further setup is necessary after installing Rtools45 to build R packages from source. When using the default installation location, R and Rtools45 may be installed simultaneously or Rtools45 may be installed when R is already running.

On ARM, binary versions of packages are currently not available from CRAN, so Rtools45 is required to install any package that needs compilation.

Additional information

A detailed tutorial on how to build R and packages using Rtools45 for R package authors and R developers is available for [R 4.5.x](#) and [R-devel](#).

From the user perspective, Rtools45 is the same as Rtools42-44. However, it uses newer versions of the compiler toolchain and libraries and headers. Maintainers of CRAN and Bioconductor packages may use [these patches](#) for reference or re-use them in their code.

A change log for Rtools45 vs Rtools44 and of individual revisions of Rtools45 is available [here](#).

**FAZER DOWNLOAD
E INSTALAR!**

```
1
2 ## mr.mlm GWAS analysis -----
3
4 install.packages("mrMLM.GUI", dependencies = T)
5
6 # download Rtools
7
8 library("mrMLM.GUI")
9 mrMLM.GUI()
10
```


☒ mrMLM ☐ Start

Multi-locus GWAS methods

1. Zhang YM, Mao Y, Xie C, Smith H, Luo L, Xu S*. Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). *Genetics* 2005;169:2267-2275
2. Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ, Zhang J, Jim M Dunwell, Xu S*, Zhang YM*. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific Reports* 2016;6:19444. (mrMLM)
3. Tamba CL, Ni YL, Zhang YM*. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Computational Biology* 2017;13(1):e1005357. (ISIS EM-BLASSO)
4. Zhang J, Feng JY, Ni YL, Wen YJ, Niu Y, Tamba CL, Yue C, Song QJ, Zhang YM*. pLARM EB: integration of least angle regression with empirical Bayes for multi-locus genome-wide association studies. *Heredity* 2017;118(6):517-524. (pLARM EB)
5. Ren WL, Wen YJ, Jim M Dunwell, Zhang YM*. pKWmEB: integration of Kruskal-Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity* 2018;120(3):208-218. (pKWmEB)
6. Wen YJ, Zhang H, Ni YL, Huang B, Zhang J, Feng JY, Wang SB, Jim M Dunwell, Zhang YM*, Wu R*. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings in Bioinformatics* 2018;19(4):700-712. (FASTmrEMMA)
7. Tamba CL, Zhang YM. A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv*, 2018. doi: 10.1101/341784. (FASTmrMLM)
8. Zhang Yuan-Ming, Jia Zhenyu, Jim M. Dunwell. Editorial: The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. *Frontiers in Plant Science* 2019, 10: 100.

§ Traits	Year	Chr	Position (pb)	SNP	* LOD Score	–Log ₁₀ P	‡ SNP effect	¶ R ² (%)	Methods
YDSD	2019	2	48,720,860	ss715647997	4.55	5.33	–147.00~ –124.64	12.43~ 17.29	1,5
		4	25,245,150	ss715650918	4.72	5.50	–513.05~ –361.05	9.81~ 16.55	1,2
	2020	7	39,295,293	ss715640487	5.35~ 6.10	6.16~ 6.94	–303.83~ –235.68	11.69~ 17.70	1,2,4,5
		8	55,278,106	ss715639359	3.99	4.74	148.49~ 238.74	5.73~ 12.37	1,2
	2021	5	5,399,253	ss715648672	4.63	5.41	–299.91~ –150.73	12.70~ 12.96	3,4
SW	C*	2	48,720,860	ss715647997	4.85	5.64	–203.83~ –175.25	13.38~ 18.32	1,4
		7	819,750	ss715645689	3.21~ 9.19	3.92~ 10.11	–7.55~ –2.99	13.48~ 21.51	2,3,4
	2019	10	24,598,176	ss715641543	3.46~ 4.95	4.18~ 5.74	–6.90~ –3.79	10.13~ 21.71	3,4
		2	3,118,774	ss715639502	3.41~ 9.10	4.13~ 10.01	–6.48~ –3.38	2.08~ 29.97	1,3,4
	2020	5	40,128,137	ss715646697	5.01~ 5.90	5.81~ 6.73	2.75~ 6.47	11.28~ 15.64	2,3
		7	819,750	ss715645689	5.75~ 10.15	6.57~ 11.09	–7.30~ –3.45	17.83~ 25.23	2,3,4,5
	2021	9	33,444,042	ss715645602	6.41~ 7.89	7.26~ 8.78	–3.03~ –2.03	5.30~ 11.75	1,4,5
		6	22,116,966	ss715639786	4.49~ 30.63	5.26~ 31.54	–6.21~ –3.52	18.43~ 57.29	4,5
	C*	7	819,750	ss715645689	3.10~ 3.38	3.80~ 4.10	–9.49~ –2.34	8.15~ 33.45	2,3,4
		9	33,444,042	ss715645602	5.17~ 7.42	5.98~ 8.29	–2.13~ –1.96	4.83~ 5.74	4,5
SDPD	2020	6	27,760,992	ss715645673	3.77~ 4.00	4.51~ 4.76	–0.35~ –0.32	4.92~ 8.24	1,5
PDPL	2019	1	51,819,821	ss715645301	3.22~ 15.14	3.93~ 16.16	0.80~ 1.77	10.11~ 43.39	1,2,3,4,5
		2	48,720,860	ss715647997	4.80~ 8.15	5.59~ 9.04	–0.97~ –0.88	8.83~ 10.74	1,4,5
		10	39,797,018	ss715646330	3.18~ 4.09	3.89~ 4.85	–1.03~ –0.61	2.92~ 8.13	1,2
	2021	10	39,797,018	ss715646330	3.13~ 5.64	3.83~ 6.46	–2.38~ –0.95	5.07~ 15.24	1,2,3,4,5
		11	48,780,038	ss715650748	6.69~ 6.86	7.54~ 7.72	–1.59~ –1.12	9.46~ 19.15	1,4
	C*	1	51,819,821	ss715645301	3.54~ 14.89	4.27~ 15.91	0.95~ 1.92	12.22~ 43.10	1,2,3,4,5
		2	48,720,860	ss715647997	4.56~ 6.89	5.34~ 7.76	–1.04~ –0.99	8.61~ 9.36	1,4,5

Métodos utilizados:

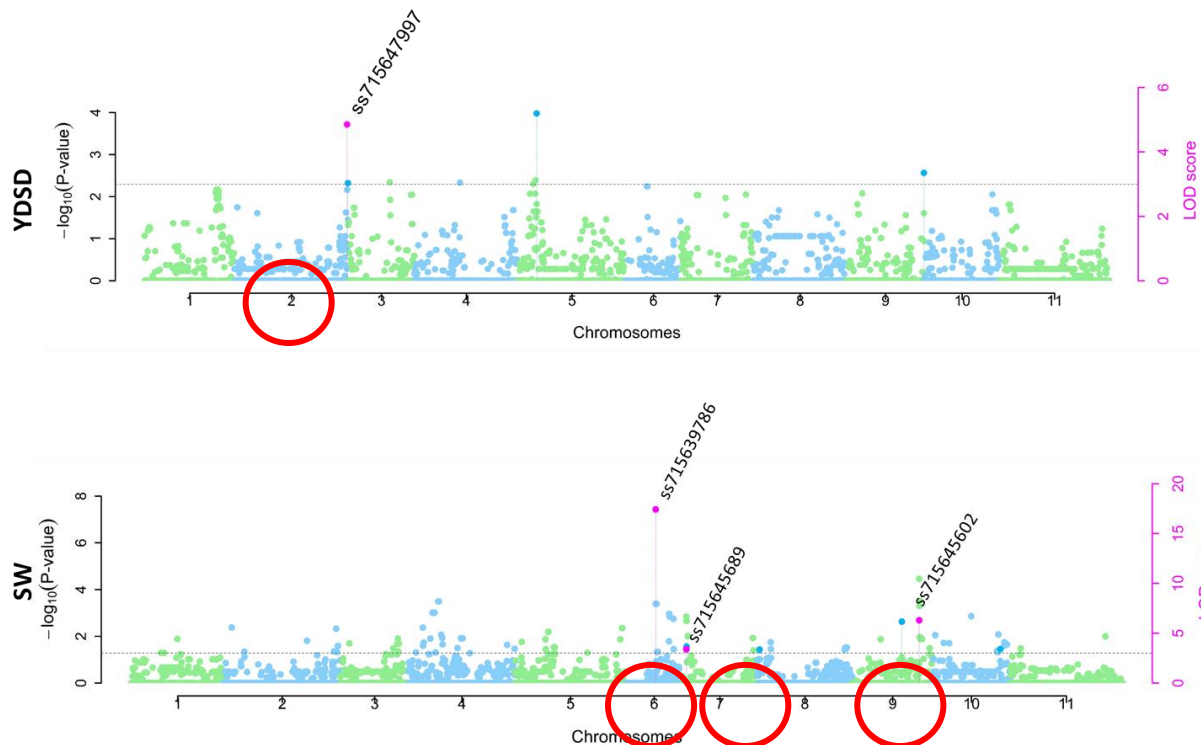
- 1, mrMLM;
- 2, FASTmrMLM;
- 3, FASTmrEMMA;
- 4, pLARmEB;
- 5, ISIS EM-BLASSO.

Pesquisar posição do SNP no NCBI – genoma de referência *Phaseolus vulgaris* (G19833 versão 1.0).

Disponível em:
https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_000499845.1/

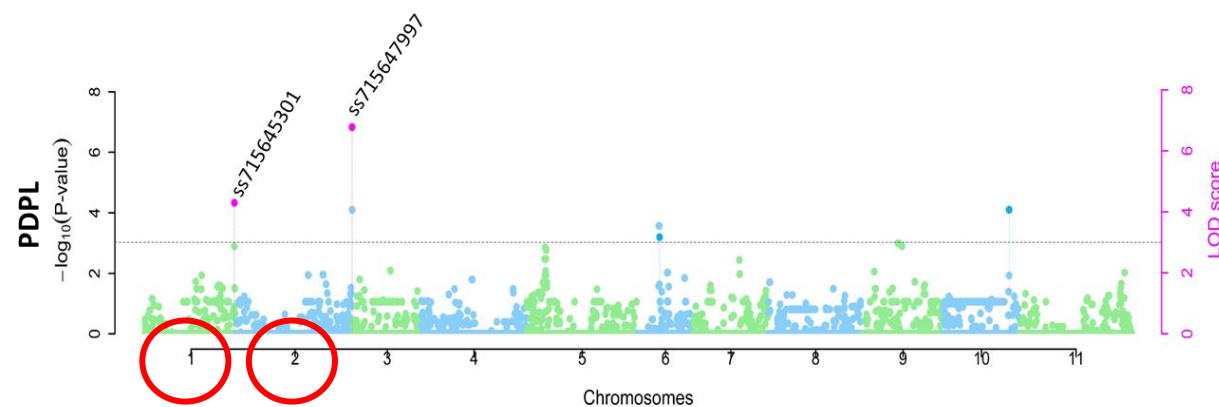
Genotipagem

SNPs para Produtividade e seus componentes: Gráfico de Manhattan plot - Análise conjunta



SNPs associados nos cromossomos: Pv01, Pv02, Pv06, Pv07 e Pv09

SNP **ss715647997** pleiotrópico no Pv02 para as características YDSD-PDPL



YDSD: Produtividade (kg ha^{-1}), **SW:** Massa de 100 grãos e **PDPL:** Número de vagens por planta.

Genes Candidatos – Genoma de referência *Phaseolus vulgaris* (G19833 versão 1.0)

YDSD	C*	2	48.720.860	ss715647997	Phvul.002G328800	serine-threonine protein kinase
SW	C*	6	22.116.966	ss715639786	Phvul.006G105000	ANKYRIN REPEAT-CONTAINING PROTEIN
		7	819.750	ss715645689	Phvul.007G011800.1	METHIONYL-TRNA SYNTHETASE
		9	33.444.042	ss715645602	Phvul.009G226000	ASPARTYL PROTEASE FAMILY PROTEIN
SDPD	2020	6	27.760.992	ss715645673	Phvul.006G166700	ABC TRANSPORTER C FAMILY MEMBER 11-RELATED
PDPL	C*	1	51.819.821	ss715645301	Phvul.001G264600	C2H2 TYPE ZINC FINGER PROTEIN
		2	48.720.860	ss715647997	Phvul.002G328800	serine-threonine protein kinase
FPIH		4	1.503.482	ss715646910	Phvul.004G014500	protein farnesyltransferase subunit beta (FNTB)
		4	1.627.690	ss715649971	Phvul.004G016000	LEUCINE-RICH REPEAT-CONTAINING PROTEIN
		6	25.791.849	ss715645752	Phvul.006G144200	SERINE-THREONINE PROTEIN KINASE
PLHT	C*	1	45.894.030	ss715647368	Phvul.001G193200	PROTEASE S28 PRO-X CARBOXYPEPTIDASE-RELATED
		1	49.657.760	ss715645852	Phvul.001G236100	POLY(A) RNA POLYMERASE GLD2
		4	1.224.240	ss715646896	Phvul.004G012300	hypothetical protein
		4	1.982.297	ss715647819	Phvul.004G019400	TRANSCRIPTION ELONGATION FACTOR B POLYPEPTIDE 3
DF	C*	1	45.894.030	ss715647368	Phvul.001G193200	PROTEASE S28 PRO-X CARBOXYPEPTIDASE-RELATED
		1	46.027.600	ss715639536	Phvul.001G194100	peroxiredoxin (alkyl hydroperoxide reductase subunit C
		7	3.770.008	ss715646355	Phvul.007G046900	16S rRNA (uracil(1498)-N(3))-methyltransferase / M(3)U(1498) specific methyltransferase
		10	7.867.881	ss715640116	Phvul.010G049800	phospholipase A1 (DAD1)
DPM	C*	1	45.746.595	ss715639271	Phvul.001G191600	beta-fructofuranosidase (E3.2.1.26, sacA)
		3	3.963.582	ss715647636	Phvul.003G038400	ACT DOMAIN-CONTAINING PROTEIN 5
		6	25.619.371	ss715648492	Phvul.006G142100	ARMADILLO/BETA-CATENIN-LIKE REPEAT-CONTAINING PROTEIN
		6	27.370.471	ss715645677	Phvul.006G162500	PPR repeat (PPR) // PPR repeat family (PPR 2) // Pentatricopeptide repeat domain (PPR 3)



Stephanie Mariel Alves

(stephaniemarielalves@gmail.com)

Doutora em Genética e Melhoramento de Plantas

Universidade Estadual de Maringá

Maringá, 87020-900, PR, Brasil

Fone: +55 44 98454 8804