

Supplementary Table S1. Different publicly available software for GWAS analyses. For each algorithm, a short description of the association method implemented and a list of advantages and drawbacks are provided.

Run mode	Software	Association method	Advantages	Drawbacks
Local	Genome-wide efficient mixed-model analysis (GEMMA); [40] [https://github.com/genetics-statistics/GEMMA]	- Linear mixed models. - Multivariate linear mixed model. - Bayesian sparse linear mixed model and variance component estimation (SNP heritability).	- Computationally fast even with very large sample sizes. - Computes exact Wald statistics and p-values. - Can perform multiple trait GWAS. - Describes the genetic architecture behind traits. - Estimates SNP heritability, genetic and environmental correlations among phenotypes. - Input data in various commonly used formats.	- Command line tool that requires some basic knowledge for installation and usage. - Bayesian sparse linear mixed model is problematic for species with a lot of LD and epistasis, for example <i>A. thaliana</i>.
	EMMAX; [66] [http://csg.sph.umich.edu/kang/emmax/download/index.html]	- Linear mixed model.	- Computationally faster than EMMA and can handle large sample sizes. - Corrects for population structure and genetic relatedness.	- Available as a command line tool.
	LIMIX; [38] [https://github.com/limix/limix]	- Linear mixed model, multi-trait model.	- Genetic and environmental factors can be modelled by different variables in the model. - Computationally efficient.	- Written in Python. - Specific format required for input data.

PLINK.; [85] [www.cog-genomics.org/plink/1.9/]

- Association tests: Cochran-Armitage trend test, Fisher's exact test, different genetic models (dominant, recessive and general), tests for stratified samples (*e.g.* Cochran-Mantel-Haenszel, Breslow-Day tests), a test for a quantitative trait; a test for differences in missing genotype rate between cases and controls; multilocus tests, using either Hotelling's T^2 statistic or a sum-statistic approach (evaluated by permutation) as well as haplotype tests, linear and logistic models.

- Available in command line (written in C/C++), R and as a GUI.
- Allows data manipulation, format conversion, data preparation for various other programs, can perform association analysis, LD calculation, distance matrix calculation, principal component analysis, imprecise fast scan for epistasis.
- Works on any type of dataset from various organisms.

- Can consume a lot of memory performing certain commands.
- Slow with large datasets.

GCTA; [65]
[<https://cns.genomics.com/software/gcta/#Overview>]

- Mixed linear model association.
- Multi-trait association.
- Meta-analysis.

- Allows data manipulation.
- Can be installed on Windows, Linux/Unix, and MacOS.
- Can be used on any species.
- Analyses the genetic architecture of complex traits.
- Variance component analysis
- LD estimation.
- Estimating the genetic relationships between individuals.
- GWAS simulation.
- Heritability estimation.

- Accepts PLINK data format only.
- Requires a large amount of memory.
- Is not very fast.
- Only command line version.

	FaST-LMM; [86] [https://github.com/fastlmm/FaST-LMM]	- Factored spectrally transformed linear mixed models.	- Faster than EMMAX and consumes less memory. - Supports tests for epistasis, corrections for cellular heterogeneity via the inclusion of principal components and heritability estimation.	- Requires Python. - Input files genotype, kinship, phenotype and covariate files in plink format.
	Multi-trait mixed model (MTMM); [33] [https://github.com/Gregor-Mendel-Institute/mtmm]	- Multi-trait linear mixed model.	- Models correlated phenotypes and considers within-trait and between-trait variance components simultaneously. - Models two traits at the same time which increases power.	- Can only analyze two phenotypes simultaneously. - Requires specific format of input data. - Works in R.
	Multi-locus mixed model (MLMM); [87] [https://github.com/Gregor-Mendel-Institute/MultLocMixMod]	- Multi-locus linear mixed model.	- Includes multiple loci in the model and can identify evidence for allelic heterogeneity (presence of other causal loci hidden by the large effect loci).	- Requires specific format of input data. - Works in R.
	TASSEL; [88] [https://sourceforge.net/projects/tassel/]	- General linear model and mixed linear model.	- Calculates and allows visualization of LD. - Calculation of principal components. - Allows other features such as: sequence alignment viewer, neighbor joining cladogram, different graphical functions and format conversion. - Accepts many different formats of input data.	- Command line and GUI interface. - GUI can be very slow with larger datasets.
	METAL; [89] [http://www.sph.umich.edu/csg/abecasis/metal/]	- Meta-analysis of genome-wide association scans.	- Improves power in complex trait gene studies. - Allows analysis of large datasets.	- Requires a specific input file format. - Command line tool that requires a bit of knowledge in script writing.
Web server	Matapax; [90] [https://matapax.mpimp-golm.mpg.de/]	- Uses GAPIT mixed model.	- User only needs to upload the phenotype file through the web browser. - Simple trait file format.	- So far only the genotype data for 1307 accessions of <i>A. thaliana</i> and a few human cancer lines available.

easyGWAS; [91]
[<https://easygwas.ethz.ch/>]

- Wilcoxon rank-sum test for homozygous genotype data; a linear regression and logistic regression for homozygous and heterozygous genotype data, linear mixed models EMMAX or FaST-LMM.
- Meta-analysis with Fisher's method, Stouffer's Z, Stouffer's weighted Z, a fixed-effect model to combine effect estimates, a random effect model.

- Various transformation methods for the phenotype data available.
- Option to make you data public or private.
- Fast computation.
- Publicly available data for many important species such as *A. thaliana*, *D. melanogaster*, *Brachypodium distachyon*.
- Can compare GWA analyses.
- Possible to perform meta-analysis.
- Only needs a web browser.

- Input phenotype and genotype data only in PLINK format.
- Must have a Dropbox account to upload a zip file of the genotype and phenotypes data.

GWA-Portal; [92]
[<https://gwas.gmi.oeaw.ac.at/>]

- Accelerated mixed model, Kruskal-Wallis test and linear regression.

- User friendly web interface.
- Only need to prepare and upload the phenotype file and run the analysis.
- An option to make your data private or public.
- Various transformation methods available for the phenotype data.
- Calculates the narrow-sense SNP heritability.
- Computationally fast.
- Interactive QQ plot and results visualization, gene annotation by TAIR, LD heatmap visualization and geographic distribution.
- Publicly available results of previous studies.
- Meta-analyses can be performed

- Only for *A. thaliana*.
- Requires a specific format of the phenotype file, ISA-Table or a CSV file format.
- Population structure accounted for only in the accelerated mixed model.

araGWAB; [93]
[<https://www.inetbio.org/aragwab/>]

- Integrates GWAS SNP p -values with functional network information.
- Within a user-defined window of distance around the significant peak in GWAS araGWAB assigns new p -values.
- Can discover other genes involved in the phenotype since one significant gene is involved in different functional gene networks.

- Analysis done within an hour.
- Only need a web browser.
- Can overcome the strict p -value threshold imposed in the GWA analysis.

- Need to upload data in a specific format.
- Only for *A. thaliana*.

GWAPP; [94]
[<https://gwapp.gmi.oeaw.ac.at/index.html#!homePage>]

- Accelerated mixed model, Kruskal-Wallis test and linear regression

- User friendly web interface.
- Only need to prepare and upload the phenotype file and run the analysis.
- An option to make your data private or public.
- Various transformation methods available for the phenotype data.
- Calculates the narrow-sense SNP heritability.
- Computationally fast.
- Interactive QQ plot and results visualization, gene annotation by TAIR, LD heatmap visualization and geographic distribution.
- Can run a conditional analysis on selected SNPs.

- Only for *A. thaliana*.
- Requires a specific format of the phenotype file, ISA-Table or a CSV file format.
- Genotype data not present for all accessions as in GWA-Portal.

GWASPro; [95] [https://www.zhaolab.org/GWASPRO/home]	- LMM and computes Wald test for p -values.	- All the computation run remotely and can be accessed via web browser. - Designed to handle large and complex datasets. - Automatically calculates the kinship matrix. - Returns original p-values, adjusted p-values based on genomic control, QQ plot and Manhattan plots.	- Need to input the genotype and phenotype files.
GWAS catalog; [96] [https://www.ebi.ac.uk/gwas/]	- A collection of GWAS studies in humans.	- One can submit results of their GWAS in humans. - Access to many GWAS studies done already in humans.	- GWAS analysis can't be performed on the website.
AraPheno; [97,98] [https://arapheno.1001genomes.org/]	- A public database of <i>A. thaliana</i> phenotypes and RNA-seq data.	- Allows meta-analyses. - Browse through published studies and retrieve information on different phenotypes. - One can submit and download different <i>A. thaliana</i> phenotypes.	- Only available for <i>A. thaliana</i>. - Can't run GWAS analysis.
