

1 Tools

We have selected four GWAS software tools to be integrated in our multiGWAS tool, two designed specifically for polyploid species as many important crops are polyploids: GWASpoly [10] and SHEsis [12], and another two designed for diploids species and extensively used in humans and plants: PLINK [8, 4] and TASSEL [3], respectively.

As MultiGWAS implements two types of GWAS analysis, naive and full, each tool is called in two different ways. The naive without any additional parameter, but the full with two parameters that take into account for population structure (Q) and relatedness (K) to prevent false associations.

1.1 GWASpoly

GWASpoly is a recent R package designed for GWAS in polyploid species that has been used in several studies in plants [2, 5, 11, 13]. It is based on the Q+K linear mixed model with biallelic SNPs that accounts for population structure and relatedness. In addition, to calculate the SNP effect for each genotypic class, GWASpoly provides a general gene action model along with four additional models: additive, simplex dominant, and duplex dominant.

MultiGWAS is using GWASpoly version 1.3. The population structure and relatedness, used in the full model, are estimated using the first five principal components and the kinship matrix, respectively, both calculated with the algorithms built in GWASpoly. For both, naive and full models, all gene action models are tested for detecting associations.

1.2 SHEsis

SHEsis is another program designed for polyploid species that includes single locus association analysis, among others. It is based on a linear regression model, and it has been used in some studies of animals and humans [9, 6].

MultiGWAS is using the version 1.0 which does not take account for population structure or relatedness, however MultiGWAS externally estimates relatedness for SHEsis by excluding individuals with cryptic first-degree relatedness using the algorithm implemented in PLINK 2.0 (see below).

1.3 PLINK

PLINK is one of the most extensively used programs for GWAS in diploids species. It was developed for humans but it is applicable to any species [7]. PLINK includes a range of analysis, including univariate GWAS using two-sample tests and linear regression models.

MultiGWAS is using two versions of PLINK: 1.9 and 2.0. Linear regression from PLINK 1.9 is used to achieve both types of analysis, naive and full. For the full analysis, population structure is estimated using the first five principal

components calculated with the PLINK 1.9 built in algorithm. But relatedness is estimated from the kinship coefficients calculated with the PLINK 2.0 built in algorithm, removing the close relatives or individuals with first-degree relatedness.

1.4 TASSEL

TASSEL is another common GWAS program based on the Java software. It was developed for maize and it has been used in several studies in plants [1, 14], but like PLINK, it is applicable to any species. For association analysis, TASSEL includes the general linear model (GLM) and mixed linear model (MLM) that accounts for population structure and relatedness.

MultiGWAS is using TASSEL 5.0, with naive GWAS achieved by the GLM, and full GWAS achieved by the MLM with two parameters: one for population structure, using the first five principal components, and another for relatedness, using the kinship matrix with centered IBS method, both calculated with built in the TASSEL built in algorithms.

References

- [1] María F. Álvarez, Myrian Angarita, María C. Delgado, Celsa García, José Jiménez-Gomez, Christiane Gebhardt, and Teresa Mosquera. Identification of Novel Associations of Candidate Genes with Resistance to Late Blight in *Solanum tuberosum* Group Phureja. *Frontiers in Plant Science*, 8:1040, 2017.
- [2] Jhon Berdugo-Cely, Raúl Iván Valbuena, Erika Sánchez-Betancourt, Luz Stella Barrero, and Roxana Yockteng. Genetic diversity and association mapping in the colombian central collection of *solanum tuberosum* L. Andigenum group using SNPs markers. *PLoS ONE*, 12(3), 2017.
- [3] Peter J Bradbury, Zhiwu Zhang, Dallas E Kroon, Terry M Casstevens, Yogesh Ramdoss, and Edward S Buckler. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19):2633–2635, 2007.
- [4] Christopher C. Chang, Carson C. Chow, Laurent C.A.M. Tellier, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1):1–16, 2015.
- [5] Luís Felipe V. Ferrão, Juliana Benevenuto, Ivone de Bem Oliveira, Catherine Cellon, James Olmstead, Matias Kirst, Marcio F. R. Resende, and Patricio Munoz. Insights Into the Genetic Basis of Blueberry Fruit-Related Traits Using Diploid and Polyploid Models in a GWAS Context. *Frontiers in Ecology and Evolution*, 6:107, 2018.

- [6] Jie Meng, Kai Song, Chunyan Li, Sheng Liu, Ruihui Shi, Busu Li, Ting Wang, Ao Li, Huayong Que, Li Li, and Guofan Zhang. Genome-wide association analysis of nutrient traits in the oyster *Crassostrea gigas*: Genetic effect and interaction network. *BMC Genomics*, 20(1):1–14, 2019.
- [7] Robert A. Power, Julian Parkhill, and Tulio De Oliveira. Microbial genome-wide association studies: lessons from human GWAS. *Nature Reviews Genetics*, 18(1):41–50, 2016.
- [8] Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A.R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I.W. De Bakker, Mark J. Daly, and Pak C. Sham. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81(3):559–575, 2007.
- [9] Hui Ping Qiao, Chun Yang Zhang, Zhi Long Yu, Qi Min Li, Yang Jiao, and Jian Ping Cao. Genetic variants identified by GWAS was associated with colorectal cancer in the Han Chinese population. *Journal of Cancer Research and Therapeutics*, 11(2):468–470, 2015.
- [10] Umesh R. Rosyara, Walter S. De Jong, David S. Douches, and Jeffrey B. Endelman. Software for Genome-Wide Association Studies in Autopolyploids and Its Application to Potato. *The Plant Genome*, 9(2):1–10, 2016.
- [11] Sanjeev Kumar Sharma, Katrin MacKenzie, Karen McLean, Finlay Dale, Steve Daniels, and Glenn J. Bryan. Linkage disequilibrium and evaluation of genome-wide association mapping models in tetraploid potato. *G3: Genes, Genomes, Genetics*, 8(10):3185–3202, 2018.
- [12] Yong Yong Shi and Lin He. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci (Cell Research (2005) 15, (97-98) DOI: 10.1038/sj.cr.7290272). *Cell Research*, 16(10):851, 2006.
- [13] Jiazheng Yuan, Benoît Bizimungu, David De Koeber, Umesh Rosyara, Zixiang Wen, and Martin Lagüe. Genome-Wide Association Study of Resistance to Potato Common Scab. *Potato Research*, 2019.
- [14] Shengkui Zhang, Xin Chen, Cheng Lu, Jianqiu Ye, Meiling Zou, Kundian Lu, Subin Feng, Jinli Pei, Chen Liu, Xincheng Zhou, Ping’an Ma, Zhaogui Li, Cuijuan Liu, Qi Liao, Zhiqiang Xia, and Wenquan Wang. Genome-wide association studies of 11 agronomic traits in cassava (*Manihot esculenta* crantz). *Frontiers in Plant Science*, 9(April):1–15, 2018.