S. Uprety (2023) Int. J. Appl. Sci. Biotechnol. Vol 11(2): 54-59.
DOI: 10.3126/ijasbt.v11i2.56114

ISSN (Online): 2091-2609
DOI Prefix: 10.3126/ijasbt

International Journal of Applied Sciences and Biotechnology (IJASBT)

Mini Review

A Prospect for Genome Wide Association Studies in Crops

Shreesha Uprety*

Institute of Agriculture and Animal Science, Kirtipur, Nepal
Permanent address: Tinpaini, Biratnagar, Province no 1, Nepal

Article Information

Received: 24 March 2023
Revised version received: 13 June 2023
Accepted: 17 June 2023
Published: 29 June 2023

Cite this article as:
S. Uprety (2023) Int. J. Appl. Sci. Biotechnol. Vol 11(2): 54-59. DOI: 10.3126/ijasbt.v11i2.56114

***Corresponding author**
Shreesha Uprety,
Institute of Agriculture and Animal Science, Kirtipur, Nepal
Permanent address: Tinpaini, Biratnagar, Province no 1, Nepal
Email: suprety.su@gmail.com

Peer reviewed under authority of IJASBT

Abstract

Genome-wide association studies (GWAS) test hundreds of thousands of genetic variants across many genomes to find those statistically associated with a specific trait or disease. A Comprehensive review of studies on biotic resistance, abiotic tolerance, yield-associated characteristics, and metabolic composition is provided. Genome-wide SNP maps have characterized population structure concerning the geographic origin and morphological type and identified patterns of ancient crop diffusion to diverse agroclimatic regions. To better understand the genomic patterns of crop diversification, nucleotide diversity variation, linkage disequilibrium, and recombination rates across the genome are quantified. Results from GWAS can be used for a range of applications. The GWAS approach has proven highly suitable for identifying common SNP-based variants with moderate to significant effects on phenotype. However, the genetic factors underlying some of these associations have been characterized. The vast majority remain unexplained. The development of next-generation sequencing and bioinformatics tools has dramatically improved and is currently being implemented to decipher the genetic diversity of targeted traits. The significant drawbacks are the need for large population size, the cost of preparing DNA samples, and less knowledge about the risk of the quality. To overcome this drawback, researchers have upgraded the statistical approaches, proper imputation of genotypes, and advanced approaches such as nested association mapping and candidate gene association mapping. The primary benefit is one-time genotyping and repeated phenotyping in different environmental conditions.

Keywords: Genome Wide Association Studies; Linkage Disequillibrium; Milestones; designs; Limitations.

Introduction

Genome-wide association studies (GWAS) help scientists identify genes associated with a particular disease (or another trait). This method studies the entire DNA set (genome) of large groups and looks for small variations, called single nucleotide polymorphisms or SNPs. A sample size with sufficient statistical power is critical to the success of genetic association studies. Genome-wide association studies require much larger sample sizes to achieve adequate statistical power.

Finding the genetic basis of complex plant traits such as flowering time, growth rate, and yield has been a significant goal in improving crops and understanding plant adaptation.

But a limitation of the ability of GWAS to identify individual genes is the occurrence of false positive results, which are an artifact of population structure. The worldwide population of natural accessions of *A. thaliana*is highly structured, and strong confounding can occur when phenotypic variation in a trait of interest overlaps with patterns of population structure. Statistical methods

developed to control population structure produce distributions of P-values that approximate the uniform distribution, although they may have reduced sensitivity. Nevertheless, GWAS in *A. thaliana*has been shown to have significant power in detecting previously known candidate genes and has also seen hundreds of loci involved in natural variation in complex traits. This new knowledge of the number of genes underlying adaptive traits and the magnitude of their effects allows us to understand better the basis of flowering time, growth rate, and yield.

A. *A. thaliana* has long been an attractive model for the study of natural variation and adaptation due to its wide distribution, the diversity of its habitats and the unsurpassed genomic resources available to this species. GWAS requires a genomic map in which the marker density is higher than the Linkage Disequillibrium (LD) range. This in turn depends on the population sample, specifically on the standing genetic diversity and the number of recombination events that shuffle that diversity.
B. Maize and rice, two of the world's most important crop species, have been the focus of intense efforts to map the ancestral genetic variation underlying agronomic traits such as grain yield, disease resistance, and plant architecture. Maize is an outcrossing plant with LDs) that break on average at approximately 2,000 bp (a distance 5 times shorter than that of *A. thaliana* (Remington *et al*., 2001). It also has a large genome (Schnable *et al*., 2009) 2.3 Gb of unique sequence and thus typing of many SNPs is needed to define a haplotype map for maize. A set of 1.6 million SNPs was designed for maize GWAS, but dense genotyping of large numbers of lines was initially cost prohibitive.
C. Rice is a self-pollinated species and, like *A. thaliana*, is a good candidate for GWAS. (Huang *et al*., 2010) identified an unbiased set of common SNPs that they used to identify strong associations between genetic loci and 14 agronomic traits, including heading date, grain size and starch quality. Here, a step forward was the use of a strategy based on second-generation sequencing technology to develop a haplotype map for 517 Chinese landraces across the rice subspecies *Oryza indica* and *Oryza japonica*. GWAS was subsequently performed using 671,355 SNPs in a subset of 373 indica lines to avoid significant confounding of population structure between subspecies (Brachi *et al*., 2011). This identified 1 to 7 loci for each agronomic trait, each explaining 6% to 68% of the variation in that trait. Several genes have also been identified in a recent GWAS of rice that has a major influence on the control of features involved in determining yield, morphology, stress tolerance, and nutritional quality (Famoso *et al*., 2011). These studies create a research platform linking genomic variation and germplasm collections to enable molecular breeding.

Checking for population structure is standard practice in GWAS, although when traits are heavily confounded, this reduces the power of the analysis and can lead to false negatives. This problem is likely to arise particularly when studying traits such as flowering time and cold tolerance that are filtered by environmental gradients that overlap with patterns of population structure. In this case, control of population structure can reduce association signals around major adaptive genes (Bergelson and Roux, 2010). In this situation, the only solution is synthetic, i.e., the restructuring of populations by crossing. Another weakness of GWAS is its lack of ability to detect rare alleles that contribute to natural variation.

Some of the confounding effects of population structure in GWAS can be avoided by adjusting the sampling strategy. Characterizing population structure before performing GWAS, together with knowledge of the ecological factors that impose selection, will help resolve some of the pitfalls of GWAS and enable the separation of adaptive variation from structured background variation

Milestones

Developed in the context of human disease genetics in mid-1990

1st GWAS publication in 2002 (Ozaki *et al*., 2002)

1st GWAS publication in plants in 2005 (Aranzana *et al*., 2005)

The first prospects for genome-wide association studies began in early 2002. This Linkage Disequillibrium-based association mapping began in humans and later in Arabidopsis, rice, grapevine, wheat, soybean, maize, tomato and other model organisms. The multi-country HapMap effort to identify a catalog of common human genetic variants represents a milestone to extend the application to other organisms to make GWAS powerful. SNPs need to be selected wisely and distributed to reflect the genetic variation. Selection of suitable and desirable markers improving strength in association signals. However, this does not necessarily mean an increased ability to detect association loci. Other disadvantages include the need for large populations, the pooling and cost of DNA sample preparation, and less knowledge about the risk of a particular trait.

Concepts Underlying Study Design

Single Nucleotide Polymorphism

This is the modern unit of genetic variation. SNPs are single base pair changes in the DNA sequence that occur with high frequency in the human genome. For genetic studies, SNPs

are typically used as markers of a genomic region, with the vast majority having minimal impact on biological systems. SNPs have functional consequences but cause amino acid changes, mRNA transcript stability, and transcription factor binding affinity changes. In particular, SNPs are a type of common genetic variation; many SNPs are present in large parts of human populations. SNPs typically have two alleles, meaning there are two commonly occurring base pair options for the SNP location in the population.

Linkage Disequilibrium

Linkage disequilibrium is a property of SNPs on a contiguous stretch of genomic sequence that describes the degree to which an allele of one SNP is inherited or correlated with an allele of another SNP in a population. The term was coined by population geneticists in an attempt to mathematically describe changes in genetic variation in a population over time. This is related to the concept of chromosomal linkage, where two markers on a chromosome remain physically linked to the chromosome for generations of a family. Recombination events in a family separate chromosomal segment from era to generation. This effect is amplified over generations and in a fixed size undergoing random mating. The rate of LD decay depends on many factors, including population size, the number of founder chromosomes in the population, and the number of generations the population has existed. As such, different human subpopulations have various degrees and patterns of LD.

Inbreeding, small population size, genetic isolation between lines, population partitioning, low recombination, rate, population admixture, natural and artificial selection, balanced selection, etc.

Factors that lead to the reduction/disruption of LD include outcrossing, high recombination rate, high mutation rate, etc.

Advantages of GWAS

1) The biological pathway of the trait may not be known.

2) The potential to discover new candidate genes that were not identified by other methodological approaches.

3) Encourage the formation of collaborative consortia in order to obtain a sufficient number of participants for the analysis who tend to continue further collaboration with the subsequent analysis.

Limitations of GWAS

Despite the tremendous success of the method, as evidenced by the wealth of information described above, GWAS currently have clear limitations, mainly issues related to population structure and low-frequency causal alleles leading to false-negative results. For example, since flowering time is a typical adaptive trait and is always confounded (i.e., highly correlated) with population structure, only one gene (ZmCCT) was detected for flowering time using a diverse association mapping panel consisting of 500 inbred lines (Liu and Yan, 2019). It is widely recognized that many false negatives occur for such confounded traits when correcting for population structure in GWAS. Another example is the demonstration that only five inbred lines in population 527 have functionally alternative alleles at the Brachytic2 locus for plant height, making identification of this locus impossible by routine association mapping analysis. Similarly, in rice, causal alleles at most of the cloned quantitative yield-related trait loci (QTL) are at low frequency in different germplasms 2% for Ghd7]; 2% for qGL3, 6% for TGW6; Two approaches have been proposed to address these issues, either the development of new statistical methods to investigate rare functional alleles or the use of artificially designed populations to equalize allelic frequencies and thereby control population structure (Alseekh *et al*., 2021).

In addition to the above issues, sometimes non-causative loci show more significant associations in GWAS than causative ones, meaning that causative genes may be distant from GWAS peaks. Such occurrence has been reported in several plant studies, including Arabidopsis, sorghum and tomato. Such misleading associations are sometimes known as synthetic associations and are thought to be due to linkage disequilibrium caused by linkage disequilibrium between commonly labeled markers and rare causative variants. This, in turn, may explain the so-called missing heritability problem of GWAS. This means that some causes do not follow the assumption of rare alleles, but rather a trait variation appears to be caused by multiple alleles within a single gene. Since mutation is constantly generates new variants, multiple independent alleles within a single gene leading to the same phenotype may be shared. As mentioned above, haplotype or gene methodologies have a high potential for identifying such situations. The aforementioned haplotype-based association mapping remains imperfect and is particularly challenging in plants. Thus, improvements in haplotype analyses will prove very beneficial for understanding the underlying genetics and its functional physiological implications.

Still why GWAS is popular?

Declining genotyping costs are likely to drive association studies away from candidate genes. It involves whole-genome resequencing of all individuals in a population, allowing assessment of point mutations, insertional deletions, and large structural variations such as copy number variation, e.g. Sequencing *Arabidobsis lyrata*. In the future, this will help to include RNA-sea data in e-QTL mapping in GWAS studies. Population selection for GWAS studies will no longer be limited to model organisms, slowly focusing more on spp that are more relevant to answering biological questions. The accuracy of GWAS depends on

one-time genotyping and repeated phenotyping under different environmental conditions.

To ensure the greatest utility of GWAS results in the future, all phenotype and genotype data should be published and stored in public databases.

Current and Future Perspective of GWAS

The power of genome-wide association studies has successfully identified many loci associated with phenotypic expression and metabolic traits in many species. However, the genetic factors underlying some of these associations have been characterized. The vast majority remains unexplained. The development of next-generation sequencing and bioinformatics tools has dramatically improved and is currently being implemented to decipher the genetic diversity of targeted traits. Multi-omic data analysis has recently supported this to improve our understanding of phenotypic diversity and its corresponding genetic basis. Combined phenotypic and transcriptomic data have been used to analyze the genetic basis of various metabolic and phenotypic traits. In addition, the development of molecular biology techniques (e.g., CRISPR/Cas9, overexpression or genetic complementation) has greatly accelerated the biological functions of causative genes behind GWAS interventions. Currently, cross-validation by association combining and linkage mapping (F2, RILs) has already been implemented in crops. Finally, although molecular and genetic validations are reliable ways to verify GWAS results, there are still accompanying issues to consider, such as; epistasis, heterosis and environmental factors. Once assembled, such factors will improve our chance of understanding the genetic regulation of complex traits and provide viable targets for crop improvement and breeding.

(i)Biotic resistance of crop species using GWAS

This is a significant concern as biotic interactions are expected to cause yield losses of 20-40% per year. Breeding efforts have, however, been successful – particularly the introgression of resistance alleles from wild species. The importance of an extensive collection of populations, including the above species, makes GWAS an attractive approach to identifying additional genes of interest for this purpose. With so many natural populations now established, it seems likely that their use, and the use of biparental and multi-parental parental populations, will release resistance from a wide range of plant-pest combinations and thus lead to persistent resistance.

(ii) Abiotic resistance of cultured species using GWAS

Similar to the studies described above aimed at creating more resistant plants, substantial research and breeding efforts have been undertaken to identify and exploit the allelic variations that confer tolerance to abiotic stress. Perhaps the most important of these is drought stress, which is estimated to cause 50% yield loss each year (Webber *et al*., 2018). While a lack of water can destroy crop yields, it is quite the opposite. Floods can have the same effect. Therefore, the development of flood-resistant rice varieties is imminent. The identification of SEMIDWARF1 gene haplotypes that facilitated this is an excellent example of the power of haplotype analysis following GWAS studies (Kuroha *et al*., 2018). Similarly, salt stress has been widely studied in rice. (AlTamimi *et al*., 2016) combined high-throughput phenotyping of plant growth and transpiration with high-density genotyping of a few indica and aus diversity panels containing 553 accessions. The study identified a previously undetected salt stress locus mapped to chromosome 11, providing new insights into rice's early response to salt and clues on how breeding can mitigate this problem. Because nitrogen fertilizers are frequently applied to fields with disastrous ecological consequences. Therefore, there is an urgent need to develop crops with high nitrogen utilization efficiency to reduce the use of chemical fertilizers and thus achieve more sustainable agricultural development. (Tang *et al*., 2019). The nitrate transporter $OsNPF6.1^{HapB}$ was highly efficient in nitrogen utilization in a recent GWAS experiment on the rice diversity panel, with haplotypes identified in more than 90% of rice cultivars on this allele.

(iii) Application of GWAS to crop species yield enhancing traits

Given the aforementioned applications of association mapping to plant resistance and tolerance to biotic and abiotic factors, it is essential to note that considerable research effort has also been devoted to elucidating the genetic basis of yield-related traits. The first study tested nearly 5,000 lines of the aforementioned Nested Association Mapping (NAM) maize population, using a simple additive model that can predict flowering time to identify many low-effect QTLs. Besides flowering time, panicle structure is a significant breeding target in rice. A recent evaluation of 49 panicle phenotypes from 242 tropical rice accessions identified 10 GWAS peaks and also demonstrated a subtle association between panicle size and yield (Crowell *et al*., 2016). The complexity of agronomic yield was also highlighted by a study of 84 agronomic traits in a panel of 809 soybean accessions, with many loci showing complex pleiotropic effects (Fang and Luo, 2019).

(iv) Metabolic composition of cultured species using GWAS

The combination of developments in sequencing with developments in mass spectrometry-based analytical systems has made understanding the genetic architecture of metabolism much easier than before. Indeed, the enormous metabolic diversity of plants has created ideal models for dissecting the genetic basis underlying metabolome regulation, with studies progressing from mutant library analysis and gene family analysis to comparisons of sister species and species series within taxa to linkage.

Association mapping based on next-generation sequencing was applied to the metabolomics study. Unlike QTL for an agricultural performance described above, genetic variants controlling natural variation in metabolite accumulation are more easily identified due to both the enormous diversity evident across experimental populations and the high accuracy of metabolite content assessments. In maize, GWAS was used to quantify the metabolite content of nearly 1000 mass elements in more than 700 lines (Zhou *et al*., 2019) and further allowed the association of metabolite traits with grain size, while a more recent study identified four times as many elements paying particular attention to benzoxazinoids and hydroxycitric acids (Wen *et al*., 2014). An earlier pioneering high-complexity study on corn kernel oil identified 74 associated loci, of which 26 were found to explain up to 83% of the phenotypic variation using a simple additive mode.

GWAS studies associations between a genome-wide set of single nucleotide polymorphisms (SNPs) and desired phenotypic traits. Quantitative assessment is based on linkage disequilibrium (LD) through genotyping and phenotyping of different individuals. In general, GWAS infers these associations using a hypothesis test with appropriate test statistics such as Pearson's χ^2-test, Fisher's exact test, F-test, or a regression model under the null hypothesis of no association.

(v) Genome-wide association study of resistance to head blight in maize

Head blight caused by the fungus *Sphacelotheca reliana* is a devastating disease of maize worldwide that results in severe quality and yield losses each year. This study is the first to perform a genome-wide association study of resistance to decapitation using a maize SNP50 Illumina array. From 45,868 single nucleotide polymorphisms in a panel of 144 inbred lines, 18 novel candidate genes were associated with resistance to head blight in maize (Wang *et al*., 2012). These candidate genes were classified into three groups, namely resistance genes, disease response genes, and other genes with possible plant disease resistance functions. The data suggest a complicated molecular mechanism of maize resistance to *S. Reliana*. This study also suggested that GWAS is a useful approach to identifying causal genetic factors for resistance to blight in maize.

***Manhattan plot*:** A scatter plot that plots p-values on a –log10 (p) scale against the genomic position of SNPs and their chromosome number. Large peaks correspond to small p values indicating that the corresponding genomic region has a strong association with the given trait.

***Quantile-quantile (Q-Q) plot*:** A plot used to evaluate how well a model used in a GWAS accounts for specific population structure and familial relatedness. SNPs in the upper right of the graph that deviate from the diagonal are most likely to be associated with the study trait.

Principal component (PC) plot: A method of estimating the effect of population structure by analyzing multivariate data in terms of the covariance structure of the data.

GWAS is projected to investigate the statistical association between phenotypes and a dense set of genetic markers that capture a substantial number of genetic variations in the genome, using many matched samples. Microarray chips are usually genotyped genetic markers. Whether a significant genetic variation in the genome, including standard, rare and structural variations, is captured by the set of markers depends on the number of features and chromosome locations.

The typical number of single nucleotide polymorphism markers used in a current GWAS depends on the exploitation of genetic recombination and allelic diversity for crop improvement, and many of the worlds' farmers depend directly on the harvests of the genetic diversity they sow for food fodder as well as the next season's seed.

Conclusions

The past year has seen remarkable progress in our ability to dissect the genetic basis of common diseases and continuum traits of biomedical significance. The GWAS approach has proven highly suitable for identifying common SNP-based variants with moderate to significant effects on phenotype. Careful implementation and appropriate interpretation led to discoveries that proved more robust than many had anticipated. An increasing number of new susceptibility loci have been identified, shedding light on the underlying mechanisms that influence disease disposition. Much is being learned about the complex relationship between changes in genome sequence and phenotypic variation. We are still unable to explain more than a small fraction of the observed familial clustering for most multifactorial traits, which highlights the need to extend the analysis to a complete range of potential gene and environmental effects. Many of the biggest challenges we will face in the coming years lie not so much in identifying the association. In GWAS, the number of statistical tests is commonly on the order of 106. At the significance level, P values < 0.01 are not significant anymore. Correction of multiple hypothesis testing is crucial.

Confounding structure leads to false positives. It requires favourable conditions like statistical power and resolution, small samples, a large number of hypothesis increased power, and testing compound hypothesis.

Many studies conducted using GWAS as a tool worldwide on different factors like temperature, effect on cobs, agronomic variants, agroclimatic diversities, flowering and grain yield traits, disease diversity, etc. A major benefit of GWAS is one time genotyping and repeated phenotyping in different environment conditions helps to study n number of traits within a short period over a large area. The rapid development of high throughput sequencing technology is that to current model organisms and will slowly become more forced on which species are more relevant for answering biological questions.

The major drawbacks are the need of large population size, cost of preparing DNA samples, and less knowledge about the risk of the trait. To overcome this drawback researchers have upgraded the statistical approaches, proper imputation of genotypes, advanced approaches such as nested association mapping and candidate gene association mapping. Despite its drawbacks, still GWAS is famous due to its dropping genotyping costs, which are likely to drive association studies away from candidate gene association studies. This will likely involve whole genome resequencing of all the individuals in a population, will allow assessment of the effect of point mutation, insertions, deletions and large structure variations.

Conflict of Interest

The authors declare that there is no conflict of interest with present publication.

References

Alseekh S, Kostova D, Bulut M and Fernie AR (2021) Genome-wide association studies: assessing trait characteristics in model and crop plants. *Cellular and Molecular Life Sciences* **78**: 5743-5754. DOI: 10.1007/s00018-021-03868-w

Al-Tamimi N, Brien C, Oakey H, Berger B, Saade S, Ho YS, Schmöckel SM, Tester M and Negrão S (2016) Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. *Nature communications* **7**: 13342. DOI: 10.1038/ncomms13342

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C and Toomajian C (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS genetics* **1**: 60. DOI: 10.1371/journal.pgen.0010060

Bergelson J and Roux F (2010) Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nature Reviews Genetics* **11**: 867-879. DOI: 10.1038/nrg2896

Brachi B, Morris GP and Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome biology***12**:1-8. DOI: 10.1186/gb-2011-12-10-232

Crowell S, Korniliev P, Falcao A, Ismail A, Gregorio G, Mezey J and McCouch S (2016) Genome-wide association and high-resolution phenotyping link Oryza sativa panicle traits to numerous trait-specific QTL clusters. *Nature communications* **7**: 10527. DOI: 10.1038/ncomms10527

Famoso AN, Zhao K, Clark RT, Tung CW, Wright MH, Bustamante C, Kochian LV and McCouch SR (2011) Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and QTL mapping. *PLoS genetics* **7**: 1002221. DOI: 10.1371/journal.pgen.1002221

Fang C. and Luo J (2019) Metabolic GWAS-based dissection of genetic bases underlying the diversity of plant metabolism. *The Plant Journal* **97**: 91-100. DOI: 10.1111/tpj.14097

Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z and Li M (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature genetics* **42**: 961-967. DOI: 10.1038/ng.695

Kuroha T, Nagai K, Gamuyao R, Wang DR, Furuta T, Nakamori M, Kitaoka T, Adachi K, Minami A, Mori Y and Mashiguchi K (2018) Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding. Science **361**: 181-186. DOI: 10.1126/science.aat1577

Liu HJ and Yan J (2019) Crop genome-wide association study: a harvest of biological relevance. *The Plant Journal* **97**: 8-18. DOI: 10.1111/tpj.14139

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y and Tanaka T (2002) Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature genetics* **32**: 650-654. DOI: 10.1038/ng1047

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM and Buckler IV ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the national academy of sciences* **98**:11479-11484. DOI: 10.1073/pnas.201394398

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA and Minx P (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**: 1112-1115. DOI: 10.1126/science.1178534

Tang W, Ye J, Yao X, Zhao P, Xuan W, Tian Y, Zhang Y, Xu S, An H, Chen G and Yu J (2019) Genome-wide associated study identifies NAC42-activated nitrate transporter conferring high nitrogen use efficiency in rice. *Nature communications* **10**: 5279. DOI: 10.1038/s41467-019-13187-1

Wang M, Yan J, Zhao J, Song W, Zhang X, Xiao Y and Zheng Y (2012) Genome-wide association study (GWAS) of resistance to head smut in maize. *Plant science* **196**: 125-131. DOI: 10.1016/j.plantsci.2012.08.004

Webber H, Ewert F, Olesen JE, Müller C, Fronzek S, Ruane AC, Bourgault M, Martre P, Ababaei B, Bindi M and Ferrise R (2018) Diverging importance of drought stress for maize and winter wheat in Europe. *Nature communications* **9**: 42-49. DOI: 10.1038/s41467-018-06525-2

Wen W, Li D, Li X, Gao Y, Li W, Li H, Liu J, Liu H, Chen W, Luo J and Yan J (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature communications* **5**: 3438. DOI: 10.1038/ncomms4438

Zhou S, Kremling KA, Bandillo N, Richter A, Zhang YK, Ahern KR, Artyukhin AB, Hui JX, Younkin GC, Schroeder FC and Buckler ES (2019) Metabolome-scale genome-wide association studies reveal chemical diversity and genetic control of maize specialized metabolites. *The Plant Cell* **31**: 937-955. DOI: 10.1105/tpc.18.00772