# Genome-wide association studies of important agronomic traits in Brassica napus: what we have learned and where we are headed

Article

Accepted Version

Liu, H., Wang, W., Yang, M., Yuan, P., Hammond, J. P. ORCID: https://orcid.org/0000-0002-6241-3551, King, G. J., Ding, G., White, P. J., Wang, S., Cai, H., Wang, C., Lu, C., Xu, F. and Shi, L. (2022) Genome-wide association studies of important agronomic traits in Brassica napus: what we have learned and where we are headed. Annual Plant Reviews online, 5 (2). pp. 151-180. ISSN 2639-3832 doi: 10.1002/9781119312994.apr0801 Available at https://centaur.reading.ac.uk/105676/

Published version at: http://dx.doi.org/10.1002/9781119312994.apr0801
To link to this article DOI: http://dx.doi.org/10.1002/9781119312994.apr0801

Publisher: Wiley

Publisher statement: This is the peer reviewed version of the following article: 'Genome-Wide Association Studies of Important Agronomic Traits in Brassica napus: What We Have Learned and Where We Are Headed', which has been published in final form at https://doi.org/10.1002/9781119312994.apr0801. This

# Genome-wide association studies of important agronomic traits in *Brassica napus*: what we have learned and where we are headed

Haijiang Liu[1,2] , Wei Wang[1,2], Meng Yang[1], Pan Yuan[1,2], John P. Hammond[3,4], Graham J. King[4], Guangda Ding[1,2], Philip J. White[1,5], Sheliang Wang [1,2], Hongmei Cai[2], Chuang Wang[2], Chungui Lu[6], Fangsen Xu[1,2], Lei Shi[1,2]*

1 National Key Lab of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

2 Key Lab of Cultivated Land Conservation, Ministry of Agriculture and Rural Affairs/ Microelement Research Centre, Huazhong Agricultural University, Wuhan 430070, China

3 School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AR, UK

4 Southern Cross Plant Science, Southern Cross University, Lismore, NSW 2480, Australia

5 The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK

6 School of Animal, Rural and Environmental Sciences, Nottingham Trent University, Nottingham, UK

*Corresponding Author：Tel: (+8627) 87286871; E-mail: leish@mail.hzau.edu.cn

**Abstract**: Oilseed rape (*Brassica napus* L.; *B. napus*) is one of the main oil crops in China as well as in the world. Genome-wide association studies (GWAS) have revolutionized the field of complex agronomic traits. In *B. napus,* these include seed yield and yield–related traits, seed oil content, abiotic and biotic stress tolerance traits over the past decade, in which hundreds of thousands to millions of genetic variants across the genomes of hundreds of individuals have been tested to identify genotype–phenotype associations. In this review, we assess the current status of GWAS in terms of genotypes, phenotypes, statistical models and candidate genes for these agronomic traits in *B. napus*. Post-GWAS, the combination of QTL mapping, transcriptomics and new statistical methods has allowed us to identify candidate genes associated with specific agronomic traits. In addition, we can use diverse populations, increase the population size or look for rare variants and structural variations of *B. napus* by whole-genome sequencing to minimize "missing heritability" effects. These approaches are essential for uncovering the genetic mechanisms defining or regulating complex agronomic traits and delivery of molecular marker assisted breeding in *B. napus* to breed new varieties that are higher yielding but resilient to our changing climate.

**Key words:** Oilseed rape, Agronomic traits, GWAS, Genotypes, Phenotypes, Genetic mechanisms

## Introduction

Oilseed rape (OSR, rapeseed, canola; *B. napus*) originated in Mediterranean region approximately 7,500 years ago by natural hybridization between *B. rapa* and *B. oleracea* (Chalhoub et al., 2014； Lu et al., 2019). It has since spread geographically and economically, becoming the second most important oil crop globally for edible oil (USDA ERS, 2020). The growing demand for *B. napus* oil has led to the continuous genetic improvement of various important agronomic traits. However, an increasing or static yield gap in many countries and our changing climate and reduced efficacy of pesticides means abiotic and biotic pressures are increasing and new resources and approaches to breeding high yielding and resilient germplasm are urgently needed.

Genome-wide association studies (GWAS), in which hundreds of thousands to millions of genetic variants across the genomes of many individuals are tested to identify genotype–phenotype associations (Figure 1), have revolutionized the field of complex agronomic traits of *B. napus* over the past decade (Tang et al., 2021). In 2010 the first GWAS was reported in *B. napus*, in which 684 amplified fragment length polymorphism (AFLP) markers in 84 canola quality winter rapeseed cultivars were used to explore the genes affecting phenological, morphological and quality traits (Honsdorf et al., 2010). To date, diverse traits of *B. napus*, ranging from molecular scale (e.g. the transcriptome) to whole plant scale (e.g. seed yield, seed oil content (SOC), root and shoot architecture traits, flowering time), and from abiotic stress tolerance (e.g. in response to phosphorus (P) deficiency, salt and drought stress etc.) to biotic stress tolerance (e.g. stem rot and clubroot), have been studied using GWAS approaches (Table 1). The number of articles on GWAS in *B. napus* has increased year by year (Figure 2), which suggests that this approach is a powerful tool to analyze the genetic structure of important agronomic traits in *B. napus*. Herein, we (1) review the advances on the *B. napus* functional genomics facilitated by GWAS; and (2) assess the challenges and opportunities of *B. napus* GWAS going forward.

### ***Genotypes used in GWAS in B. napus***

Single nucleotide polymorphism (SNP) chips and whole- genome sequencing (WGS) are the most widely used genotyping methods currently used to identify genetic variants within genotypes from populations used for GWAS. To date, the 60k SNP chips have been used in approximately 54% of *B. napus* GWAS (Table 1). However, with the development of WGS technology, GWAS based on highdepth WGS is gradually increasing (Wu et al., 2019; Lu et al., 2019; Tang et al., 2021). Compared with SNP chips, WGS is capable of detecting more genetic variants, especially rare and structural variations, but the cost is higher, and the analysis method is more complicated (Table 2).

GWAS is primarily designed to test SNVs (single-nucleotide variants) for association with traits. However, other types of genetic variants can also be detected by GWAS. For example, structural variations (SVs), including small insertions and deletions (InDels) and presence–absence variation (PAV), were prevalent in the *B. napus* genome and some studies show they make significant contributions to *B. napus* phenotypic variations (Gabur et al., 2018; Song et al., 2020). A hundred and fifteen significant PAV were associated with blackleg resistance in *B. napus*, of which, 41 were significant at false discovery rate (FDR) ≤ 0.1 (Gabur et al., 2018). Recently, more than one million PAVs were identified and enriched in associations with silique length, seed weight and flowering time of *B. napus* by performing *de novo* assembly and annotation of eight representative *B. napus* genomes (Song et al., 2020). Compared to SNPs, PAVs contribute much more to these phenotypes (Song et al., 2020).

Besides SNVs and SVs, transcriptional variation is also widely used in GWAS of *B. napus* (Harper et al., 2012; Lu et al 2014; Alcock et al., 2017; Alcock et al., 2018; Tang et al., 2021). GWAS for transcriptomic variation is also called transcriptome-wide association studies (TWAS) or associative transcriptomics. GWAS with both SNPs and gene expression markers (GEMs) identified *Bna.HACG*, a gene that controls glucosinolate biosynthesis in *B. napus* seeds (Harper et al., 2012). GWAS combined with TWAS revealed two candidate genes, *Bna.A10.ACA8* and *Bna.C02.MGT7*, responsible for the leaf accumulation of calcium and magnesium, respectively (Alcock et al., 2017). Subsequently, a total of five GEMs markers associated with leaf potassium concentration were identified, and 9 candidate genes were predicted (Alcock et al., 2018). Recently, a combination of GWAS and TWAS was employed to uncover the regulatory networks associated with SOC in *B. napus* and each expression quantitative trait loci (eQTL) can largely explain the variation of gene expression signatures (Tang et al., 2021).

### ***Phenotypes used in GWAS in B. napus***

Natural population is the most commonly used population for GWAS analysis in *B. napus*, and generally consists of three types of rapeseed, i.e. spring, winter and semi-winter (Wu et al., 2019; Lu et al., 2019; Tang et al., 2021). However, in some studies, a few *B. napus* types within the experimental population were probably not adapted to the climate of the growing area. For example, spring type rapeseed is not suitable for the cold climate because it will cause low seed yield, but in some GWAS, the spring, winter, and semi-winter rapeseed were planted in the same growing area without considering the impact of climate to study the yield and yield-related traits (Sun et al., 2016b; Lu et al., 2017; Li et al., 2021). Consequently, the differences in the seed yield and yield related traits are mostly attributed to the growing environment, rather than the yield-related genetic factors. Therefore, researchers should construct a reasonable experimental population according to the adaptability (e.g. climate) of the plants. Some researchers have also noticed this problem and made a reasonable design for the experimental population (Kumar et al., 2018). In France, the spring type rapeseed does not grow well, which has resulted in a compromised investigation into the impact of blackleg disease. Considering this, a population of 166 winter *B. napus* were planted in France to conduct GWAS analysis of blackleg disease (Kumar et al., 2018).

In *B. napus* GWAS, some phenotypic traits, such as flowering time, seed yield, plant height, branch number, and seed weight, can be directly measured (Figure 3a); some other traits such as tolerance to biotic and abiotic stresses, cannot be directly investigated. For the latter, some parameters, such as root length under P deficiency (Wang et al., 2017b), shoot fresh weight under salt stress (Wan et al., 2017), root to shoot length ratio under water stress (Zhang et al., 2015), radicle length under lead (Pb) stress (Zhang et al., 2020), seed vigor under low-temperature stress (Luo et al., 2021), the stem rot disease index (Wei et al., 2016) and root fresh weight under drought stress (Khanzada et al., 2020) were used to conduct GWAS. Although these parameters could indicate the ability of plants to tolerate stress they increase the complexity and the difficulty of uncovering the genetic mechanism in response to abiotic and biotic stress by GWAS. To overcome this, one opportunity is to quantify the metabolites directly related to the stress in *B. napus* and perform metabolite-based GWAS (Figure 3b). Metabolite-based GWAS has previously been used to identify 1,459 significant locus–trait associations in maize, with more than 58.5% identified as significant loci for 983 metabolites in maize kernel supported by expression QTLs, and some (14.7%) co-located with QTLs in previous linkage analysis, and two genes were further validated by mutant and transgenic analysis (Wen et al., 2014). Metabolite association studies will be helpful to understand the functional mechanism of complex traits and identify the candidate gene(s). GWAS of metabolites of *B. napus* under abiotic and biotic stress can be performed in the future.

Four strategies are useful to improve the precision of an experiment: (1) Blocking, (2) Randomization, (3) Replication, and (4) Repeat measurements. The first two are smart strategies because they do not require large samples. The last two are costly strategies, but they are necessary if random and measurement errors must be estimated and controlled, such as one-year/multi-location experiments, one-location/multi-year experiments, multi-year / multi-location experiments. Best linear unbiased prediction (BLUP) analysis has been used successfully to deal with multi-year phenotypic values of agronomic traits, such as plant height and branch number (Liu et al., 2021b), silique number (Li et al., 2020), seed oil content (Wang et al., 2018), seed phytate content (Liu, et al., 2021a) in rapeseed, and ionomic traits in rice (Yang et al., 2018). R package ‘lme4’ (https://cran.r-project.org/web/packages/lme4/index.html) is used to calculate the BLUP values of investigated traits.

### *GWAS of important agronomic traits in B. napus*

GWAS of seed yield and seed yield related traits, including plant height, first branch height, inflorescence length, silique length, seeds per silique and seed weight of *B. napus* were first reported in 2014 (Cai et al., 2014; Table 1). A panel of 192 inbred lines of *B. napus* were genotyped using 740 AFLP markers, and a total of 43 AFLP markers were identified associated with yield and yield-related traits (Cai et al., 2014; Table 1). Five common markers were associated with both first branch height and plant height, and one marker was associated with both silique length and seed weight in the three-year experiment (Cai et al., 2014; Table 1). In a larger study, using a panel of 157 *B. napus* accessions and 690953 SNPs, 20 SNPs were significantly associated with seed weight on A01, A04, A09, C02, and C06 chromosomes (Dong et al., 2018; Table 1). Recently, six candidate genes for silique number were identified through a combination of GWAS and transcriptome profiling (Li et al., 2020a; Table 1). This study also highlighted that natural variation of silique number is largely affected by *B. napus* biomass and nutrient accumulation, which highlighted a potential new approach for the genetic improvement of silique number in *B. napus* (Li et al., 2020a). Recently, GWAS combined with QTL analysis revealed the genetic architecture of seed yield at high planting density in *B. napus* and identified two candidate genes (*BnaA02.TCP1* and *BnaA02.HY5*) for it (Menendez et al., 2021). Overexpression of *BnaA02.TCP1* in *Arabidopsis thaliana* promoted growth independently of plant density, and overexpression of *BnaA02.HY5* increased seed yield and biomass at high planting density (Menendez et al., 2021).

One hundred and forty three SNPs were significantly associated with flowering time by a GWAS study with 368 *B. napus* accessions and 11,804 SNPs, and *Bna.A02.FT* and *Bna.C02.FT* were identified as candidate genes (Raman et al., 2019). In another GWAS study for flowering time with 991 *B. napus* accessions and more than nine million SNPs, *Bna.A02.FT* was also detected (Wu et al., 2019).

Subsequently, *Bna.A02.FLC and Bna.A02.FT* were confirmed as candidate genes for flowering time by GWAS study of flowering time with a panel of 950 F1 hybrids from 475 spring-type canola (Jan et al., 2019). In these studies, *Bna.A02.FT* was mapped by different research groups using different *B. napus* accessions, showing its dominant effect in controlling the variation of flowering time in *B. napus*.

For SOC, 39 SNPs were identified by a combined GWAS (227 *B. napus* accessions and 34,000 high-quality SNP markers) and a QTL mapping approach, with explained phenotypic variations ranging from 10.23% to 24.45%. SNPs located on A05, A08 and A10 co-located with previously reported QTLs associated with SOC (Sun et al., 2016c). Herein, the QTL means the significant SNPs with close proximity (Sun et al., 2016c; Li et al., 2021). In another GWAS study with 521 *B. napus* cultivars and 52157 SNPs, 50 SNPs were significantly associated with SOC, of which, 29 were newly discovered (Liu et al., 2016b). Recently, 27 loci for SOC were identified, and two candidate genes of *BnaA05.PMT6* and *BnaC05.PMT6* were confirmed as negative regulators of SOC by a combined GWAS and TWAS using a panel of 505 *B. napus* accessions and more than 10 million SNPs from WGS (Tang et al., 2021). This is the first study that used GWAS/TWAS to identify and confirm candidate genes in *B. napus* using knockout mutants and overexpression lines.

GWAS of root related traits under P deficiency identified 285 SNP loci associated with these traits, and a haplotype of '*BnA03Hap*' on A03 chromosome, which will be important for breeding high P efficient varieties (Wang et al., 2017b). Nine SNPs associated with Pb accumulation in *B. napus* were identified by GWAS and the expression of five candidate genes (*GSTUs*, *BCATs*, *UBP13*, *TBR* and *HIPP01)* were verified by Quantitative Real-time PCR (qRT-PCR) (Zhang et al., 2020). GWAS was undertaken to uncover the genetic basis of low-temperature stress in *B. napus* using more than 8 million SNPs by WGS, and 22 loci were identified. Underlying these loci, five candidate genes, *BnaA03g40290D,* *BnaA06g07530D, BnaA09g06240D, BnaA09g06250D,* and *BnaC02g10720D* associated with low-temperature stress were confirmed (Luo et al., 2021).

In biotic interaction, GWAS identified 17 significant SNPs associated with stem rot resistance on chromosomes A08 and C06 by using an association panel with 347 *B. napus* accessions, with a *tau class glutathione S-transferase* gene cluster identified by combing GWAS and transcriptome analysis (Wei et al., 2016). Eight co-located SNPs for blackleg (*Leptosphaeria maculans*) resistance in *B. napus* were identified by combining GWAS analysis and QTL mapping (Raman et al., 2016). Subsequently, sixteen significant SNPs associated with blackleg disease were identified by combining GWAS analysis and QTL mapping, of which nine were co-located with the genomic regions in a previous study (Kumar et al., 2018).

In total, thousands of genetic variations have been identified for various traits in *B. napus* by GWAS in recent years. If the genetic variations associated with seed yield and yield-related traits are also associated with other traits, such as biotic and abiotic stress resistance traits, there may be a gene controlling both traits (pleiotropy) near the location of genetic variation.

## *A tip of the iceberg – current GWAS in B. napus*

### *Interplay of genetics and environmental factors on complex agronomic traits*

Complex agronomic traits result from the interplay of genetics and environmental factors. For example, seed yield in *B. napus* is dependent on complex interactions between genetic predisposition, abiotic stress (e.g. water or nutrients deficiency) and biotic stress (e.g. *Sclerotinia Sclerotiorum*, Clubroot Disease) (Zhang et al., 2015; Wang et al., 2017b; Fredua-Agyeman et al., 2020). For example, P is an essential nutrient in the regulation of plant growth and development and P deficiency reduces seed yield significantly. However, in most of the GWAS studies on the seed yield of *B. napus*, soil nutrient availability is not reported. Given there are significant differences in the P use efficiency among different varieties in an association panel of *B. napus* (Wang et al., 2017b) or if the soil used for phenotyping is severely P deficient, the seed yield trait will contain a component of the adaptation to deficient P conditions, and the genes associated with seed yield may also be associated with the P use efficiency. With most GWAS only focusing on specific phenotypes, even minimizing environmental variables, there is the possibility that the measured trait and SNP associations are the product of these interactions. In addition, different growth stages, different planting location, and the impact of abiotic and biotic stresses may affect the final yield in different ways and through different genetic and metabolic interactions (Figure 4). GWAS findings today therefore represent the starting point for the exploration of a wider range of phenotype-environment interactions through GWAS and is likely to lead to additional discoveries (Figure 4).

### ***Populations and population size***

Most populations used in GWAS of *B. napus* are natural populations (Table 1). The abundant diversity of natural populations can make GWAS reach single variance level resolution. However, the population structure, kinship and rare variants in natural populations may affect the statistical power of GWAS (Flint-Garcia et al., 2005). It is difficult to detect rare variants underlying target traits if they are significantly affected by the population structure (Flint-Garcia et al., 2005). To overcome false associations and improve detection ability of rare alleles in crops, multi-parent cross populations have been developed, including nested association mapping (NAM), random-open-parent association mapping (ROAM), and multi-parent advanced generation Inter-Crosses (MAGIC) (Yu and Buckler, 2006; Dell'Acqua et al., 2015; Xiao et al., 2017). Recently, a *B. napus* NAM population consisting of 2425 $F_6$ RILs derived from crosses between the common parent Zhongshuang11 (ZS11) and 15 diverse semi-winter oilseed rape and spring oilseed rape founder lines, has been developed to dissect the genetic architecture of some important agronomic traits, such as silique length, seed weight and flowering time (Hu et al., 2018; Song et al., 2020). In future research, the complementary advantages of natural and artificial populations will further accelerate the dissection of important agronomic traits in *B. napus* (Figure 4). In addition, larger sample sizes in association populations are necessary for the identification of additional loci and rare variants because sample size is the primary limiting factor in genetic variation discovery. A population of 200 individuals can detect QTLs that explain 20% of the phenotypic variance with 90% power. However, if the QTLs only explained 5% of the phenotypic variation, the population would need to have more than 1000 individuals with the same power of detection (Gatti et al., 2014). Therefore, using GWAS to identify a new trait, where its phenotypic variance is not clear, and increasing the population size could improve the detection efficiency of the phenotypic variance (Figure 4).

***Statistical models***

Statistical models are an important part hidden in the GWAS "iceberg". GLM and MLM are two models commonly used in *B. napus* GWAS (Figure 4, Table 1). GLM only uses the population structure, and the outputs have a high level of false positives. MLM uses both the population structure and kinship; however, the calculation speed is slower and the computational resources are higher compared with GLM. In recent years, large-scale whole-genome resequencing of *B. napus* natural populations had been carried out (Wu et al., 2018; Lu et al., 2019; Tang et al., 2021). Several models, such as EMMXA (Wu et al., 2018;), mrMLM (Lu et al., 2019; Qian et al., 2021), GEMMA (Luo et al., 2021) and Fast-LMM (Tang et al., 2021), have been used to conduct GWAS based on WGS data, which have faster calculation speeds and higher powers of detection. In addition, a method has been developed to predict candidate genes in each QTL based on multi-layered data and information of *Arabidopsis* (Tang et al., 2021). The use and development of new statistical models will greatly improve the identification of candidate genes in the future (Figure 4).

### *Rare variation and structural variants*

GWAS has been used to identify thousands genes for human diseases and agronomic traits in crops，which is accelerating our understanding of the genetic components of these traits, such as age-related macular degeneration in human, grain weight in rice and SOC in oilseed rape (Klein et al., 2005; Si et al., 2016; Tang et al., 2021). However, only a small portion of phenotypic variation can be explained in almost all the studied traits using GWAS, especially in humans (Génin et al., 2020). For example, across 5000 human GWAS, 68,000 SNP-trait associations were identified in humans (the GWAS catalog, April 2018). These SNPs only explain a modest proportion of the estimated heritability of all the traits except for those such as age-related macular degeneration and exfoliation glaucoma (Manolio et al., 2009; Manolio et al., 2013; Buniello et al. 2019). This is known as the “missing heritability”. For example, in *B. napus*, 285 root related SNPs have been identified, but only 5.8% of them explained 10% of the phenotypic variation (Wang et al., 2017b). The first reason for this is GWAS is not capable of capturing all the genetic associations with target traits, and most of the variants with small effects are missed. The second reason is that the rare variations which are not captured by SNP-chip based GWAS may be the main factor(s) controlling the traits. For example, the final genotype files used for SNP-chip based GWAS in *B. napus* do not contain rare variants because the genotype files are usually filtered according to minor-allele frequency (MAF) > 0.05, which may lead to the rare variants that are associated with the phenotype not being represented and identified. In addition, structural variants are not represented in SNP chips, which are completely ignored in early GWAS analysis of complex traits. Structural variants are important genome variations in *B. napus*, which encompass more nucleotides on the genome than SNPs (Song et al., 2020). Incorporation and analysis of structural variants-GWAS are proposed to promote the genetic analysis of complex agronomic traits in *B. napus* in the future.

### *Connecting GWAS to biological mechanisms underlying complex traits*

Although most of *B. napus* GWAS have focused on candidate gene mining, few new genes have been cloned or validated through functional molecular approaches. This could be attributed to the limitations of GWAS, such as target traits being associated with unrelated loci if they are significantly correlated to population structure and kinship, which is a common cause of false positives (Flint-Garcia et al., 2005; Larsson et al., 2013). The values of LD decay of *B. napus* are associated with the population size, such as 298 kb in 300 inbred lines (Zhou et al., 2018), 238 kb in 505 inbred lines (Liu et al., 2021a), 0.7 Mb in 280 inbred lines (Li et al., 2021), 1.6 Mb in 177 inbred lines (Dakouri et al., 2021). Nevertheless, the LD decays observed in *B. napus* are larger than that of other species, e.g. 50 kb in Arabidopsis (Nordborg et al., 2005), 200 kb in rice (Li et al., 2017) and 160 kb in maize (Li et al., 2019). Large LD decay in *B. napus* indicates an association locus contains more than 100 genes, and thus it is difficult to pinpoint the causal genes associated with the significant loci.

Consequently, loci identified by GWAS associated with traits may only explain a small fraction of the heritability of the target traits and may represent a false association or not necessarily causal variants and candidate genes (Tam et al., 2019). These shortcomings highlight the need to be cautious in the interpretation of GWAS (Tam et al., 2019). Here, a method to narrow down the range of candidate genes are summarized. Firstly, GWAS can be combined with other methods, such as transcriptomics, to predict the candidate genes. Twenty seven percent of the published papers on GWAS of *B. napus* indicate that the differentially expressed genes in the target interval are causal genes (Table 1). Secondly, candidate gene association analysis, especially the differentially expressed candidate genes, are performed to identify whether there are SNPs significantly related to the target phenotype. If the SNP is significantly associated with the phenotype, its position within the candidate gene (e.g. the promoter, 5'UTR, intron or exon) and its subsequent impact on expression level or protein sequence should be explored. For example, there are two genes (*BnaA10g09290D*, *BnaC08g26640D*) identified by GWAS as candidate genes for controlling plant height under low phosphorus stress (Liu et al., 2021b). Candidate gene association analysis demonstrated that chrA10_8216680 (T/A), located in the exon region of the candidate gene *BnaA10g09290D,* resulted in amino acid changes from isoleucine to asparagine, and the SNP of chrC08_27999778 (A/T) located in the exon region of the candidate gene *BnaC08g26640D* resulted in amino acid changes from isoleucine to asparagine. Both SNPs showed strong association with the trait; therefore, these two SNPs should be the direct effect sites underlying the loci that contribute to the phenotypic difference in plant height under low phosphorus stress (Liu et al., 2021b). In short, if the candidate gene harbors a SNP that causes a significant differentiation in expression levels or protein sequence and that also shows strong association with the target phenotype, the gene is most likely to be the causal gene.

In addition, GWAS researchers should be encouraged to take on the challenge and go beyond identification of loci, with the functional validation, through molecular approaches of causal candidate genes (Figure 4). Recently, 424 significant SNPs were associated with trichome appearance in a *B. napus* GWAS, and a candidate gene, *BnaA06.GL1.a*, was identified by the combination of GWAS with RNA-seq. Functional validation using *Arabidopsis* lines carrying the *35S:BnaA.GL1.a* cassette formed trichomes on the stems and leaves, confirming its role in trichome development and indicating that *B. napus* might have a similar regulatory network that controls the development of trichome to *Arabidopsis* (Xuan et al., 2020). Twenty-seven loci for SOC were detected and two candidate genes *BnaA05.PMT6* and *BnaC05.PMT6* were identified. In the knockout mutants of *BnaA05.PMT6* and *BnaC05.PMT6*, the SOC of *B. napus* increased by 3.0% (Tang et al., 2021). Recently, GWAS analysis combined with the transcriptome identified the candidate gene *BnaA07.FAX1-1* responsible for the natural variation of total plant dry matter. The *B. napus BnaA07. FAX1-1* overexpression lines grown under field conditions showed significantly higher biological yield, seed yield, and seed oil content compared with the wild type (Xiao et al., 2021). Fifteen SNPs were identified to be associated with the seed glucosinolates by GWAS and four candidate genes were predicted by GWAS combined with TWAS, comprehensive bioinformatics, and pocket algorithm analyses. Among these four genes, knocking-out only the *B. napus BnaC02.GTR2* resulted in lower glucosinolates acid and higher oil content in seeds, and therefore this gene is likely to be the causal gene (Tan et al., 2021).

### *The home of GWAS: From laboratory to field*

To highlight the power of GWAS, a successful case from laboratory to field was demonstrated in the development of maize varieties with increased concentrations of vitamin A (VA). The candidate alleles of *LcyE* and c*rtRB1* were detected by a candidate gene association study (Harjes et al., 2008), and then were introgressed into high yielding maize varieties by molecular marker assisted breeding. The maize with high VA concentrations is now being consumed by malnourished children in Africa, with the prevalence of VA deficiency declining in this region (Fiedler et al., 2014). *B. napus* GWAS have detected many favorable alleles and haplotypes, which can be used for further molecular marker-assisted breeding. For example, eight successive SNPs located near the peak SNP Bn-A03-p3052882 on A03 of *B. napus* formed nine haplotypes, designated '*BnA03Hap*' haplotypes, and the inbred lines carring GAAAAAGG at '*BnA03Hap6*' have higher than average values in primary root length, total root length, lateral root length, and root dry weight than those with the other haplotypes (Wang et al., 2017b). Recently, GWAS of plant height of *B. napus* at low P supply identified133 SNPs for both plant height and branch number and proposed three candidate genes (*BnaA10g09290D*, *BnaC08g26640D* and *BnaA02g33340D*) for plant height (Liu et al., 2021b). Candidate gene association and haplotype analysis identified several P -efficient haplotypes, e.g. *BnaA10g09290Hap1* and *BnaC08g26640Hap1*, which have more branch number and higher plant height than P -inefficiency haplotypes (Liu et al., 2021b). One hundred and nineteen significant SNPs were identified to be associated with phytate concentration in *B. napus*, and a low phytate haplotype *('BnaA09g10220ConcHap1')* of the candidate gene (*BnaA9.MRP5*) was proposed (Liu et al., 2021a). In the future, genomic region-based or gene-based GWAS will provide greater confidence in associations between loci and traits, and will contribute to the subsequent breeding of higher yielding and more resilient varieties of *B. napus* (Figure 4).

## Conclusions

The emergence of GWAS has caused a remarkable shift in our capacity to understand the genetic basis of plant agronomic and economic traits. At present, a large number of papers have reported the application of GWAS in the analysis of various agronomic traits of *B. napus*. These studies have identified large-scale genetic variation in seed yield and yield related traits, SOC, root architecture traits in response to P deficiency and established the reliability of association analysis. The combination of QTL mapping, transcriptomics, metabolomics or other methods may help to narrow down the candidate genes. In addition, using more diverse populations, increasing the population size or identifying rare or structural variations of *B. napus* by WGS to minimize "missing heritability" effects, will help to reveal more of the 'GWAS discoveries' iceberg.

## ACKNOWLEDGMENTS

This work was supported by the National Nature Science Foundation of China (Grants No. 32172662 and 31972498). We also acknowledge the Natural and Fundamental Research Funds for the Central Universities of China (Grant No. 2662019PY013). PJW was supported by the Rural and Environment Science and Analytical Services Division (RESAS) of the Scottish Government.

## Conflict of interest

The authors declare no competing financial interest.

## Author contributions

HJL, PJW and LS developed the concept. HJL performed data analysis. HJL, JH and LS wrote the manuscript. WW, MY, PY, GJK, GDD, SLW, HMC, CW, CGL and FSX revised the manuscript.

## Reference

Alcock, T.D., Havlickova, L., He, Z., Bancroft, I., White, P.J., Broadley, M.R. and Graham, N.S. (2017) Identification of Candidate Genes for Calcium and Magnesium Accumulation in Brassica napus L. by Association Genetics. *Front Plant Sci* **8**, 1968.

Alcock, T.D., Havlickova, L., He, Z., Wilson, L., Bancroft, I., White, P.J., Broadley, M.R. and Graham, N.S. (2018) Species-Wide Variation in Shoot Nitrate Concentration, and Genetic Loci Controlling Nitrate, Phosphorus and Potassium Accumulation in Brassica napus L. *Front Plant Sci* **9**, 1487.

Arifuzzaman, M., Oladzadabbasabadi, A., McClean, P. and Rahman, M. (2019) Shovelomics for phenotyping root architectural traits of rapeseed/canola (Brassica napus L.) and genome-wide association mapping. *Mol Genet Genomics* **294**, 985-1000.

Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P.L., Amode, R., Guillen, J.A., Riat, H.S., Trevanion, S.J., Hall, P., Junkins, H., Flicek, P., Burdett, T., Hindorff, L.A., Cunningham, F. and Parkinson, H. (2019) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012.

Bus, A., Korber, N., Parkin, I.A., Samans, B., Snowdon, R.J., Li, J. and Stich, B. (2014) Species- and genome-wide dissection of the shoot ionome in Brassica napus and its relationship to seedling development. *Front Plant Sci* **5**, 485.

Cai, D., Xiao, Y., Yang, W., Ye, W., Wang, B., Younas, M., Wu, J. and Liu, K. (2014) Association mapping of six yieldrelated traits in rapeseed (Brassica napus L.). *Theor Appl Genet* **127**, 85-96.

Harjes, C.E., Rocheford, T.R., Bai, L., Brutnell, T.P, Kandianis, C.B., Sowinski, S.G., Stapleton, A.E., Vallabhaneni, R., Williams, M, Wurtzel, E.T, Yan, J.B., Buckler. E.S. (2008) Natural Genetic Variation in Lycopene Epsilon Cyclase Tapped for Maize Biofortification. *Science* **319**, 331-333.

Chalhoub, B., Denoeud, F., Liu, S., et al. 2014, Early allopolyploid evolution in the post-Neolithic B. napus oilseed genome, Science, 345, 950–3.

Chen, L., Wan, H., Qian, J., Guo, J., Sun, C., Wen, J., Yi, B., Ma, C., Tu, J., Song, L., Fu, T. and Shen, J. (2018) Genome-Wide Association Study of Cadmium Accumulation at the Seedling Stage in Rapeseed (Brassica napus L.). *Front Plant Sci* **9**, 375.

Dakouri, A., Lamara, M., Karim, M.M., Wang, J., Chen, Q., Gossen, B.D., Strelkov, S.E., Hwang, S.F., Peng, G. and Yu, F. (2021) Identification of resistance loci against new pathotypes of Plasmodiophora brassicae in Brassica napus based on genome-wide association mapping. *Sci*

*Rep* **11**, 6599.

Dell'Acqua, M., Gatti, D.M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A.L., Aung, H.H., Nelissen, H., Baute, J., Frascaroli, E., Churchill, G.A., Inze, D., Morgante, M. and Pe, M.E. (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. *Genome Biol* **16**, 167.

Dong, H., Tan, C., Li, Y., He, Y., Wei, S., Cui, Y., Chen, Y., Wei, D., Fu, Y., He, Y., Wan, H., Liu, Z., Xiong, Q., Lu, K., Li, J. and Qian, W. (2018) Genome-Wide Association Study Reveals Both Overlapping and Independent Genetic Loci to Control Seed Weight and Silique Length in Brassica napus. *Front Plant Sci* **9**, 921.

Duan, X., Wang, X., Jin, K., Wang, W., Liu, H., Liu, L., Zhang, Y., Hammond, J.P., White, P.J., Ding, G., Xu, F. and Shi, L. (2021) Genetic Dissection of Root Angle of Brassica napus in Response to Low Phosphorus. *Front Plant Sci* **12**, 697872.

Farid, M., Yang, R.C., Kebede, B, Rahman, H. (2020) Evaluation of Brassica oleracea accessions for resistance to Plasmodiophora brassicae and identification of genomic regions associated with resistance. *Genome* **63**, 91-101.

Fiedler, J.L., Afidra, R., Mugambi, G., Tehinse, J., Kabaghe, G., Zulu, R., Lividini, K., Smitz, M.F., Jallier, V., Guyondet, C. and Bermudez, O. (2014) Maize flour fortification in Africa: markets, feasibility, coverage, and costs. *Ann N Y Acad Sci* **1312**, 26-39.

Fikere, M., Barbulescu, D.M., Malmberg, M.M., Spangenberg, G.C., Cogan, N.O.I. and Daetwyler, H.D. (2020) Meta-analysis of GWAS in canola blackleg (Leptosphaeria maculans) disease traits demonstrates increased power from imputed whole-genome sequence. *Sci Rep* **10**, 14300.

Flint-Garcia, S.A., Thuillet, A.C., Yu, J., Pressoir, G., Romero, S.M., Mitchell, S.E., Doebley, J., Kresovich, S., Goodman, M.M. and Buckler, E.S. (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J* **44**, 1054-1064.

Fredua-Agyeman, R., Yu, Z., Hwang, S.F. and Strelkov, S.E. (2020) Genome-Wide Mapping of Loci Associated With Resistance to Clubroot in Brassica napus ssp. napobrassica (Rutabaga) Accessions From Nordic Countries. *Front Plant Sci* **11**, 742.

Fu, F., Zhang, X., Liu, F., Peng, G., Yu, F. and Fernando, D. (2020) Identification of resistance loci in Chinese and Canadian canola/rapeseed varieties against Leptosphaeria maculans based on genome-wide association studies. *BMC Genomics* **21**, 501.

Gabur, I., Chawla, H.S., Liu, X., Kumar, V., Faure, S., von Tiedemann, A., Jestin, C., Dryzska, E., Volkmann, S., Breuer, F., Delourme, R., Snowdon, R. and Obermeier, C. (2018) Finding invisible quantitative trait loci with missing data. *Plant Biotechnol J* **16**, 2102-2112.

Gacek, K., Bayer, P.E., Bartkowiak-Broda, I., Szala, L., Bocianowski, J., Edwards, D. and Batley, J. (2016) Genome-Wide Association Study of Genetic Control of Seed Fatty Acid Biosynthesis in Brassica napus. *Front Plant Sci* **7**, 2062.

Gatti, D.M., Svenson, K.L., Shabalin, A., Wu, L.Y., Valdar, W., Simecek, P., Goodwin, N., Cheng, R., Pomp, D., Palmer, A., Chesler, E.J., Broman, K.W. and Churchill, G.A. (2014) Quantitative trait locus mapping methods for diversity outbred mice. *G3 (Bethesda)* **4**, 1623-1633.

Genin, E. (2020) Missing heritability of complex diseases: case solved? *Hum Genet* **139**, 103-113.

Guan, M., Huang, X., Xiao, Z., Jia, L., Wang, S., Zhu, M., Qiao, C., Wei, L., Xu, X., Liang, Y., Wang, R., Lu, K., Li, J. and Qu, C. (2018) Association Mapping Analysis of Fatty Acid Content in Different Ecotypic Rapeseed Using mrMLM. *Front Plant Sci* **9**, 1872.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P. and Bancroft, I. (2012) Associative transcriptomics of traits in the polyploid crop species Brassica napus. *Nat Biotechnol* **30**, 798-802.

Hatzig, S., Breuer, F., Nesi, N., Ducournau, S., Wagner, M.H., Leckband, G., Abbadi, A. and Snowdon, R.J. (2018) Hidden Effects of Seed Quality Breeding on Germination in Oilseed Rape (Brassica napus L.). *Front Plant Sci* **9**, 419.

Hatzig, S.V., Frisch, M., Breuer, F., Nesi, N., Ducournau, S., Wagner, M.H., Leckband, G., Abbadi, A. and Snowdon, R.J. (2015) Genome-wide association mapping unravels the genetic control of seed germination and vigor in Brassica napus. *Front Plant Sci* **6**, 221.

He, Y., Hu, D., You, J., Wu, D., Cui, Y., Dong, H., Li, J. and Qian, W. (2019) Genome-wide association study and protein network analysis for understanding candidate genes involved in root development at the rapeseed seedling stage. *Plant Physiol Biochem* **137**, 42-52.

Honsdorf, N., Becker, H.C and Ecke, W. (2010) Association mapping for phenological, morphological, and quality traits in canola quality winter rapeseed (Brassica napus L.). *Genome* 53, 899-907.

Hu, J., Guo, C., Wang, B., Ye, J., Liu, M., Wu, Z., Xiao, Y., Zhang, Q., Li, H., King, G.J. and Liu, K. (2018) Genetic Properties of a Nested Association Mapping Population Constructed With Semi-Winter and Spring Oilseed Rapes. *Front Plant Sci* **9**, 1740.

Huang, L., Min, Y., Schiessl, S., Xiong, X., Jan, H.U., He, X., Qian, W., Guan, C., Snowdon, R.J., Hua, W., Guan, M. and Qian, L. (2021) Integrative analysis of GWAS and transcriptome to reveal novel loci regulation flowering time in semi-winter rapeseed. *Plant Sci* **310**, 110980.

Jan, H.U., Guan, M., Yao, M., Liu, W., Wei, D., Abbadi, A., Zheng, M., He, X., Chen, H., Guan, C., Nichols, R.A., Snowdon, R.J., Hua, W. and Qian, L. (2019) Genome-wide haplotype analysis improves trait predictions in Brassica napus hybrids. *Plant Sci* **283**, 157-164.

Jin, S., Zhang, S., Liu, Y., Jiang, Y., Wang, Y., Li, J. and Ni, Y. (2020) A combination of genome-wide association study and transcriptome analysis in leaf epidermis identifies candidate genes involved in cuticular wax biosynthesis in Brassica napus. *BMC Plant Biol* **20**, 458.

Jorge, T.F., Rodrigues, J.A., Caldana, C., Schmidt, R., van Dongen, J.T., Thomas-Oates, J. and Antonio, C. (2016) Mass spectrometry-based plant metabolomics: Metabolite responses to abiotic stress. *Mass Spectrom Rev* **35**, 620-649.

Khanzada, H., Wassan, G.M., He, H., Mason, A.S., Keerio, A.A., Khanzada, S., Faheem, M., Solangi, A.M., Zhou, Q., Fu, D., Huang, Y. and Rasheed, A. (2020) Differentially evolved drought stress indices determine the genetic variation of Brassica napus at seedling traits by genome-wide association mapping. *Journal of Advanced Research* **24**, 447-461.

Kittipol, V., He, Z., Wang, L., Doheny-Adams, T., Langer, S. and Bancroft, I. (2019) Genetic architecture of glucosinolate variation in Brassica napus. *J Plant Physiol* **240**, 152988.

Klein, R., Zeiss, C., Chew, E.Y., Tsai, J.Y., Sackler, R.S., Haynes, C., Henning, A.K., SanGiovanni, J.P., Mane, S.M., Mayne, S.T., Bracken, M.B., Ferris, F.L., Ott, J., Barnstable, C., Hoh., J. (2005) Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science* **308**, 385-389

Korber, N., Bus, A., Li, J., Parkin, I.A., Wittkop, B., Snowdon, R.J. and Stich, B. (2016) Agronomic and Seed Quality Traits Dissected by Genome-Wide Association Mapping in Brassica napus. *Front Plant Sci* **7**, 386.

Korber, N., Wittkop, B., Bus, A., Friedt, W., Snowdon, R.J. and Stich, B. (2012) Seedling development in a Brassica napus diversity set and its relationship to agronomic performance. *Theor Appl*

*Genet* **125**, 1275-1287.

Kumar, V., Paillard, S., Fopa-Fomeju, B., Falentin, C., Deniot, G., Baron, C., Vallee, P., Manzanares-Dauleux, M.J. and Delourme, R. (2018) Multi-year linkage and association mapping confirm the high number of genomic regions involved in oilseed rape quantitative resistance to blackleg. *Theor Appl Genet* **131**, 1627-1643.

Laperche, A., Aigu, Y., Jubault, M., Ollier, M., Guichard, S., Glory, P., Strelkov, S.E., Gravot, A. and Manzanares-Dauleux, M.J. (2017) Clubroot resistance QTL are modulated by nitrogen input in Brassica napus. *Theor Appl Genet* **130**, 669-684.

Larsson, S.J., Lipka, A.E. and Buckler, E.S. (2013) Lessons from Dwarf8 on the strengths and weaknesses of structured association mapping. *PLoS Genet* **9**, e1003246.

Li, F., Chen, B., Xu, K., Gao, G., Yan, G., Qiao, J., Li, J., Li, H., Li, L., Xiao, X., Zhang, T., Nishio, T. and Wu, X. (2016a) A genome-wide association study of plant height and primary branch number in rapeseed (Brassica napus). *Plant Sci* **242**, 169-177.

Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., Harper, A.L., Trick, M., Liu, S., Gao, G., Wang, N., Yan, G., Qiao, J., Li, J., Li, H., Xiao, X., Zhang, T. and Wu, X. (2014) Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (Brassica napus L.). *DNA Res* **21**, 355-367.

Li, H., Cheng, X., Zhang, L., Hu, J., Zhang, F., Chen, B., Xu, K., Gao, G., Li, H., Li, L., Huang, Q., Li, Z., Yan, G. and Wu, X. (2018) An Integration of Genome-Wide Association Study and Gene Co-expression Network Analysis Identifies Candidate Genes of Stem Lodging-Related Traits in Brassica napus. *Front Plant Sci* **9**, 796.

Li, K., Wang, J., Kuang, L., Tian, Z., Wang, X., Dun, X., Tu, J. and Wang, H. (2021) Genome-wide association study and transcriptome analysis reveal key genes affecting root growth dynamics in rapeseed. *Biotechnol Biofuels* **14**, 178.

Li, L., Luo, Y., Chen, B., Xu, K., Zhang, F., Li, H., Huang, Q., Xiao, X., Zhang, T., Hu, J., Li, F. and Wu, X. (2016b) A Genome-Wide Association Study Reveals New Loci for Resistance to Clubroot Disease in Brassica napus. *Front Plant Sci* **7**, 1483.

Li, S., Zhu, Y., Varshney, R.K., Zhan, J., Zheng, X., Shi, J., Wang, X., Liu, G. and Wang, H. (2020a) A systematic dissection of the mechanisms underlying the natural variation of silique number in rapeseed (Brassica napus L.) germplasm. *Plant Biotechnol J* **18**, 568-580.

Li, Z., Liu, X., Xu, X., Liu, J., Sang, Z., Yu, K., Yang, Y., Dai, W., Jin, X. and Xu, Y. (2020b) Favorable haplotypes and associated genes for flowering time and photoperiod sensitivity identified by comparative selective signature analysis and GWAS in temperate and tropical maize. *The Crop Journal* **8**, 227-242.

Lin, H., Li, X., Guo, Z., Lv, Y., Cen, X., Ding, X., Wu, H., Li, X., Huang, J. and Xiong, L. (2017) Genetic control of the root system in rice under normal and drought stress conditions by genome-wide association study. *PLOS Genetics* **13**.

Liu, H., Li, X., Zhang, Q., Yuan, P., Liu, L., King, G.J., Ding, G., Wang, S., Cai, H., Wang, C., Xu, F. and Shi, L. (2021a) Integrating a genome-wide association study with transcriptomic data to predict candidate genes and favourable haplotypes influencing Brassica napus seed phytate. *DNA Res* **28**.

Liu, H., Wang, J., Zhang, B., Yang, X., Hammond, J.P., Ding, G., Wang, S., Cai, H., Wang, C., Xu, F. and Shi, L. (2021b) Genomewide association study dissects the genetic control of plant height and branch number in response to lowphosphorus stress in Brassica napus. *Ann Bot*.

Liu, J., Wang, W., Mei, D., Wang, H., Fu, L., Liu, D., Li, Y. and Hu, Q. (2016a) Characterizing Variation of Branch Angle and Genome-Wide Association Mapping in Rapeseed (Brassica napus L.). *Front Plant Sci* **7**, 21.

Liu, S., Fan, C., Li, J., Cai, G., Yang, Q., Wu, J., Yi, X., Zhang, C. and Zhou, Y. (2016b) A genome-wide association study reveals novel elite allelic variations in seed oil content of Brassica napus. *Theor Appl Genet* **129**, 1203-1215.

Liu, S., Huang, H., Yi, X., Zhang, Y., Yang, Q., Zhang, C., Fan, C. and Zhou, Y. (2020) Dissection of genetic architecture for glucosinolate accumulations in leaves and seeds of Brassica napus by genome-wide association study. *Plant Biotechnol J* **18**, 1472-1484.

Lu, G., Harper, A.L., Trick, M., Morgan, C., Fraser, F., O'Neill, C. and Bancroft, I. (2014) Associative transcriptomics study dissects the genetic architecture of seed glucosinolate content in Brassica napus. *DNA Res* **21**, 613-625.

Lu, K., Peng, L., Zhang, C., Lu, J., Yang, B., Xiao, Z., Liang, Y., Xu, X., Qu, C., Zhang, K., Liu, L., Zhu, Q., Fu, M., Yuan, X. and Li, J. (2017) Genome-Wide Association and Transcriptome Analyses Reveal Candidate Genes Underlying Yield-determining Traits in Brassica napus. *Front Plant Sci* **8**, 206.

Lu, K., Wei, L., Li, X., Wang, Y., Wu, J., Liu, M., Zhang, C., Chen, Z., Xiao, Z., Jian, H., Cheng, F., Zhang, K., Du, H., Cheng, X., Qu, C., Qian, W., Liu, L., Wang, R., Zou, Q., Ying, J., Xu, X., Mei, J., Liang, Y., Chai, Y.R., Tang, Z., Wan, H., Ni, Y., He, Y., Lin, N., Fan, Y., Sun, W., Li, N.N., Zhou, G., Zheng, H., Wang, X., Paterson, A.H. and Li, J. (2019) Whole-genome resequencing reveals Brassica napus origin and genetic loci involved in its improvement. *Nat Commun* **10**, 1154.

Lu, K., Xiao, Z., Jian, H., Peng, L., Qu, C., Fu, M., He, B., Tie, L., Liang, Y., Xu, X. and Li, J. (2016) A combination of genome-wide association and transcriptome analysis reveals candidate genes controlling harvest index-related traits in Brassica napus. *Sci Rep* **6**, 36452.

Luo, T., Zhang, Y., Zhang, C., Nelson, M.N., Yuan, J., Guo, L. and Xu, Z. (2021) Genome-Wide Association Mapping Unravels the Genetic Control of Seed Vigor under Low-Temperature Conditions in Rapeseed (Brassica napus L.). *Plants (Basel)* **10**.

Luo, X., Ding, Y., Zhang, L., Yue, Y., Snyder, J.H., Ma, C. and Zhu, J. (2017a) Genomic Prediction of Genotypic Effects with Epistasis and Environment Interactions for Yield-Related Traits of Rapeseed (Brassica napus L.). *Front Genet* **8**, 15.

Luo, X., Xue, Z., Ma, C., Hu, K., Zeng, Z., Dou, S., Tu, J., Shen, J., Yi, B. and Fu, T. (2017b) Joint genome-wide association and transcriptome sequencing reveals a complex polygenic network underlying hypocotyl elongation in rapeseed (Brassica napus L.). *Sci Rep* **7**, 41561.

Manolio, T.A. (2013) Bringing genome-wide association findings into clinical use. *Nat Rev Genet* **14**, 549-558.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A. and Visscher, P.M. (2009) Finding the missing heritability of complex diseases. *Nature* **461**, 747-753.

Menendez, Y.C., Sanchez, D.H., Snowdon, R.J., Rondanini, D.P. and Botto, J.F. (2021) Unraveling the impact on agronomic traits of the genetic architecture underlying plant-density responses in canola. *J Exp Bot* **72**, 5426-5441.

Nordborg, M., Hu, T.T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., Jakobsson, M., Kim, S., Morozov, Y., Padhukasahasram, B., Plagnol, V., Rosenberg, N.A., Shah, C., Wall, J.D., Wang, J., Zhao, K., Kalbfleisch, T., Schulz, V., Kreitman, M. and Bergelson, J. (2005) The pattern of polymorphism in Arabidopsis thaliana. *PLoS Biol* **3**, e196.

Qian, M., Fan, Y., Li, Y., Liu, M., Sun, W., Duan, H., Yu, M., Chang, W., Niu, Y., Li, X., Liang, Y., Qu, C., Li, J. and Lu, K. (2021) Genome-wide association study and transcriptome comparison reveal novel QTL and candidate genes that control petal size in rapeseed. *J Exp Bot* **72**, 3597-3610.

Qu, C., Jia, L., Fu, F., Zhao, H., Lu, K., Wei, L., Xu, X., Liang, Y., Li, S., Wang, R. and Li, J. (2017) Genome-wide association mapping and Identification of candidate genes for fatty acid composition in Brassica napus L. using SNP markers. *BMC Genomics* **18**, 232.

Raman, H., McVittie, B., Pirathiban, R., Raman, R., Zhang, Y., Barbulescu, D.M., Qiu, Y., Liu, S. and Cullis, B. (2020) Genome-Wide Association Mapping Identifies Novel Loci for Quantitative Resistance to Blackleg Disease in Canola. *Front Plant Sci* **11**, 1184.

Raman, H., Raman, R., Coombes, N., Song, J., Diffey, S., Kilian, A., Lindbeck, K., Barbulescu, D.M., Batley, J., Edwards, D., Salisbury, P.A. and Marcroft, S. (2016) Genome-wide Association Study Identifies New Loci for Resistance to Leptosphaeria maculans in Canola. *Front Plant Sci* **7**, 1513.

Raman, H., Raman, R., Qiu, Y., Yadav, A.S., Sureshkumar, S., Borg, L., Rohan, M., Wheeler, D., Owen, O., Menz, I. and Balasubramanian, S. (2019) GWAS hints at pleiotropic roles for FLOWERING LOCUS T in flowering time and yield-related traits in canola. *BMC Genomics* **20**, 636.

Schiessl, S., Iniguez-Luy, F., Qian, W. and Snowdon, R.J. (2015) Diverse regulatory factors associate with flowering time and yield responses in winter-type Brassica napus. *BMC Genomics* **16**, 737.

Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., Zhou, T., Lu, T., Zhu, J., Shangguan, Y., Chen, E., Gong, C., Zhao, Q., Jing, Y., Zhao, Y., Li, Y., Cui, L., Fan, D., Lu, Y., Weng, Q., Wang, Y., Zhan, Q., Liu, K., Wei, X., An, K., An, G. and Han, B. (2016) OsSPL13 controls grain size in cultivated rice. *Nat Genet* **48**, 447-456.

Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D., Wang, B., Lu, S., Zhou, R., Xie, W.Z., Cheng, Y., Zhang, Y., Liu, K., Yang, Q.Y., Chen, L.L. and Guo, L. (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of Brassica napus. *Nat Plants* **6**, 34-45.

Sun, C., Wang, B., Wang, X., Hu, K., Li, K., Li, Z., Li, S., Yan, L., Guan, C., Zhang, J., Zhang, Z., Chen, S., Wen, J., Tu, J., Shen, J., Fu, T. and Yi, B. (2016a) Genome-Wide Association Study Dissecting the Genetic Architecture Underlying the Branch Angle Trait in Rapeseed (Brassica napus L.). *Sci Rep* **6**, 33673.

Sun, C., Wang, B., Yan, L., Hu, K., Liu, S., Zhou, Y., Guan, C., Zhang, Z., Li, J., Zhang, J., Chen, S., Wen, J., Ma, C., Tu, J., Shen, J., Fu, T. and Yi, B. (2016b) Genome-Wide Association Study Provides Insight into the Genetic Control of Plant Height in Rapeseed (Brassica napus L.). *Front Plant Sci* **7**, 1102.

Sun, F., Liu, J., Hua, W., Sun, X., Wang, X. and Wang, H. (2016c) Identification of stable QTLs for seed oil content by combined linkage and association mapping in Brassica napus. *Plant Sci* **252**, 388-399.

Tam, V., Patel, N., Turcotte, M., Bosse, Y., Pare, G. and Meyre, D. (2019) Benefits and limitations of genome-wide association studies. *Nat Rev Genet* **20**, 467-484.

Tan, Z., Xie, Z., Dai, L., Zhang, Y., Hu, Z., Tang, S., Wan, L., Yao, X., Guo, L. and Hong, D. (2021) Genome- and transcriptome-wide association studies reveal the genetic basis and the breeding history of seed glucosinolate content in Brassica napus. *Plant Biotechnol J*.

Tang, S., Zhao, H., Lu, S., Yu, L., Zhang, G., Zhang, Y., Yang, Q.Y., Zhou, Y., Wang, X., Ma, W., Xie, W. and Guo, L. (2021) Genome- and transcriptome-wide association studies provide insights into the genetic basis of natural variation of seed oil content in Brassica napus. *Mol Plant*.

USDA ERS (2020) https://www.ers.usda.gov/data-products/oil-crops-yearbook/.

Wan, H., Chen, L., Guo, J., Li, Q., Wen, J., Yi, B., Ma, C., Tu, J., Fu, T. and Shen, J. (2017) Genome-Wide Association Study Reveals the Genetic Architecture Underlying Salt Tolerance-Related Traits in Rapeseed (Brassica napus L.). *Front Plant Sci* **8**, 593.

Wang, B., Wu, Z., Li, Z., Zhang, Q., Hu, J., Xiao, Y., Cai, D., Wu, J., King, G.J., Li, H. and Liu, K. (2018) Dissection of the genetic architecture of three seed-quality traits and consequences for breeding in Brassica napus. *Plant Biotechnol J* **16**, 1336-1348.

Wang, H., Wang, Q., Pak, H., Yan, T., Chen, M., Chen, X., Wu, D. and Jiang, L. (2021) Genome-wide association study reveals a patatin-like lipase relating to the reduction of seed oil content in Brassica napus. *BMC Plant Biol* **21**, 6.

Wang, J., Jian, H., Wei, L., Qu, C., Xu, X., Lu, K., Qian, W., Li, J., Li, M. and Liu, L. (2015) Genome-Wide Analysis of Seed Acid Detergent Lignin (ADL) and Hull Content in Rapeseed (Brassica napus L.). *PLoS One* **10**, e0145045.

Wang, J., Xian, X., Xu, X., Qu, C., Lu, K., Li, J. and Liu, L. (2017a) Genome-Wide Association Mapping of Seed Coat Color in Brassica napus. *J Agric Food Chem* **65**, 5229-5237.

Wang, T., Wei, L., Wang, J., Xie, L., Li, Y.Y., Ran, S., Ren, L., Lu, K., Li, J., Timko, M.P. and Liu, L. (2020) Integrating GWAS, linkage mapping and gene expression analyses reveals the genetic control of growth period traits in rapeseed (Brassica napus L.). *Biotechnol Biofuels* **13**, 134.

Wang, X., Chen, Y., Thomas, C.L., Ding, G., Xu, P., Shi, D., Grandke, F., Jin, K., Cai, H., Xu, F., Yi, B., Broadley, M.R. and Shi, L. (2017b) Genetic variants associated with the root system architecture of oilseed rape (Brassica napus L.) under contrasting phosphate supply. *DNA Res* **24**, 407-417.

Wassan, G.M., Khanzada, H., Zhou, Q., Mason, A.S., Keerio, A.A., Khanzada, S., Solangi, A.M., Faheem, M., Fu, D. and He, H. (2021) Identification of genetic variation for salt tolerance in Brassica napus using genome-wide association mapping. *Mol Genet Genomics* **296**, 391-408.

Wei, D., Cui, Y., Mei, J., Qian, L., Lu, K., Wang, Z.M., Li, J., Tang, Q. and Qian, W. (2019) Genome-wide identification of loci affecting seed glucosinolate contents in Brassica napus L. *J Integr Plant Biol* **61**, 611-623.

Wei, L., Jian, H., Lu, K., Filardo, F., Yin, N., Liu, L., Qu, C., Li, W., Du, H. and Li, J. (2016) Genome-wide association analysis and differential expression analysis of resistance to Sclerotinia stem rot in Brassica napus. *Plant Biotechnol J* **14**, 1368-1380.

Wei, L., Jian, H., Lu, K., Yin, N., Wang, J., Duan, X., Li, W., Liu, L., Xu, X., Wang, R., Paterson, A.H. and Li, J. (2017) Genetic and transcriptomic analyses of lignin- and lodging-related traits in Brassica napus. *Theor Appl Genet* **130**, 1961-1973.

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. *Nat Commun* **5**, 3438.

Werner, C.R., Voss-Fels, K.P., Miller, C.N., Qian, W., Hua, W., Guan, C.Y., Snowdon, R.J. and Qian, L.

(2018) Effective Genomic Selection in a Narrow-Genepool Crop with Low-Density Markers: Asian Rapeseed as an Example. *Plant Genome* **11**.

Wood, I.P., Pearson, B.M., Garcia-Gutierrez, E., Havlickova, L., He, Z., Harper, A.L., Bancroft, I. and Waldron, K.W. (2017) Carbohydrate microarrays and their use for the identification of molecular markers for plant cell wall composition. *Proc Natl Acad Sci U S A* **114**, 6860-6865.

Wu, D., Liang, Z., Yan, T., Xu, Y., Xuan, L., Tang, J., Zhou, G., Lohwasser, U., Hua, S., Wang, H., Chen, X., Wang, Q., Zhu, L., Maodzeka, A., Hussain, N., Li, Z., Li, X., Shamsi, I.H., Jilani, G., Wu, L., Zheng, H., Zhang, G., Chalhoub, B., Shen, L., Yu, H. and Jiang, L. (2019) Whole-Genome Resequencing of a Worldwide Collection of Rapeseed Accessions Reveals the Genetic Basis of Ecotype Divergence. *Mol Plant* **12**, 30-43.

Wu, J., Zhao, Q., Liu, S., Shahid, M., Lan, L., Cai, G., Zhang, C., Fan, C., Wang, Y. and Zhou, Y. (2016a) Genome-wide Association Study Identifies New Loci for Resistance to Sclerotinia Stem Rot in Brassica napus. *Front Plant Sci* **7**, 1418.

Wu, Z., Wang, B., Chen, X., Wu, J., King, G.J., Xiao, Y. and Liu, K. (2016b) Evaluation of Linkage Disequilibrium Pattern and Association Study on Seed Oil Content in Brassica napus Using ddRAD Sequencing. *PLoS One* **11**, e0146383.

Xiao, Y., Liu, H., Wu, L., Warburton, M. and Yan, J. (2017) Genome-wide Association Studies in Maize: Praise and Stargaze. *Mol Plant* **10**, 359-374.

Xiao, Z., Tang, F., Zhang, L., Li, S., Wang, S., Huo, Q., Yang, B., Zhang, C., Wang, D., Li, Q., Wei, L., Guo, T., Qu, C., Lu, K., Zhang, Y., Guo, L., Li, J. and Li, N. (2021) The Brassica napus fatty acid exporter FAX1-1 contributes to biological yield, seed oil content, and oil quality. *Biotechnol Biofuels* **14**, 190.

Xiao, Z., Zhang, C., Tang, F., Yang, B., Zhang, L., Liu, J., Huo, Q., Wang, S., Li, S., Wei, L., Du, H., Qu, C., Lu, K., Li, J. and Li, N. (2019) Identification of candidate genes controlling oil content by combination of genome-wide association and transcriptome analysis in the oilseed crop Brassica napus. *Biotechnol Biofuels* **12**, 216.

Xu, L., Hu, K., Zhang, Z., Guan, C., Chen, S., Hua, W., Li, J., Wen, J., Yi, B., Shen, J., Ma, C., Tu, J. and Fu, T. (2016) Genome-wide association study reveals the genetic architecture of flowering time in rapeseed (Brassica napus L.). *DNA Res* **23**, 43-52.

Xuan, L., Yan, T., Lu, L., Zhao, X., Wu, D., Hua, S. and Jiang, L. (2020) Genome-wide association study reveals new genes involved in leaf trichome formation in polyploid oilseed rape (Brassica napus L.). *Plant Cell Environ* **43**, 675-691.

Yao, M., Guan, M., Yang, Q., Huang, L., Xiong, X., Jan, H.U., Voss-Fels, K.P., Werner, C.R., He, X., Qian, W., Snowdon, R.J., Guan, C., Hua, W. and Qian, L. (2021) Regional association analysis coupled with transcriptome analyses reveal candidate genes affecting seed oil accumulation in Brassica napus. *Theor Appl Genet* **134**, 1545-1555.

Yao, M., Guan, M., Zhang, Z., Zhang, Q., Cui, Y., Chen, H., Liu, W., Jan, H.U., Voss-Fels, K.P., Werner, C.R., He, X., Liu, Z., Guan, C., Snowdon, R.J., Hua, W. and Qian, L. (2020) GWAS and co-expression network combination uncovers multigenes with close linkage effects on the oleic acid content accumulation in Brassica napus. *BMC Genomics* **21**, 320.

Yang, M., Lu, K., Zhao, F. J., et al. (2018) Genome-wide association studies reveal the genetic basis of ionomic variation in rice. *Plant Cell* **30**, 2720-2740.

Yong, H.Y., Wang, C., Bancroft, I., Li, F., Wu, X., Kitashiba, H. and Nishio, T. (2015) Identification of a gene controlling variation in the salt tolerance of rapeseed (Brassica napus L.). *Planta* **242**, 313-

326.

Yu, J. and Buckler, E.S. (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* **17**, 155-160.

Zhang, F., Xiao, X., Xu, K., Cheng, X., Xie, T., Hu, J. and Wu, X. (2020) Genome-wide association study (GWAS) reveals genetic loci of lead (Pb) tolerance during seedling establishment in rapeseed (Brassica napus L.). *BMC Genomics* **21**, 139.

Zhang, J., Mason, A.S., Wu, J., Liu, S., Zhang, X., Luo, T., Redden, R., Batley, J., Hu, L. and Yan, G. (2015) Identification of Putative Candidate Genes for Water Stress Tolerance in Canola (Brassica napus). *Front Plant Sci* **6**, 1058.

Zheng, M., Peng, C., Liu, H., Tang, M., Yang, H., Li, X., Liu, J., Sun, X., Wang, X., Xu, J., Hua, W. and Wang, H. (2017) Genome-Wide Association Study Reveals Candidate Genes for Control of Plant Height, Branch Initiation Height and Branch Number in Rapeseed (Brassica napus L.). *Front Plant Sci* **8**, 1246.

Zhou, Q., Han, D., Mason, A.S., Zhou, C., Zheng, W., Li, Y., Wu, C., Fu, D. and Huang, Y. (2018) Earliness traits in rapeseed (Brassica napus): SNP loci and candidate genes identified by genome-wide association analysis. *DNA Res* **25**, 229-244.

Zhu, Q., King, G.J., Liu, X., Shan, N., Borpatragohain, P., Baten, A., Wang, P., Luo, S. and Zhou, Q. (2019) Identification of SNP loci and candidate genes related to four important fatty acid composition in Brassica napus using genome wide association study. *PLoS One* **14**, e0221578

Table 1. GWAS of important agronomic traits in *B. napus* and the candidate genes

(Please see Suppl. file)

Table 2. Benefits and limitations using SNP chip and WGS in GWAS analysis of *B. napus*.

| Factors | SNP chip | WGS |
|---|---|---|
| Cost | Relatively inexpensive | Expensive (>US$200 per sample) |
| Reliability | Reliable, very comprehensive solutions | Less comprehensive solution and less accurate technology |
| Genomic coverage | About tens of thousands of SNP markers, and the location of SNP markers is fixed, new mutations, and rare mutations and indel markers are not included | Covering the whole genome, from SNP markers, rare variation to almost all genetic variation in the genome |
| GWAS analysis | Relatively simple, with comprehensive solutions, and can be completed in Windows system | Relatively difficult; requires Linux system and R language |
| Computational costs | Relatively low; personal computers can meet the needs of computing | Greater costs to store, process, analyse and interpret the resulting data |
| Research objectives | (1) New traits that have not been studied; (2) Detection of common-variant associations in a large sample size | (1) Traits that have been studied by SNP chip; (2) rare and structure variants |

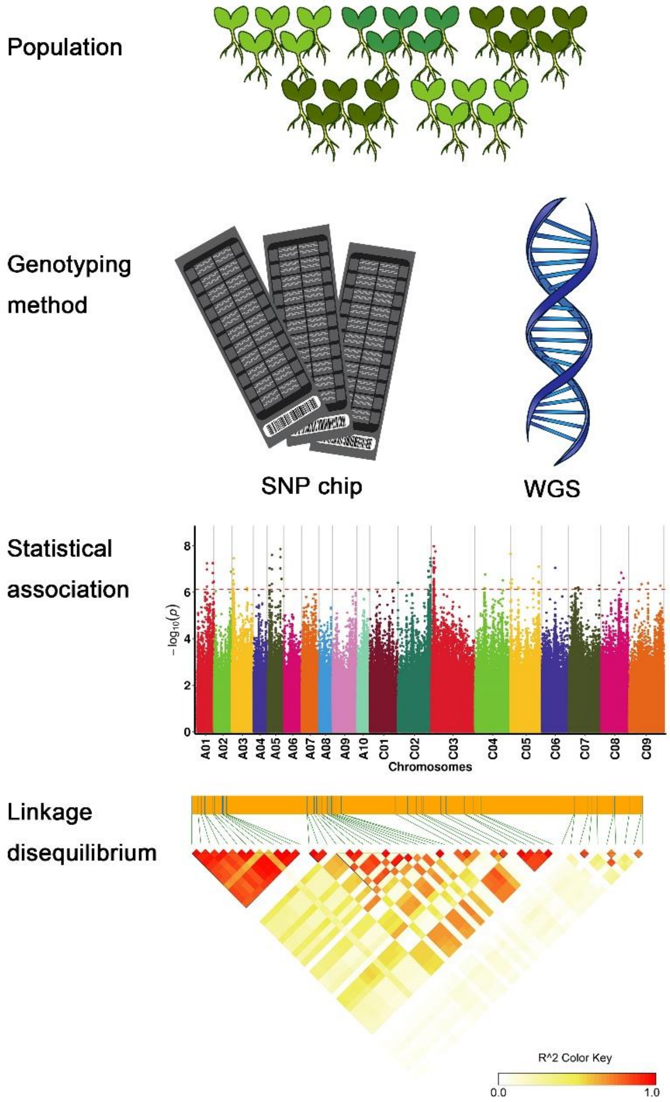

**Fig 1**. **GWAS study design in *B. napus***. The aim of GWAS is to detect associations between genotype frequency and trait status. The first step of GWAS involves in collecting natural populations and identifying the status of the studied traits. Genotyping can be performed using SNP-chip or WGS. Association analysis is used to identify regions of the genome associated with the target traits at genome- wide. Candidate genes are usually not directly identified but are in linkage disequilibrium with the significant SNPs. GWAS, genome- wide association study; WGS, whole-genome sequencing; SNP, single nucleotide polymorphism.

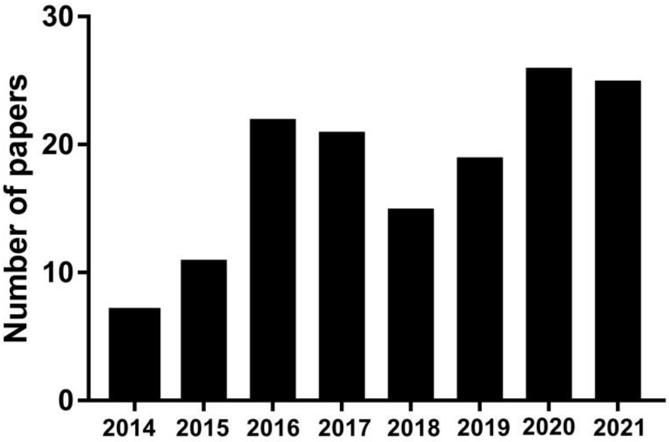

**Fig 2. The increase of publications on the GWAS of *B. napus* since the release of the reference genome of *B. napus* cultivar *Darmor-bzh* in 2014.** The data are from NCBI pubmed (https://pubmed.ncbi.nlm.nih.gov/). The key words of "genome-wide association study" and "*Brassica napus*" were searched in September 30, 2021.

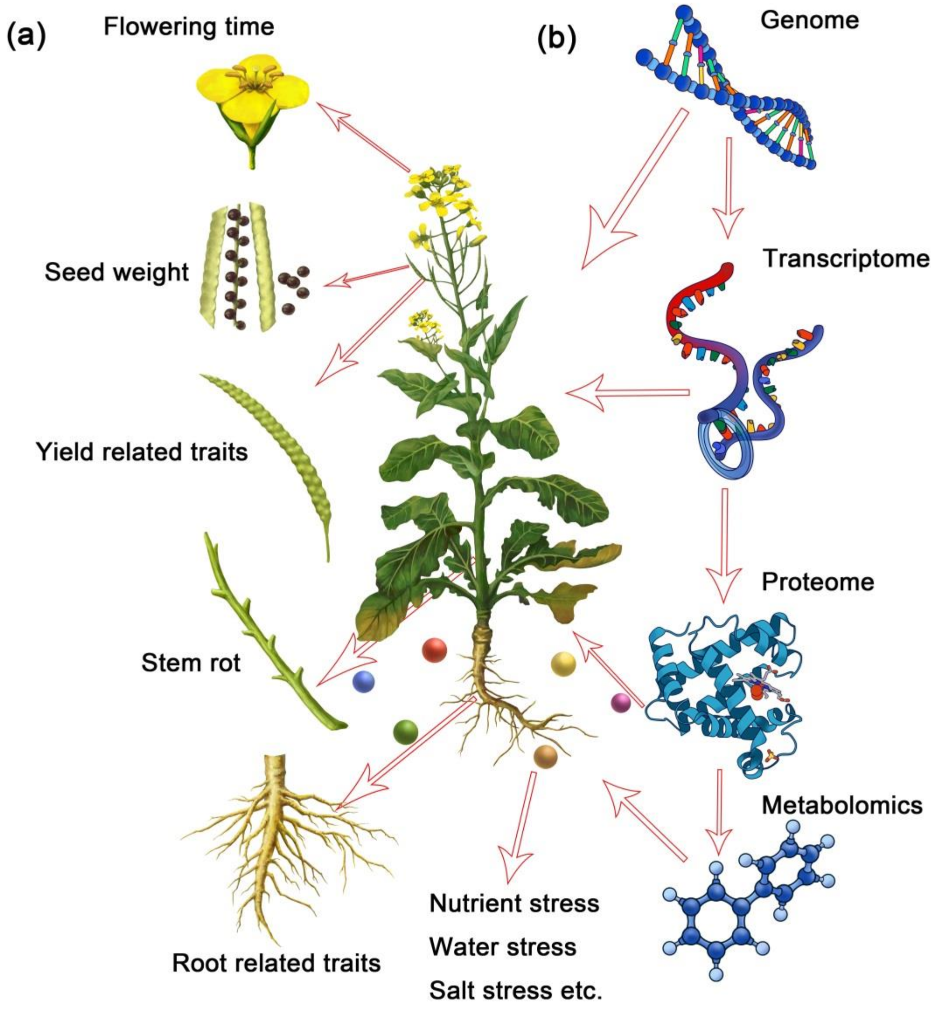

**Fig 3. The traits of *B. napus* used in the GWAS.** (a) The important agronomic traits today. (b) The omics traits tomorrow.

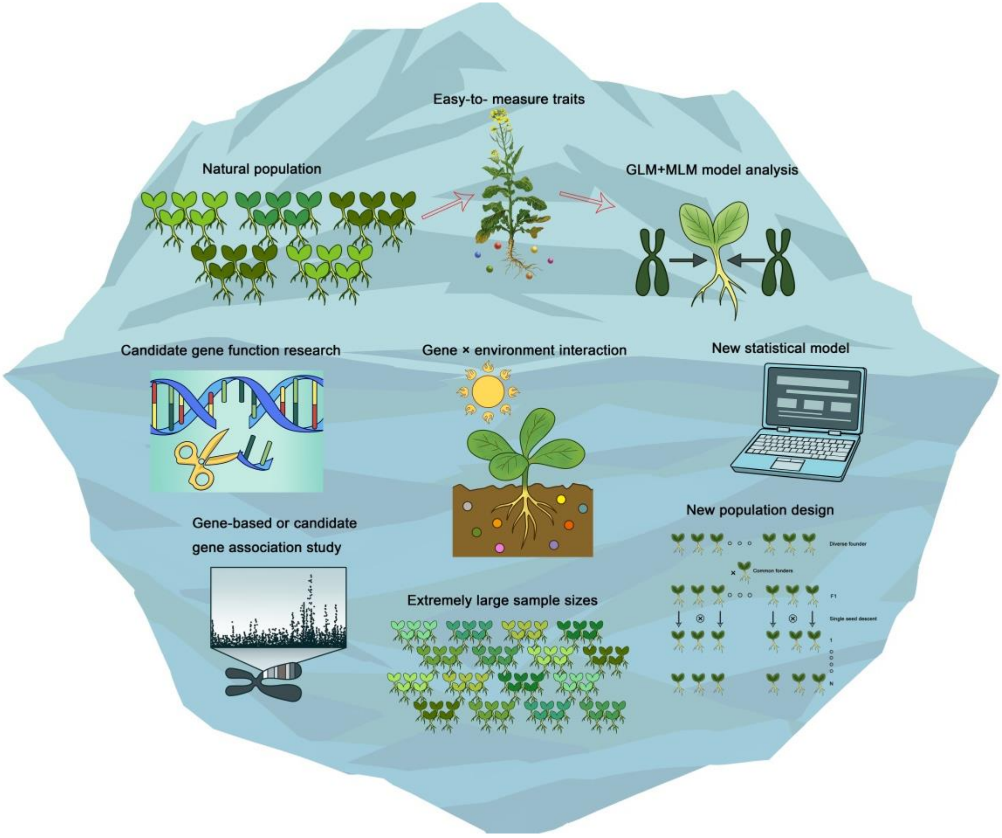

**Fig 4. A tip of the iceberg - current GWAS in *B. napus***. The discoveries that can be made in *B. napus* using GWAS are represented by an iceberg. The portion of the iceberg above water represents the discoveries that have been made in *B. napus* by GWAS to date, using easy- to-measure traits (for example seed yield, seed weight and flowering time, etc.), predominantly natural populations; and GLM and MLM models. Most of the iceberg is submerged. The submerged part includes the main exploration directions in the future, including the interaction between gene and environment, larger population sizes, new statistical model, new population design, gene function study, and molecular marker assisted breeding.