

Genome-Wide Association Studies Dissect the G×E Interaction for Agronomic Traits in a Worldwide Collection of Safflower (*Carthamus Tinctorius* L.)

Huanhuan Zhao (✉ huan.zhao@agriculture.vic.gov.au)

Agriculture Victoria

Keith W. Savin

Agriculture Victoria

Yongjun Li

Agriculture Victoria

Edmond J Breen

Agriculture Victoria

Pankaj Maharjan

Grains Innovation Park

Josquin F. Tibbits

Agriculture Victoria

Surya Kant

Grains Innovation Park

Matthew J. Hayden

Agriculture Victoria

Hans D. Daetwyler

Agriculture Victoria

Research Article

Keywords: Safflower, agronomic traits, G × E interaction, GWAS, mixed linear model

Posted Date: October 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-926995/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Safflower (*Carthamus tinctorius L.*) has been cultivated worldwide for centuries, originally as a source of textile dyes. Modern safflower breeding has focused on high grain and oil yield and broad adaptability. Here, a genome-wide association study was conducted using a globally diverse Genebank collection of 406 accessions, which included landraces, breeding lines and elite cultivars. We explored the genetic architecture and genotype-by-environment interaction ($G \times E$) patterns of grain yield (YP), days to flowering (DF), plant height (PH), 500 seed weight (SW), seed oil content (OL), and crude protein content (PR) in four environments (sites) that differed in water availability.

Results: Phenotypic variation within the global collection was observed for all traits under differed water stress environments. Two mixed linear models were adopted, and YP exhibited low overall genetic correlations ($r_{G_{\text{overall}}}$) across sites, while SW and OL had high $r_{G_{\text{overall}}}$ and high pairwise genetic correlations ($r_{G_{ij}}$) across all pairwise sites. Ninety-two marker-trait associations (MTAs) were identified using three methods, single locus genome-wide association studies (GWAS) using a mixed linear model (MLM), the Bayesian multi-locus method (BayesR), and meta-GWAS. MTAs with large effects across all sites were detected for OL, SW, and PR, and MTAs specific for the different water stress sites was identified for all traits. Five MTAs were associated with multiple traits, 4 of 5 MTAs were variously associated with the three traits of SW, OL, and PR and marker effects were consistent with phenotypic observations in different environments. The thresholds of different GWAS methods used in the study affected the number of MTAs identified for complex traits.

Conclusions: This study provided insights into the phenotypic variability and genetic architecture of important safflower agronomic traits under different environments. This knowledge is essential to breed for high grain and oil yield and local adaption in safflower.

Background

Safflower (*Carthamus tinctorius L.*) is a member of the Compositae family, grown as a vegetable, cut flower, herbal medicine, animal feed, birdseed, and oilseed, etc. in over 60 geographical regions covering the Middle East, Africa, America, Europe, and Asia [1]. In recent years, with a growing demand for healthy cooking oil and clean biofuel and bio-lubricants, safflower has emerged as a modern industrial oilseed crop due to its higher oleic and linoleic acid content compared to other oilseed crops [2, 3]. In 2019, FAO data showed safflower seed production world-wide was approximately 0.6 million tonnes, and the top 4 largest growers (Kazakhstan, United States, Russian Federation, and Mexico) produce over 75% of total production[4]. The Australian safflower growing area is currently about 40,000 ha, down from its peak of 74,688 ha in 1979 [5]. As a potential crop that could grow in a drier environment, safflower is gaining more research attention [6].

To date, genetic analyses for agronomic traits in safflower have largely been undertaken using conventional family-based methods. This has allowed the identification of genes and quantitative trait

loci (QTL) for traits such as plant height, seed oil content, days to flowering, etc. [7–9]. Association mapping approaches have also been used to identify QTL in safflower. Six marker-trait associations (MTAs) for PH and five MTAs for DF were identified in an association study using microsatellite markers [10], while another study using AFLP markers detected four MTAs for PH under drought conditions [11]. The *Fad2* gene family (Fatty acid desaturases, FAD) in safflower have been sequenced with genes being isolated and cloned [12, 13]. By using safflower core collections with SSR markers, an association study discovered several MTAs for oil content, oleic acid content and linoleic acid content [10]. However, no genome-wide association studies (GWAS) based on single nucleotide polymorphisms (SNPs) markers have been reported in safflower.

Statistical methods used in GWAS analysis are important for identifying MTAs for complex traits [14, 15]. Single-locus GWAS with mixed linear models (MLM-GWAS) has been widely used to detect the MTAs for agronomic traits in a variety of plants, including wheat [16], rapeseed [17], soybean [18], etc. To increase power to discover SNP with small effects and reduce false-positive associations, summary statistic-based methods (meta-GWAS) have been adopted in some studies [19, 20]. In canola, a meta-GWAS analysis identified 79 genomic regions conferring potential candidate resistance to canola blackleg disease, more significant SNPs than single-locus GWAS [21]. Differing from single-locus MLM-GWAS testing one marker at a time, multi-locus GWAS have been applied by fitting all loci simultaneously to improve fine-mapping [22, 23]. As a multi-locus Bayesian method, BayesR simultaneously accommodates all SNPs in the model, and SNPs effects were a mixture of four normal distributions, which include SNPs with 0, small and moderate effects. In each distribution, fewer SNPs explain the gradually more genetic variance [24, 25]. BayesR has been used to identify QTL or associations in dairy cattle and wheat [26, 27]

The variation in phenotypes among genotypes in different environments is evaluated as the extent of the genotype-by-environment interaction ($G \times E$), which is also referred to as the traits phenotypic plasticity [28]. Identifying $G \times E$ interaction patterns and their genetic basis under multi-environment trials can deepen the knowledge of the genetic architecture of traits [29, 30]. In a canola study, 12 environment-stable QTL and 43 environment-specific QTL were detected for flowering time in three different ecological conditions, which provided new insights into the genetic regulatory network underlying the control of flowering time [31]. Few studies investigating $G \times E$ interaction patterns have been reported in safflower; those studies were carried out to evaluate genotypes and yield stability [32, 33].

In Australia, crop production is challenged by spatial drought patterns due to seasonal rainfall and high temperatures [34]. Therefore, understanding the $G \times E$ interaction and genetic basis underlying grain yield and related agronomic traits are important for safflower breeding. In this study, a globally diverse Genebank collection of 406 accessions was grown in 4 different field environments (2 trials in one location but with different field management in 2017 and 2 locations in 2018). We examined the $G \times E$ patterns for grain yield (YP), plant height (PH), days to flowering (DF), 500 seed weight (SW), seed protein (PR), and seed oil content (OL). GWAS was conducted for each environment using three methods. The aims were to: 1) assess genetic variability in the different environments and the level of $G \times E$ interaction

for the agronomic traits measured; 2) identify MTAs for these traits and the genetic basis of their G × E interaction.

Results

Phenotypic variation and correlations

406 globally diverse safflower accessions were evaluated in four field trials. The phenotypic distributions and means for YP, PH, DF, SW, PR, and OL are shown in Fig. 1. Mean YP was the highest at site 1 (1.89kg/plot), a third less at site 2 (1.21 kg/plot), and halved at sites 3 and 4 (0.66 and 0.72kg/plot) (Supplementary Table s2). The distribution for YP was much narrower at sites 3 and 4 compared with sites 1 and 2 for YP (Fig. 1). PH showed a similar distribution pattern to YP, with lower means at sites 3 and 4 (~ 60cm) compared to sites 1 and 2 (~ 115cm). DF had the highest mean value and narrowest distribution at site 2 (~ 160days, Fig. 1). There were differences in trait means between sites for the three seed traits (SW, PR, and OL), but they were more subtle than those observed for YP and PH. The SW mean was higher at sites 3 and 4 (~ 20.6g) than at sites 1 and 2 (19.56g and 18.77g, respectively). The mean of PR ranged from 15.14–15.92% across four sites, and OL decreased about 1% with different water stress environments, from 31.83% at site 1 down to 29.69% at site 3. Similarly, the distributions for the three seed traits did not change dramatically across the four sites (Fig. 1, Supplementary Table s2).

Pearson correlations between traits at each site showed that YP was positively correlated with PH and negatively correlated with PR, while SW was negatively correlated with OL, PR at all sites (Fig. 2). PR was positively correlated with OL at sites 3 and 4, and DF is highly correlated with PH at sites 1, 3, and 4.

G × E interaction

G × E interactions for each trait were determined through combined site analysis. The linear mixed model (model 2) assumes homogeneous genetic and residual variances across sites, and genetic correlations ranged from 0 to 1 between sites. The model including G × E interaction effects had a higher log-likelihood than the model excluding G × E interactions for all traits (Supplementary Table s3). The significance test showed that G × E interaction effects were significant for all traits (Table 1). High overall genetic correlations (r_{Goverall}) among sites were observed for SW and OL (0.95 and 0.94, respectively), indicating low G × E interaction for those traits, while low r_{Goverall} for YP (0.48) indicated a strong G × E interaction. G × E levels were moderate for PH, DF, and PR, with r_{Goverall} value ranging from 0.67 to ~ 0.79 (Table 1).

Table 1

Variance components and genetic correlation (r_{Goverall} , with standard error in parenthesis) were estimated with the homogenous linear mixed model (model 2) for the six traits with combined environments (σ_G^2 is the genetic variance of accessions, σ_{GE}^2 is the variance for G \times E interaction).

Variance component \ Trait	YP	PH	DF	SW	OL	PR
σ_G^2 Accessions (σ_G^2)	0.057	77.500	8.413	9.114	10.11	0.341
σ_{GE}^2 ENV.Accessions (σ_{GE}^2)	0.061*	32.947*	4.230*	0.517*	0.59*	0.089*
σ_e^2 Residual (σ_e^2)	0.133	53.749	3.469	0.907	0.908	0.312
r_{Goverall}	0.48 (0.02)	0.70 (0.02)	0.67 (0.02)	0.95 (0.01)	0.94 (0.01)	0.79 (0.02)
*: G \times E interaction covariance was significant at 0.05 level.						

To account for differences in G \times E interactions across sites, a linear mixed model assuming heterogeneous genetic and residual variances (model 3) was adopted. The AIC was lower for all traits, suggesting that model 3 fitted the data better (Supplementary Table s4). The genetic correlation (r_{Gij}) between pairwise site combinations varied for all traits; however, traits with high r_{Goverall} also had high r_{Gij} between pairwise sites (Table 2). SW and OL had high pairwise genetic correlations between all sites with r_{Gij} values > 0.9 , especially between sites 3 and 4, which a r_{Gij} (0.99) not significantly different from 1. According to Robertson [35], a correlation of performance between environments ≤ 0.8 indicates a considerable re-ranking of individuals. As all traits had $r_{\text{Gij}} \geq 0.80$ between sites 3 and 4, those two sites could be treated as a single site with limited re-ranking. Genetic correlations for PH were uniformly high between pairwise sites ($r_{\text{Gij}} \geq 0.80$), while the genetic correlations for DF and YP varied, with those for YP being the most variable.

Table 2

Genetic correlations (r_{Gij} , with standard error in parenthesis) between sites for six traits were calculated with model 3.

Site pair		Trait					
ENV_i	ENV_j	YP	PH	DF	SW	OL	PR
1	2	0.69 (0.04)	0.84 (0.02)	0.6 (0.04)	0.96 (0.01)	0.94 (0.01)	0.89 (0.02)
1	3	0.45 (0.06)	0.81 (0.05)	0.85 (0.02)	0.93 (0.01)	0.95 (0.01)	0.7 (0.04)
1	4	0.68 (0.05)	0.81 (0.03)	0.85 (0.02)	0.93 (0.01)	0.95 (0.01)	0.67 (0.04)
2	3	0.39 (0.07)	0.87 (0.04)	0.35 (0.05)	0.93 (0.01)	0.9 (0.01)	0.78 (0.03)
2	4	0.46 (0.06)	0.8 (0.03)	0.36 (0.05)	0.92 (0.01)	0.91 (0.01)	0.73 (0.04)
3	4	0.82 (0.06)	0.87 (0.05)	0.99 (0)*	0.99 (0)*	0.99 (0)*	0.94 (0.02)

* Genetic correlation is not significantly differ from value of 1 at 0.05 level.

Genome-wide associations

A total of 1780 SNPs out of 1806 filtered SNPs were positioned on the 12 pseudochromosomes of the draft safflower genome assembly, with about 100–200 SNPs per pseudochromosome. LD decayed rapidly over a short physical distance, followed by a slower decline over longer pairwise distances (Supplementary Figure s1). The heatmap of the genomic relationship matrix (**G**) revealed a strong population structure among some accessions (Fig. 3), which was consistent with the previous observation [36].

Combined QQ plots for the four sites showed that the inclusion of G matrix in the GWAS effectively accounted for the observed population structure (Fig. 4, Supplementary Figure s2). A relaxed significance threshold of $-\log_{10}(p) \geq 2$ was used to denote MTAs in the MLM-GWAS, which resulted in the identification of between 41 and 71 putative MTAs for each trait across the four sites. For the meta-GWAS, the number of significant SNPs was more than twice the number found in the MLM-GWAS for OL, SW, and PR. Fewer MTAs were detected using the BayesR method, especially for PR (Table 3). SNPs with large effects in the BayesR analysis typically overlapped with SNPs above the significance threshold in the MLM-GWAS (Fig. 4, Supplementary Figure s2).

A total of 92 significant MTAs were detected by all three GWAS approaches (Table 3, Supplementary Tables s5-s10). Heatmap of pairwise LD between significant MTAs for each trait were plotted to show that MTAs were not tightly linked (Supplementary Figure s3). By comparing MLM-GWAS and BayesR results,

the significant MTAs were further classified as site-specific or shared among sites, where sites 3 and 4 were treated as a single site. Traits with low overall $G \times E$ (high genetic correlation) had a higher percentage of shared MTAs across sites for SW and OL, while for traits with high to moderate overall $G \times E$ traits (PH, DF and YP) limited shared MTA between pairwise sites were observed. 5 and 6 MTAs shared across all sites were observed for SW and OL, respectively. Some MTAs had large effects (> 0.15), such as Locus 6195 and 28935 for OL and Locus 3057 and 27064 for SW. PR had 3 shared MTAs, with 2 being shared across all sites. Locus 3057 showed large effects across all sites. 7 MTAs were shared between two sites, and 5 of 7 were between sites 1 and 2 for PH. Two and one MTAs were shared between two sites for DF and YP, respectively. All site-specific MTAs were consistent in effect direction but varied in magnitude across sites, except the MTAs for Locus 25179 and 6025 associated with DF and PH, respectively, which had opposite directions with near zero effect at non-significant sites. The number of site-specific MTAs observed for each trait differed among sites. Twelve out of 18 site-specific SNPs for OL were observed at site 3 (or site 4), six of 9 site-specific MTAs for SW were observed at site 1, 8 of 14 site-specific MTAs for DF were at site 1. (Table 3, Supplementary Tables s5 - s10).

Table 3
Number of MTAs identified by different GWAS methods

Trait	MLM-GWAS	Meta-GWAS	BayesR	Common MTAs	Shared MTAs (≥ 2 sites)	Site-specific MTAs
DF	52	62	38	16	2	14
OL	54	116	38	30	12	18
PH	71	71	32	17	7	10
PR	55	119	5	5	3	2
SW	41	150	27	20	11	9
YP	59	48	10	4	1	3

Five MTAs had significant associations with more than one trait. Locus 5628 was associated with PH at sites 2 and 3 and DF at site 1. Locus 9819 was associated with DF at site 1, OL at site 3, and PR at sites 2 and 3. Locus 17302 and 3057 increased SW but decreased PR and OL at site 2 and site 3 (or site 4), respectively. Locus 28935 was associated with low SW and high OL at sites 2 and 3 (or site 4) (Table 4).

Table 4

MTAs with effects on multiple traits with their MAF, physical position, Z-score, and p-value from the meta-GWAS

SNP	MAF	Pseudochromosome	Physical position	Trait	Site	Z-score	P-value
Locus5628	0.39	1	83,711,358	DF	1	4.167	3.08E-05
				PH	2,3	4.638	3.51E-06
Locus9819	0.1	8	79,766,528	DF	1	4.598	4.26E-06
				PR	2,3,4	5.709	1.14E-08
				OL	3	4.763	1.91E-06
Locus17302	0.1	12	1,089,790	PR	2	-4.555	5.24E-06
				SW	1,2	4.97	6.68E-07
				OL	2	-3.701	0.000215
Locus3057	0.06	7	15,110,222	PR	1,2,3,4	-8.97	2.95E-19
				SW	1,2,3,4	7.021	2.2E-12
				OL	3,4	-4.45	8.58E-06
Locus28935	0.08	8	23,744,954	SW	2,3,4	-5.623	1.88E-08
				OL	1,2,3,4	6.767	1.32E-11

Discussion

Understanding $G \times E$ is an important initial step to develop strategies for a breeding program in the target environment(s). Our results showed that $G \times E$ patterns differed between safflower traits. The identification of site-shared and site-specific MTAs in GWAS provides knowledge to broaden our understanding of the genetic basis of $G \times E$ interactions for important safflower traits.

Different $G \times E$ interaction patterns were observed for safflower traits

The seasonal rainfall in the Victoria Wimmera region (Horsham and Wonwondah) had an impact on safflower agronomic traits. Safflower is normally sown in Winter in Australia to maximize the usage of the available water from Winter and early Spring rain [37]. In our study, we observed that water stress during flowering decreased safflower grain yield substantially. Further, insufficient Spring rain heavily reduced safflower production via poor biomass accumulation. Similar grain yield instability induced by rainfall patterns has also been reported in other crops [38]. Besides grain yield, a 1% decrease of OL was observed under differing water stress, consistent with previous studies that oil content decreased under

drought stresses [39, 40]. The positive correlation between OL and PR (0.14 to ~ 0.46 across sites) agreed with a previous study where a positive correlation ($r = 0.476$) was reported in safflower [41]. The negative correlation between SW with PR and OL across all four sites suggests a negative relationship between carbohydrate accumulation and protein and oil accumulation. This could be similar to the competition between carbohydrate and protein accumulation in cereal crops [42, 43].

The overall $G \times E$ interactions were significant for all the traits studied. However, there were different levels of $G \times E$ for the different traits. The heterogeneous model further revealed detailed $G \times E$ patterns for each trait, which indicated the rank changes of accessions between sites. The high $G \times E$ observed for YP across sites was consistent with studies in other crops [44, 45]. Low to moderate pairwise genetic correlation indicated re-ranking for YP was high among sites. The cultivars and breeding lines performed well at site 1 (19 accessions out of the 50 top YP accessions) but not at the other three sites, suggesting that introgression of water stress tolerance from landraces could improve safflower yield stability. The low level of $G \times E$ for OL with limited re-ranking across sites observed in our study was also indicated in a soybean study [46]. According to the BLUEs for each site, about half of the top 30 accessions with high OL were cultivars and breeding lines, reflecting breeding efforts to improve OL in safflower cultivars. Although a moderate r_{Overall} was observed for DF and PH, the pairwise genetic correlation showed that lines were reranked strongly for DF at site 2 compared with other sites. The genetic divergence of DF among the accessions in response to water stress at flowering implied DF is important in developing drought tolerant varieties [47].

GWAS identified MTAs for safflower traits

GWAS has been widely used to study the genetic basis of the important agronomy traits with diverse germplasm in crops [48]. Multi-environment trials normally were combined to present the overall phenotypic variation for GWAS to detecting the associations between markers and traits [49, 50]. However, with diverse germplasm, the phenotypic variation displayed under differed environments can be used to measure the plasticity of the traits or trait $G \times E$ level with proper statistical models [51, 52]. Environmental stable and environmental specific MTAs can help our understanding of the genetic basis of trait $G \times E$, and it also will enrich our knowledge of the genetic architecture of the important agronomy traits [53, 54]. In our study, GWAS was carried out with a globally diverse safflower collection for six agronomic traits in four field trials that differed with water availability and 92 MTAs were identified. MTAs shared across sites were identified for traits with low $G \times E$, and site-specific MTAs were discovered for all traits with more site-specific MTAs than shared MTAs identified for moderate overall $G \times E$ traits.

A high number of significant MTAs were identified for seed oil content (OL) by all three GWAS approaches, of which 18 were shared across sites, and 12 were site-specific, indicating the complex genetic control of this trait. Studies with canola showed that 24 candidate genes were involved in fatty acid biosynthesis [55]. In safflower, a transcriptome study showed that a significant number of transcription factors were involved in oil accumulation in safflower seeds [56]. The six MTAs shared across four sites will be of interest to safflower breeders and geneticists as sources of genetic variation to improve the seed oil content in safflower under different growing conditions. Similarly, numerous MTA

(total 20) were identified for seed weight (SW), of which 11 were shared across sites. Three MTAs explained more than 10–20% phenotypic variance across sites will provide useful information for breeders to modify SW in safflower (Supplementary Table s8).

The molecular basis for $G \times E$ interactions could be due to site-specific QTLs, gene expression or differences in the magnitude of expression across environments [52, 54]. In our study, all site-specific MTAs showed differing allelic effects across sites for each trait (Supplementary Table s5-s10), however, the effects were significant in some environments but not in other environments. We postulate that the site-specific MTAs identified in our study could be related to the $G \times E$ observed for each trait. In our study, both PH and DF showed moderate overall $G \times E$, and a higher number of site-specific MTAs were identified. This is consistent with the studies in other crops that both traits are highly influenced by the drought conditions [57, 58]. Markers associated with PH and DF under drought conditions in safflower have been reported [11]. Further studies are needed to verify if they are linked with the MTAs identified in our study. Only one MTA was identified for DF at site 2, which could be related to the narrow phenotypic variation observed at site 2 [59]. MTAs identified for PH and DF under differed water stress environments could be used in breeding for drought resistant safflower varieties. Few MTAs for PR and YP were detected by all three GWAS methodologies. However, those that were identified explained a high proportion of the phenotypic variation for each trait, indicating their potential importance for genetic improvement.

Correlations between traits can be caused by pleiotropy or a close linkage of loci associated with the traits [60]. Shared major genes or QTL for flowering time and plant height have been reported in soybean [61, 62]. In our study, Locus 5628 was associated with DF at site 1 and PH at site 2 and 3, suggesting the MTA is likely tightly linked to different QTL affecting both traits, rather than being a single QTL with pleiotropic effects. In canola, a QTL affecting both OL and PR in repulsion was reported, suggesting the PR and OL biosynthesis pathways interfere and/or compete with one another [63]. In maize, high OL and high PR were achieved using the *opaque2* modifier genes. However, a yield reduction was noted [64]. In our study, we identified four MTAs affecting three traits, one MTAs influencing both SW and OL, one MTAs associated with PR and OL, and two MTAs interfering SW, OL, and PR. The allelic effects of those MTAs were consistent with the correlation observed in the field among the three traits that PR and OL are positively correlated, and both traits are negatively correlated with SW. This suggested that safflower breeding for PR and OL may differ from canola and maize. However, balancing seed weight and seed quality (OL and PR) would be a challenge. There were other strong phenotypic correlations, such as YP with PH and YP with PR, but associated markers were not identified. The reason could be the low number of significant SNPs that were observed for grain yield.

The interplay of GWAS results and genetic architectures

SW and OL are known as highly heritable traits in many crops, while yield is more quantitative in nature [65, 66]. The number of MTAs identified by the three GWAS methods did not fully reflect the complexity of the trait genetic architecture. One reason for this could be the thresholds used by the three methods. The p-value used in our single locus MLM-GWAS was relaxed, and a significant number of candidate MTAs

were observed for all traits. With meta-GWAS, we reported significant SNPs number at $-\log_{10}P$ value at 3 instead of 2, which detected more significant SNPs for each trait indicating the increased power (Supplementary Table s11). However, multi-locus BayesR, which can improve association mapping resolution by removing multiple SNPs being in LD with the same QTL, could detect SNPs with larger effects [26, 67]. We observed fewer MTAs for all traits with the BayesR methods with the arbitrary threshold of 0.7 posterior probability of a SNP having an effect. This threshold may have been too stringent for polygenic traits such as YP and PR. Only 10 MTAs associated with YP and 5 MTAs associated with PR were detected with BayesR, which explained 5–28% of the phenotypic variance.

Heterogenous model fit the data better

Mixed linear models are widely used for $G \times E$ analysis in crop research [68]. Falconer & Mackay (1996) suggested that the same trait measured in different environments should be considered as different (but correlated) traits [69]. In our study, the homogenous model 1 combined the four sites together and estimated the overall GXE pattern with only three parameters. However, the heterogeneous model 2 treated each site as an independent environment, and a total of 26 parameters were estimated. The increased number of parameters allowed dissection of $G \times E$ among individual environments to reveal hidden patterns of genetic correlation between sites. Furthermore, the AIC, BIC, and Logl were improved for all traits, indicating model 2 was better able to fit the data [70]. These findings agreed with Malosetti et al. (2013), who compared different models to study $G \times E$ interactions, and concluded that sophisticated mixed models are necessary to allow for heterogeneity of genetic variances and correlations across environments [51].

Conclusions

Two mixed linear models were applied to analyze the $G \times E$ pattern for grain yield (YP), days to flowering (DF), plant height (PH), 500 seed weight (SW), seed oil content (OL), and seed protein content (PR) in a globally diverse safflower collection grown in four field trials. The heterogenous mixed linear model (MLM) fitted data better and provided a detailed estimation of the $G \times E$ pattern. We observed that different water stress conditions impacted the performance of each of these traits differently, with low overall $G \times E$ observed for OL and SW and high overall $G \times E$ for YP. Ninety-two MTAs were identified with large effects MTAs detected for OL, SW, and PR across all sites. Site-specific MTAs were detected for all traits with differed allelic effects, suggesting these MTAs could be associated with trait $G \times E$. Five MTAs were associated with multiple traits. The uniform GWAS thresholds used in the study could have impacted the number of significant SNP identified for complex traits. This study has provided new insights into the genetic architecture of the traits studied, and it presents opportunities to exploit the MTA identified in breeding programs to increase yield stability and local adaptation in safflower.

Materials And Methods

Plant material and phenotyping

406 globally diverse safflower accessions were sourced from the Australian Grain Genebank (AGG), including elite cultivars, breeding lines, and landraces. Accessions information and the field trial experiment design are previously described [36]. In brief, with a randomized complete block design, all accessions were sown at two field sites in two consecutive years (2017 and 2018, a total of four sites that differed in water availability). Sites 1 and 2 were sown in 2017 at the same location (Horsham, Victoria) in a low rainfall zone. Site 1 was flood irrigated before sowing and considered an optimal site with a full soil water profile. Site 2 was rainfed, with soil water stress starting in late spring (during the flowering stage). Sites 3 and 4 were sown in 2018. Site 3 was at the same location as sites 1 and 2 but was rainfed and experienced soil water stress during the entire growing season, with minimal rain in the early spring and high temperature towards the end of the season. Site 4 was located in a higher rainfall zone (Wonwondah, Victoria) and received more rain overall than site 3, but also experienced soil water stress.

Days to flowering (DF) was recorded as the number of days from sowing to 25% of the plot flowering. Plant height (PH) was measured at the late flowering stage from the ground surface to the top of the plot canopy cm. Seed weight (SW) was measured as random 500 achenes from the whole plot in grams. Grain yield was measured as yield per plot (YP) in kilograms harvested by machine. Seed protein (PR) and seed oil content (OL) were determined by near-infrared reflectance spectroscopy (NIR, Foss Pacific Pty Ltd, Denmark) with calibration by the Dumas nitrogen combustion method for protein (TruMac, Leco Corporation St Joseph USA), and the Soxhlet extraction for oil (Soxtec 25050, FOSS, Hilleroed, Denmark). The NIR prediction models R-squared (R^2) and standard error of prediction (SEP) were 0.93 and 0.7% for seed protein content, 0.95 and 1.2% for seed total oil content.

Statistical analysis of phenotypic data

Summary statistics were calculated for each trait at each site. The best linear unbiased estimates (BLUE) for each trait at each site were calculated by a single site linear mixed model with safflower accessions fitted as fixed effects. The model was illustrated as:

$$Y_{mijk} = \mu + g_m + R_j + r_i + c_k + \epsilon_{mijk} \quad [1]$$

, where Y_{mijk} is the phenotypes of accessions m in rep j at row i , column k , μ is the overall mean, g_m is the fixed accession genetic effect, R_j is the replicate effect; r_i is the row effect, c_k is the column effect, and ϵ_{mijk} is the residual, including the AR1×AR1 covariance structure to adjust spatial variation.

Pearson's correlation at each site was calculated based on the BLUEs of each trait. BLUEs were used as the "phenotypes" for the GWAS.

To assess the G × E level for each trait, the four sites were combined, and the genetic effect associated with accessions was decomposed into two components: the genetic effect of accessions and the interaction effect between accessions among sites (G × E effect). The genetic effect of accessions and

interaction effect between accessions with sites were assumed to be homogenous for all sites, and the linear mixed model was:

$$Y_{ijk} = \mu + S_i + R_j + G_k + SG_{ik} + \epsilon_{ijk} [2]$$

, where Y_{ijk} is the phenotype of accession k in rep j at site i , μ is the overall mean, S_i is the fixed i -th site effect, R_j is the fixed replicate effect, G_k is the random accession genetic effect, SG_{ik} is the random interaction effect between accessions and sites, and ϵ_{ijk} is the residual. Two models, including and excluding the $G \times E$ effect for each trait as a random effect, were compared. A Log-Likelihood Ratio test was used to test the significance of the $G \times E$ effect for each trait [71]. The genetic correlation among the four sites ($r_{G\text{overall}}$) was estimated as the ratio of the genetic effect of accessions to total genetic variance, calculated as $r_{G\text{overall}} = \sigma_G^2 / (\sigma_G^2 + \sigma_{GE}^2)$, where σ_G^2 is the genetic variance of accessions, σ_{GE}^2 is the variance for $G \times E$ interaction. High genetic correlation among sites indicated low $G \times E$ interaction, while low genetic correlation indicated high $G \times E$ interaction [72].

A heterogeneous variance structure was also fitted in the linear mixed model, which assumes that accessions genetic effect is independent at each site. It can be illustrated as:

$$Y_{ijk} = \mu + S_i + SR_{ij} + SG_{ik} + \epsilon_{ijk} [3]$$

Where the terms are the same as above, with the site as a fixed effect and the accession and trial replicate effects both nested within sites as a random effect with different variance for each site. The residual variance was also nested within the site, with the AR1×AR1 covariance structure used to adjust spatial variation across columns and rows within each site. The genetic correlation of the accession effect between two sites was calculated as $r_{Gij} = \sigma_{GiGj} / \sqrt{\sigma_{Gi}^2 * \sigma_{Gj}^2}$, where the σ_{Gi}^2 and σ_{Gj}^2 are the variance of the accession genetic effect at sites i and j , respectively. The σ_{GiGj} is the covariance of the accessions genetic effect at sites i and j . Similar to the above, high genetic correlations between two sites indicated a low $G \times E$ interaction. The significance of the genetic correlation between two sites was tested for deviation from 1 using Likelihood Ratio Tests. If r_{Gij} significantly differed from 1, it suggested the ranking of accessions at the two sites was different. Akaike Information Criterion (AIC) was used to compare the fitness of models 2 and 3.

SNP genotyping and population structure

349 accessions were genotyped using a genotyping-by-sequencing assay as described in [36]. SNPs were filtered for a missing data rate < 30% and minor allele frequency (MAF) > 0.01 and imputed with LinkImpute [73]. The physical position of the filtered SNP was determined by mapping their flanking sequences to a draft safflower genome assembly (unpublished data) with 12 main scaffolds (pseudochromosomes). Linkage disequilibrium (LD) was calculated for all pairwise SNP using PLINK [74]. 318 safflower accessions with both imputed genotyping data and phenotyping data were used for

the association study. Population structure was evaluated from the genomic relationship matrix (GRM) according to VanRaden [75], which was calculated using filtered SNP.

Genome wide association study

Single site GWAS was conducted for each trait using the BLUEs of each trait as the “phenotypes” (Supplementary Table s1). First, a single SNP regression model, referred to as MLM-GWAS, implemented in the GCTA software [76], was performed with the GRM fitted to account for population structure. Manhattan and quantile-quantile (Q-Q) plots generated with an R script [77] were used to visualize associations for each trait. Second, the Bayesian multi-locus approach-BayesR was performed using the Markov Chain Monte Carlo (MCMC) method with 50,000 iterations and 25,000 burn-in. SNPs with large effects were declared if they had a non-zero effect with at least a 0.7 posterior probability, averaged over 5 runs [24]. And third, meta-GWAS implemented in the software Metal [78] was performed in which the single site MLM-GWAS was treated as an independent study. SNP identified by all three methods were considered candidate MTAs for each trait.

Abbreviations

YP

grain yield

DF

Days to flowering

PH

Plant height

SW

500 Seed weight

PR

seed protein content

OL

seed oil content

r_{Goverall}

genetic correlations across sites

r_{Gij}

genetic correlations between all pairwise sites

MTA

Marker trait association

QTL

Quantitative trait loci

SNP

Single-nucleotide polymorphism

GWAS

Genome-wide association study
MLM
Mixed linear model
BayesR
Bayesian multi-locus method
FAD
Fatty acid desaturases
LD
Linkage disequilibrium
BIC
Bayesian Information Content
AIC
Akaike Information Criterion
BLUE
The best linear unbiased estimate
NIR
Near-infrared spectrometry
 R^2
R-squared
SEP
standard error of prediction
MAF
minor allele frequency
Q-Q plots
quantile-quantile
MCMC
Markov Chain Monte Carlo method

Declarations

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Availability of data and materials

The phenotypic datasets supporting the conclusions of this article is(are) included within the article and the attached additional files. And the genotype dataset, please see the previously published paper:

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by Agriculture Victoria Research, Victoria state government, Australia.

Author's contributions

HZ, MH, and HD conceived and designed the experiment. PM, JT, SK, and MH performed and supervised the phenotyping, genotyping, genome assembly; KS performed SNP annotation and previewed the manuscript; YL assisted with the statistical analysis; EB assisted with BayesR analysis; HZ performed the data analysis and wrote the manuscript; MH and HD supervised the study and data interpretation. All authors revised the manuscript.

Acknowledgments

The authors thank the Australian Grains Genebank for providing seed for the safflower accessions. The collective efforts of the Agriculture Victoria field and lab technical staff are most appreciated.

References

1. Knowles P, Ashri A: **Safflower – *Carthamus Tinctorius* (Compositae)**. In: *Evolution of Crop Plants*. Edited by Smartt J, Simmonds NW. Harlow, UK: Longman; 1995.
2. Fernández-Martínez J, del Río M, de Haro A: **Survey of safflower (*Carthamus tinctorius* L.) germplasm for variants in fatty acid composition and other seed characters**. *Euphytica* 1993, **69**(1):115-122.
3. Khalid N, Khan RS, Hussain MI, Farooq M, Ahmad A, Ahmed I: **A comprehensive characterisation of safflower oil for its potential applications as a bioactive food ingredient - A review**. *Trends in Food Science & Technology* 2017, **66**(Supplement C):176-186.
4. [<http://www.fao.org/faostat/en/#data>]
5. Jochinke D, Wachsmann N, Potter T, Norton R: **Growing safflower in Australia: Part 1 - History, experiences and current constraints on production**. In: *The 7th international safflower conference* Waga Wagga, Australia; 2008.
6. Li D, Mündel HH: **Safflower, *Carthamus tinctorius* L. promoting the conservation and use of underutilized and neglected crops 7**. Rome, Italy: Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute; 1996.
7. Kotecha A: **Inheritance and Association of Six Traits in Safflower**¹. *Crop Science* 1979, **19**(4):cropsci1979.0011183X001900040022x.

8. Ramachandram M, Goud JV: **Genetic analysis of seed yield, oil content and their components in safflower (*Carthamus tinctorius* L.)**. *Theor Appl Genet* 1981, **60**(3):191-195.
9. Hamdan Y, García-Moreno M, Fernández-Martínez J, Velasco L, Pérez-Vich B: **Mapping of major and modifying genes for high oleic acid content in safflower**. *Molecular Breeding* 2012, **30**:1279-1293.
10. Ambreen H, Kumar S, Kumar A, Agarwal M, Jagannath A, Goel S: **Association Mapping for Important Agronomic Traits in Safflower (*Carthamus tinctorius* L.) Core Collection Using Microsatellite Markers**. *Frontiers in Plant Science* 2018, **9**(402).
11. Ebrahimi F, Majidi MM, Arzani A, Mohammadi-Nejad G: **Association analysis of molecular markers with traits under drought stress in safflower**. *Crop and Pasture Science* 2017, **68**(2):167-175.
12. Cao S, Zhou XR, Wood CC, Green AG, Singh SP, Liu L, Liu Q: **A large and functionally diverse family of *Fad2* genes in safflower (*Carthamus tinctorius* L.)**. *BMC Plant Biol* 2013, **13**:5.
13. Wood CC, Okada S, Taylor MC, Menon A, Mathew A, Cullerne D, Stephen SJ, Allen RS, Zhou XR, Liu Q *et al*: **Seed-specific RNAi in safflower generates a superhigh oleic oil with extended oxidative stability**. *Plant Biotechnol J* 2018, **16**(10):1788-1796.
14. Wang MH, Cordell HJ, Van Steen K: **Statistical methods for genome-wide association studies**. *Semin Cancer Biol* 2019, **55**:53-60.
15. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM *et al*: **Mixed linear model approach adapted for genome-wide association studies**. *Nat Genet* 2010, **42**(4):355-360.
16. Ledesma-Ramírez L, Solís-Moya E, Iturriaga G, Sehgal D, Reyes-Valdes MH, Montero-Tavera V, Sansaloni CP, Burgueño J, Ortiz C, Aguirre-Mancilla CL *et al*: **GWAS to Identify Genetic Loci for Resistance to Yellow Rust in Wheat Pre-Breeding Lines Derived From Diverse Exotic Crosses**. *Frontiers in Plant Science* 2019, **10**(1390).
17. Qu C, Jia L, Fu F, Zhao H, Lu K, Wei L, Xu X, Liang Y, Li S, Wang R *et al*: **Genome-wide association mapping and Identification of candidate genes for fatty acid composition in *Brassica napus* L. using SNP markers**. *BMC Genomics* 2017, **18**(1):232.
18. Leamy LJ, Zhang H, Li C, Chen CY, Song B-H: **A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*)**. *BMC Genomics* 2017, **18**(1):18.
19. Pasaniuc B, Price AL: **Dissecting the genetics of complex traits using summary association statistics**. *Nature Reviews Genetics* 2017, **18**(2):117-127.
20. Joukhadar R, Thistlethwaite R, Trethowan R, Keeble-Gagnère G, Hayden MJ, Ullah S, Daetwyler HD: **Meta-analysis of genome-wide association studies reveal common loci controlling agronomic and quality traits in a wide range of normal and heat stressed environments**. *Theor Appl Genet* 2021, **134**(7):2113-2127.
21. Fikere M, Barbulescu DM, Malmberg MM, Spangenberg GC, Cogan NOI, Daetwyler HD: **Meta-analysis of GWAS in canola blackleg (*Leptosphaeria maculans*) disease traits demonstrates increased power from imputed whole-genome sequence**. *Sci Rep* 2020, **10**(1):14300.

22. Tamba CL, Ni Y-L, Zhang Y-M: **Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies.** *PLOS Computational Biology* 2017, **13**(1):e1005357.
23. Kaler AS, Gillman JD, Beissinger T, Purcell LC: **Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize.** *Frontiers in Plant Science* 2020, **10**(1794).
24. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, Mason BA, Goddard ME: **Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels.** *Journal of dairy science* 2012, **95**(7):4114-4129.
25. Daetwyler HD, Bansal U, Bariana H, Hayden M, Hayes B: **Genomic prediction for rust resistance in diverse wheat landraces.** *TAG Theoretical and applied genetics Theoretische und angewandte Genetik* 2014, **127**.
26. Pasam RK, Bansal U, Daetwyler HD, Forrest KL, Wong D, Petkowski J, Willey N, Randhawa M, Chhetri M, Miah H *et al.*: **Detection and validation of genomic regions associated with resistance to rust diseases in a worldwide hexaploid wheat landrace collection using BayesR and mixed linear model approaches.** *Theor Appl Genet* 2017, **130**(4):777-793.
27. Xiang R, Breen EJ, Prowse-Wilkins CP, Chamberlain AJ, Goddard ME: **Bayesian genome-wide analysis of cattle traits using variants with functional and evolutionary significance.** *bioRxiv* 2021:2021.2005.2005.442705.
28. Bradshaw AD: **Evolutionary significance of phenotypic plasticity in plants.** *Adv Genet* 1965, **13**:115-155.
29. Kusmec A, de Leon N, Schnable PS: **Harnessing Phenotypic Plasticity to Improve Maize Yields.** *Frontiers in Plant Science* 2018, **9**(1377).
30. Das A, Parihar AK, Saxena D, Singh D, Singha KD, Kushwaha KPS, Chand R, Bal RS, Chandra S, Gupta S: **Deciphering Genotype-by- Environment Interaction for Targeting Test Environments and Rust Resistant Genotypes in Field Pea (*Pisum sativum* L.).** *Frontiers in Plant Science* 2019, **10**(825).
31. Li B, Zhao W, Li D, Chao H, Zhao X, Ta N, Li Y, Guan Z, Guo L, Zhang L *et al.*: **Genetic dissection of the mechanism of flowering time based on an environmentally stable and specific QTL in *Brassica napus*.** *Plant science : an international journal of experimental plant biology* 2018, **277**:296-310.
32. Jamshidmoghaddam M, Pourdard SS: **Genotype × environment interactions for seed yield in rainfed winter safflower (*Carthamus tinctorius* L.) multi-environment trials in Iran.** *Euphytica* 2013, **190**(3):357-369.
33. Alizadeh K, Mohammadi R, Shariati A, Eskandari M: **Comparative analysis of statistical models for evaluating of genotype x environment interaction in rainfed safflower.** *Agricultural Research* 2017, **6**.
34. Chenu K, Dehifard R, Chapman SC: **Large-scale characterization of drought pattern: a continent-wide modelling approach applied to the Australian wheatbelt—spatial and temporal trends.** *New Phytol* 2013, **198**(3):801-820.

35. Robertson A: **The Sampling Variance of the Genetic Correlation Coefficient.** *Biometrics* 1959, **15**(3):469-485.
36. Zhao H, Li Y, Petkowski J, Kant S, Hayden MJ, Daetwyler HD: **Genomic prediction and genomic heritability of grain yield and its related traits in a safflower genebank collection.** *The Plant Genome* 2020, **n/a**(n/a):e20064.
37. Wachsmann NJ, D.; Potter, T.; Norton, R.: **Growing safflower in Australia: part 2 - Agronomic research and suggestions to increase yields and production.** In: *Safflower: unexploited potential and world adaptability 7th International Safflower Conference*, . Edited by Knights SEP, T.D. Wagga Wagga, New South Wales, Australia.: Agri-MC marketing and Communication; 2008: 1-8.
38. Sadras V, Dreccer MF: **Adaptation of wheat, barley, canola, field pea and chickpea to the thermal environments of Australia.** *Crop and Pasture Science* 2015, **66**(11):1137-1150.
39. Ebrahimi F, Majidi MM, Arzani A, Mohammadi-Nejad G: **Association analysis of molecular markers with traits under drought stress in safflower.** *Crop and Pasture Science* 2017, **68**(2):167-175, 169.
40. Joshan Y. SB, Jabbari H., Mozafari H., Moaveni P.: **Effect of drought stress on oil content and fatty acids composition of some safflower genotypes.** *Plant Soil Environ* 2019, **65**:4.
41. Oz M: **Relationship between Sowing Time, Variety, and Quality in Safflower.** *Journal of Chemistry* 2016, **2016**:9835641.
42. Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, Graner A: **Genome-wide association studies for agronomical traits in a world wide spring barley collection.** *BMC Plant Biology* 2012, **12**(1):16.
43. Bjarnason M, Vasal SK: **Breeding of Quality Protein Maize (QPM).** In: *Plant Breeding Reviews*. 1992: 181-216.
44. Tolessa T, Keneni G, Gela TS, Jarso M, Bekele Y: **Genotype × Environment Interaction and Performance Stability for Grain Yield in Field Pea (*Pisum sativum* L.) Genotypes.** *International Journal of Plant Breeding* 2013, **7**:116-123.
45. He S, Thistlethwaite R, Forrest K, Shi F, Hayden MJ, Trethowan R, Daetwyler HD: **Extension of a haplotype-based genomic prediction model to manage multi-environment wheat data using environmental covariates.** *Theoretical and Applied Genetics* 2019, **132**(11):3143-3154.
46. Sudarić A, Šlmić D, Vratarić M: **Characterization of genotype by environment interactions in soybean breeding programmes of southeast Europe.** *Plant Breeding* 2006, **125**(2):191-194.
47. Bhandari A, Sandhu N, Bartholome J, Cao-Hamadoun T-V, Ahmadi N, Kumari N, Kumar A: **Genome-Wide Association Study for Yield and Yield Related Traits under Reproductive Stage Drought in a Diverse indica-aus Rice Panel.** *Rice* 2020, **13**(1):53.
48. Liu HJ, Yan J: **Crop genome-wide association study: a harvest of biological relevance.** *Plant J* 2019, **97**(1):8-18.
49. Landers DA, Stapleton AE: **Genetic interactions matter more in less-optimal environments: a Focused Review of "Phenotype uniformity in combined-stress environments has a different genetic**

- architecture than in single-stress treatments” (Makumburage and Stapleton, 2011). *Frontiers in Plant Science* 2014, **5**(384).
50. Leamy LJ, Zhang H, Li C, Chen CY, Song BH: **A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*)**. *BMC Genomics* 2017, **18**(1):18.
51. Malosetti M, Ribaut J-M, van Eeuwijk FA: **The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis**. *Frontiers in Physiology* 2013, **4**(44).
52. Des Marais DL, Hernandez KM, Juenger TE: **Genotype-by-Environment Interaction and Plasticity: Exploring Genomic Responses of Plants to the Abiotic Environment**. *Annual Review of Ecology, Evolution, and Systematics* 2013, **44**(1):5-29.
53. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB: **Genetic architecture: the shape of the genetic contribution to human traits and disease**. *Nature Reviews Genetics* 2018, **19**(2):110-124.
54. Li X, Guo T, Mu Q, Li X, Yu J: **Genomic and environmental determinants and their interplay underlying phenotypic plasticity**. *Proceedings of the National Academy of Sciences* 2018, **115**(26):6679.
55. Qu C, Jia L, Fu F, Zhao H, Lu K, Wei L, Xu X, Liang Y, Li S, Wang R *et al*: **Genome-wide association mapping and identification of candidate genes for fatty acid composition in *Brassica napus* L. using SNP markers**. *BMC genomics* 2017, **18**(1):232-232.
56. Li D, Wang Q, Xu X, Yu J, Chen Z, Wei B, Wu W: **Temporal transcriptome profiling of developing seeds reveals candidate genes involved in oil accumulation in safflower (*Carthamus tinctorius* L.)**. *BMC Plant Biology* 2021, **21**(1):181.
57. Yuan Y, Cairns JE, Babu R, Gowda M, Makumbi D, Magorokosho C, Zhang A, Liu Y, Wang N, Hao Z *et al*: **Genome-Wide Association Mapping and Genomic Prediction Analyses Reveal the Genetic Architecture of Grain Yield and Flowering Time Under Drought and Heat Stress Conditions in Maize**. *Frontiers in Plant Science* 2019, **9**(1919).
58. Ma X, Feng F, Wei H, Mei H, Xu K, Chen S, Li T, Liang X, Liu H, Luo L: **Genome-Wide Association Study for Plant Height and Grain Yield in Rice under Contrasting Moisture Regimes**. *Frontiers in Plant Science* 2016, **7**(1801).
59. Stich B, Melchinger A: **An introduction to association mapping in plants**. *Cab Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, v5, 1-9 (2010) 2010, **5**.
60. Chen Y, Lübberstedt T: **Molecular basis of trait correlations**. *Trends in Plant Science* 2010, **15**(8):454-461.
61. Cober ER, Morrison MJ: **Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean**. *Theor Appl Genet* 2010, **120**(5):1005-1012.
62. Fang C, Ma Y, Wu S, Liu Z, Wang Z, Yang R, Hu G, Zhou Z, Yu H, Zhang M *et al*: **Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean**. *Genome Biology* 2017, **18**(1):161.

63. Chao H, Wang H, Wang X, Guo L, Gu J, Zhao W, Li B, Chen D, Raboanatahiry N, Li M: **Genetic dissection of seed oil and protein content and identification of networks associated with oil content in Brassica napus.** *Sci Rep* 2017, **7**:46295.
64. Vanous A, Gardner C, Blanco M, Martin-Schwarze A, Wang J, Li X, Lipka AE, Flint-Garcia S, Bohn M, Edwards J *et al*: **Stability Analysis of Kernel Quality Traits in Exotic-Derived Doubled Haploid Maize Lines.** *The Plant Genome* 2019, **12**(1):170114.
65. Ward BP, Brown-Guedira G, Kolb FL, Van Sanford DA, Tyagi P, Sneller CH, Griffey CA: **Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia.** *PLoS One* 2019, **14**(2):e0208217-e0208217.
66. Xiao Z, Zhang C, Tang F, Yang B, Zhang L, Liu J, Huo Q, Wang S, Li S, Wei L *et al*: **Identification of candidate genes controlling oil content by combination of genome-wide association and transcriptome analysis in the oilseed crop Brassica napus.** *Biotechnology for Biofuels* 2019, **12**(1):216.
67. Kemper KE, Reich CM, Bowman PJ, vander Jagt CJ, Chamberlain AJ, Mason BA, Hayes BJ, Goddard ME: **Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions.** *Genetics Selection Evolution* 2015, **47**(1):29.
68. Smith AB, Cullis BR, Thompson R: **The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches.** *The Journal of Agricultural Science* 2005, **143**(6):449-462.
69. Falconer DS, Mackay TFC: **Introduction to Quantitative Genetics, 4th Edition**, 4th edition edn: Longman Group Ltd; 1996.
70. H. A: **Information Theory and an Extension of the Maximum Likelihood Principle.** In: *Selected Papers of Hirotugu Akaike Springer Series in Statistics (Perspectives in Statistics)*. Edited by Parzen E. TK, Kitagawa G. New York, NY: Springer; 1998.
71. Kendall MG, Stuart A: **The advantage theory of statistics** London: Griffin and Co.; 1979.
72. Li Y, Wilcox P, Telfer E, Graham N, Stanbra L: **Association of single nucleotide polymorphisms with form traits in three New Zealand populations of radiata pine in the presence of genotype by environment interactions.** *Tree Genetics & Genomes* 2016, **12**(4):63.
73. Money D, Gardner K, Migicovsky Z, Schwaninger H, Zhong GY, Myles S: **LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms.** *G3* 2015, **5**(11):2383-2390.
74. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ *et al*: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**(3):559-575.
75. VanRaden PM: **Efficient methods to compute genomic predictions.** *Journal of dairy science* 2008, **91**(11):4414-4423.
76. Yang J, Lee SH, Goddard ME, Visscher PM: **GCTA: a tool for genome-wide complex trait analysis.** *Am J Hum Genet* 2011, **88**(1):76-82.

77. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB *et al*: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness.** *Nature Genetics* 2006, **38**(2):203-208.
78. Willer CJ, Li Y, Abecasis GR: **METAL: fast and efficient meta-analysis of genomewide association scans.** *Bioinformatics* 2010, **26**(17):2190-2191.

Figures

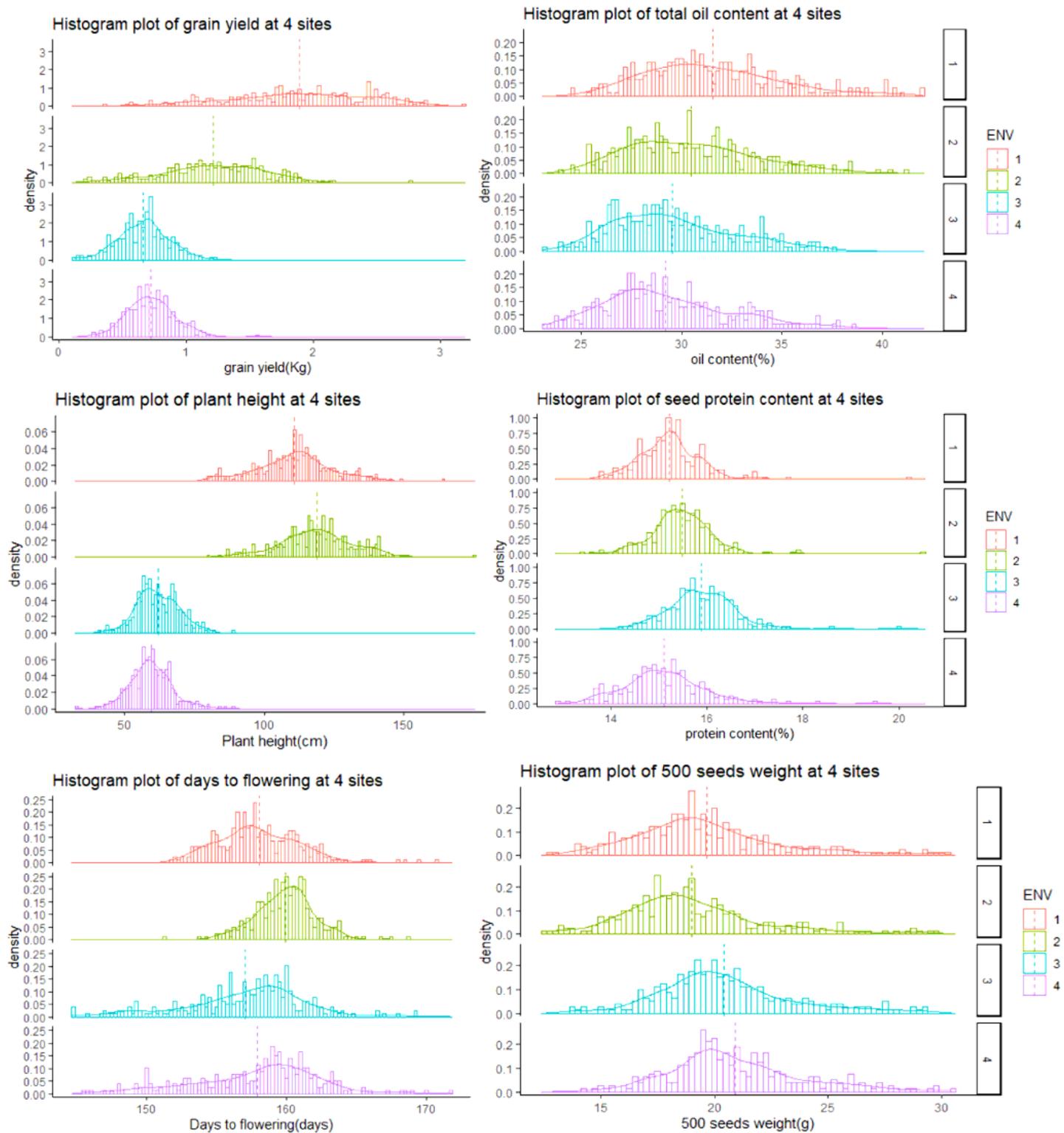


Figure 1

Distribution of grain yield (YP), plant height (PH), days to flowering (DF), 500 seed weight (SW), seed protein (PR), and oil content (OL) among the 406 safflower accessions at each of the four field trial sites (dashed line shows the trait mean).

Site	Trait	OL	PH	PR	SW	YP
ENV1	DF	0.10	0.46	-0.09	-0.22	0.10
	OL		-0.11	0.19	-0.51	0.15
	PH			-0.22	0.01	0.44
	PR				-0.38	-0.37
	SW					0.03
ENV2	DF	0.29	-0.08	0.20	-0.35	-0.16
	OL		-0.08	0.20	-0.51	0.03
	PH			-0.32	0.13	0.48
	PR				-0.48	-0.41
	SW					0.21
ENV3	DF	0.03	0.58	0.07	-0.17	0.10
	OL		-0.06	0.51	-0.51	-0.03
	PH			-0.08	-0.05	0.42
	PR				-0.55	-0.29
	SW					0.18
ENV4	DF	0.02	0.62	-0.02	-0.15	0.19
	OL		-0.11	0.64	-0.52	0.10
	PH			-0.11	0.00	0.34
	PR				-0.47	-0.21
	SW					0.07

Figure 2

Pearson correlation between traits at each site (red color indicates a highly negative correlation while green color indicates a highly positive correlation).

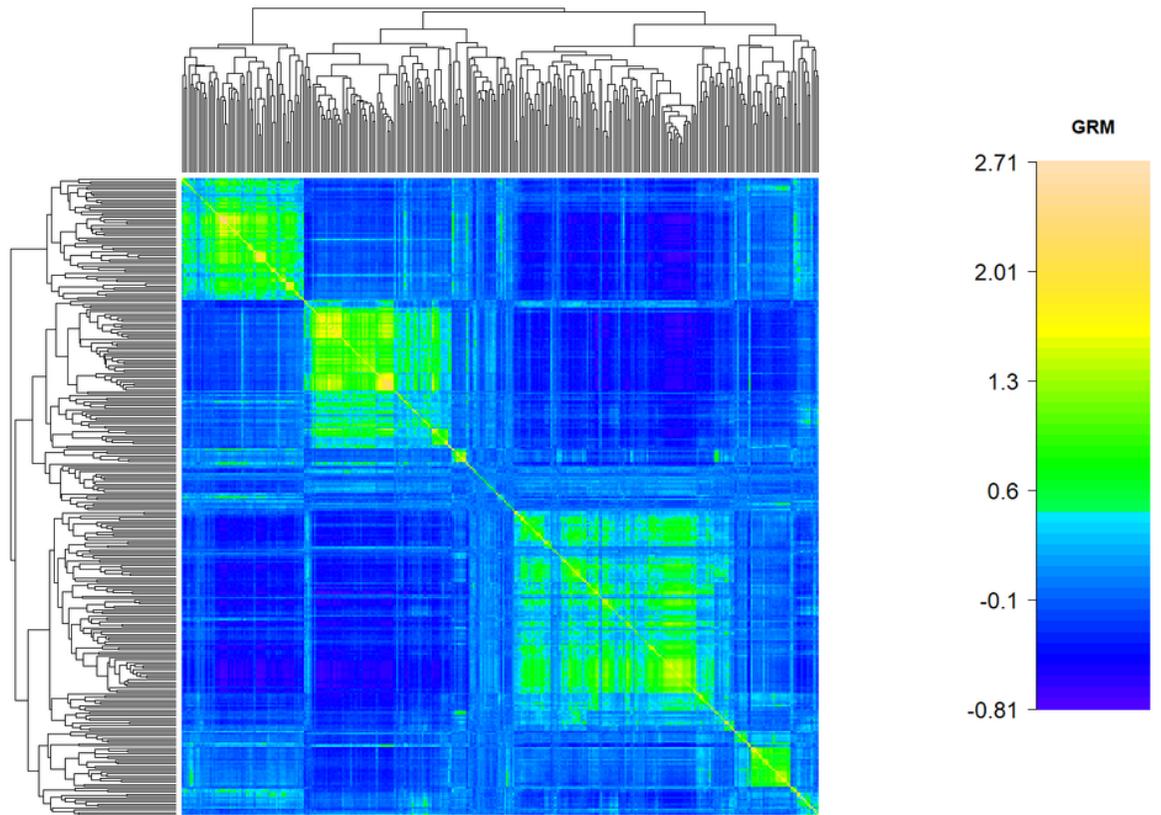


Figure 3

Heatmap of genomic relationships among 318 safflower accessions. Blue indicates less related, while orange means more related.

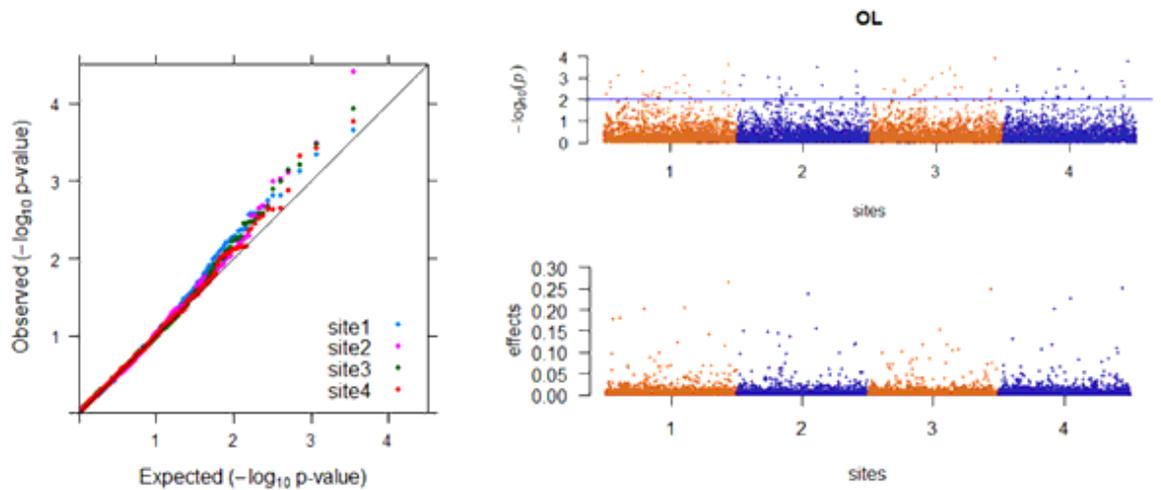


Figure 4

Left, QQ plot for seed oil content (OL) at the four trial sites. Right, negative log10 p values are plotted for all SNPs across four sites. The blue line indicates the threshold of significance for the MLM-GWAS

method (upper) and SNP effects for OL in the BayesR method (bottom). SNP are sorted according to physical position in the 12 pseudochromosomes and are in the same order for each of the four trial sites.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Zhaoetal2021supplementary.docx](#)
- [Zhaoetal2021supplementary.xlsx](#)