Identification of QTL for seed yield and agronomic traits in 944 soybean (Glycine max) RILs from a diallel cross of early-maturing varieties

Xintian Zhu | Willmar L. Leiser | Volker Hahn | Tobias Würschum

Abstract

Increasing soybean yield plays a key role in meeting the high demand for protein in Europe and other countries. The aim of this study was to dissect the genetic architecture underlying seed yield, plant height, protein yield and thousand-seed weight in early-maturing soybean. To this end, we performed QTL mapping based on 944 RILs derived from a half-diallel crossing design of five parents. We identified five to eight QTL for each of the four agronomic traits and some explained a considerable proportion of the genotypic variance. The three major QTL showed pleiotropic effects on two or more traits. Fine characterization revealed the maturity genes E1 and E3, and the stem growth habit gene Dt2 as likely candidates underlying these QTL. In general, the allele increasing seed yield also resulted in taller plants, which needs to be considered during selection due to an increased risk of lodging. Collectively, our results underline the strong effect of some loci like the E1 gene on a range of traits including seed yield, making them attractive targets for a marker-assisted selection.

Keywords

maturity genes, plant height, protein yield, QTL mapping, seed yield, soybean

1 | INTRODUCTION

Soybean is the most important and most widely grown legume crop worldwide, being used for human consumption and animal feed. The global demand for soybean has been constantly growing in the last decade, but only very little is produced in Europe (2.6% of world production; FAOSTAT, 2020). Europe is consequently highly dependent on soybean imports that currently account for around 95% of the annual demand. The majority of soybean in Europe is used for animal feeding, mainly for poultry and pork with their requirement for high protein (Tillie & Rodriguez-Cerezo, 2015). In addition, tofu (soybean curd), the most common soybean food product in Europe, is a plant protein alternative to meat, and is becoming increasingly popular with European consumers due to the trend towards vegetarian and vegan lifestyle (Kurasch et al., 2018; Kurasch, Leiser, et al., 2018; Würschum et al., 2019). One solution to reducing Europe’s high dependency on protein imports is to increase the acreage of legume cultivation. Soybean is well suited for this, and in recent years, soybean cultivars of early maturity groups have been promoted in Central and even Northern Europe (Hahn & Würschum, 2014; Kurasch, Hahn, Leiser, Vollmann, et al., 2017). A central issue is the adaptation to higher latitudes, with their different photoperiod and reduced growing season, which is mediated by maturity loci. These E loci confer photoperiod insensitivity, and E1, E2, E3 and E4, have

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2021 The Authors. Plant Breeding published by Wiley-VCH GmbH.
been shown to be essential for the spread of soybean to regions of higher latitude (Jiang et al., 2014). In recent years, the annual average yield of soybean was about 2.0 Mg/ha in Europe, which is lower than the worldwide average yield of 2.7 Mg/ha (FAOSTAT, 2020). In Germany, soybean yields typically range between 2.5 and 4.0 Mg/ha, but need to be increased further for soybean to be economically competitive with traditionally grown crops such as wheat. Thus, the improvement of soybean yield remains a major target in soybean breeding in Europe and other regions of higher latitude.

Seed yield is a highly quantitative trait. At present, Soybase (https://soybase.org/; Grant et al., 2009) lists over 150 QTL related to seed yield. Most of these QTL were identified by QTL mapping using recombinant inbred lines (RILs) or backcross populations (Du et al., 2009; Fox et al., 2015; Gai et al., 2007; Guzman et al., 2007; Hnetkovsky et al., 1996; Kim et al., 2012; Mansur et al., 1996; Orf et al., 1999; Rossi et al., 2013; Wang et al., 2004; Wang et al., 2014; Zhang et al., 2004). Given the genetic complexity of seed yield, many studies not only focused on yield but also on yield-related agronomic traits, like plant height (Chang et al., 2018; Li et al., 2020; Orf, Chase, Adler, et al., 1999; Specht et al., 2001; Sun et al., 2006; Xue et al., 2019; Yin et al., 2017). Thousand-seed weight is another important yield component and positively correlated with yield (Burris et al., 1973; Smith & Camper, 1975). Many QTL were identified in previous studies (Fasoula et al., 2004; Han et al., 2012; Hirata et al., 2014; Kato et al., 2014; Mian et al., 1996; Sun et al., 2012; Wu et al., 2018; Yan et al., 2014), but until now only few of the underlying genes have been cloned (Lu et al., 2017).

The aim of this study was to investigate the genetic architecture of seed yield, the seed yield component trait thousand-seed weight, plant height and protein yield. The study was based on 944 RILs derived from eight crosses among five early-maturing European soybean cultivars. In particular, our objectives were to (a) perform QTL mapping for the four traits, (b) fine-map identified QTL, (c) analyse QTL co-localizations for six traits by also considering the quality traits protein content and oil content, and (d) draw conclusions for the potential of marker-assisted selection for the investigated traits in soybean breeding.

2 | MATERIALS AND METHODS

2.1 | Plant materials

A total of 944 F5:8 recombinant inbred lines (RILs) of soybean (Glycine max [L.] Merr.) were derived from eight families, which were constructed in a half-diallel mating design using five parental lines (Kurasch et al., 2017). The five varieties used as parents are ‘Gallec’ (P1, e1-nl e2-ns E3-Ha E4), ‘Primus’ (P2, e1-nl e2-ns E3-Ha E4), ‘Protina’ (P3, e1-nl e2-ns e3-Mo E4), ‘Sultana’ (P4, e1-nl e2-ns E3-Ha e4-SORE-1) and ‘Sigalia’ (P5, e1-as e2-ns e3-tr E4), that all show a good performance in Central Europe (the allelic state at the maturity loci E1 - E4 is given in brackets). The number of RILs in each of the eight families was as follows: ‘P1 × P2’ (104), ‘P1 × P3’ (117), ‘P1 × P5’ (234), ‘P2 × P5’ (80), ‘P2 × P3’ (117), ‘P2 × P4’ (106), ‘P3 × P5’ (94), ‘P4 × P5’ (92).

2.2 | Phenotypic data

The field trials and the phenotypic data of the four traits, seed yield, plant height, protein yield and thousand-seed weight, have been described previously (Kurasch, Hahn, Leiser, Starck, et al., 2017). In brief, all RILs and their parents were classified according to their maturity date relative to each other as the earliest, middle and latest maturing genotypes, and as these groups then grown in three trials. The three trials were grown with overlapping genotypes between trials, 38 lines were included in Trial 1 and Trial 2, and 66 lines were grown in both Trial 2 and Trial 3. The trials were grown in a partially replicated design with 20% of the lines grown in replication. At each location a different set of lines was replicated. The three trials were grown at three locations in Germany (Eckartsweier, 48°31’47’’N, 7°51’18’’E; Hohenheim, 48°42’42’’N, 9°12’41’’E; Neuenstein, 49°12’23’’N, 9°34’54’’E) in 2014, in yield plots with 4 rows and 9 m × 6 m) and 65 seeds m−2.

The harvested seeds were dried to reach the same moisture content before evaluating dry matter seed yield and thousand-seed weight. Plant height was measured based on the distance between the ground and the last trifoliate leaf at the beginning of maturity and protein yield was derived by protein content multiplied with seed yield. Best linear unbiased estimates (BLUEs) were calculated across locations and were used for QTL mapping (Figure S1). Heritability was estimated following the BLUP-based approach suggested by Piepho and Möhring (2007). The estimated heritability for seed yield, plant height, protein yield and thousand-seed weight was 0.77, 0.91, 0.70 and 0.94, respectively.

2.3 | Linkage map construction

The pipeline of linkage map construction was as described in our previous study (Zhu et al., 2020). In brief, we obtained sequence data by Genotyping-by-sequencing (Elshire et al., 2011) and then, the raw data were processed and converted into ABH format. Next, the construction of the genetic linkage map was performed by using the R packages R/qtl and ASMap (Broman et al., 2003; Taylor & Butler, 2017) and finally, the R package Mapfuser was used to build a consensus map (van Muijen et al., 2017). Our final linkage map included 20 linkage groups, with a cumulative distance of 3,202.68 cM and a total marker number of 10,893 distributed at 3,605 unique positions. The average distance between unique marker positions across the whole consensus map was 0.89 cM.

2.4 | QTL mapping

We chose PlabMQTL to perform single family QTL mapping by using composite interval mapping and an additive genetic model (Utz, 2012). The number of cofactors was chosen by the modified Bayesian information criterion (Baierl et al., 2006). QTL scanning was carried out at intervals with 1cM and an empirical LOD
threshold was determined by 2000 random permutations for each trait with a genome-wide error rate of $\alpha \leq 0.1$. The support interval was specified as a 1-LOD fall-off (Darvasi & Soller, 1997). Two-hundred fivefold cross-validation runs were conducted to assess the frequency of QTL detection. The proportion of explained genotypic variance was assessed as $P_G = \frac{R_{adj}^2}{h^2}$, where $R_{adj}^2$ is the adjusted explained phenotypic variance by this QTL and $h^2$ is the heritability of the trait.

Multi-family mapping was conducted with the R package mmpR with an additive model that adds connection between families via shared parents (parental model) (Garin et al., 2017). An allele effect was estimated for each parental line, which is then more independent from the genetic background as estimated across all families containing this parent (Blanc et al., 2006). Here, we assumed that the residual terms are cross-specific, fitted by REML using ASReml-R (Butler et al., 2009). The process of QTL detection included simple interval mapping to select cofactors, followed by composite interval mapping to identify QTL. The confidence interval was calculated with -log10(p-value) drop by 1.0 from the CIM profile. The empirical LOD threshold was determined by 1,000 permutation runs at a genome-wide error rate of $\alpha \leq 0.05$ for each trait, where the parental model was still used, but the heterogenous residual term (HRT) fitted by REML was replaced by default HRT for time-saving. The same was used in five-fold cross-validation to determine the LOD threshold with 1,000 runs and to obtain the QTL detection frequency.

For genome-wide association mapping, we used a model incorporating a kinship matrix and a fixed effect defining the biparental family to which a genotype belongs as described previously (Zhu et al., 2020). Multiple testing was controlled by a Bonferroni-corrected significance threshold ($p < .05$). The explained genotypic variance was calculated by fitting all significantly associated markers one by one in a linear model ($P_G$-Single) or together in a joint linear model ordered by the strength of their association ($P_G$-Joint). The allele substitution effect was obtained from the regression coefficient of the linear model when only one marker was considered.

3 | RESULTS

In this study, we performed QTL mapping for seed yield, plant height, protein yield and thousand-seed weight, using 944 RILs derived from a half-diallel experimental mating design of five typical early-maturing European soybean varieties. Six, five, seven and eight QTL were identified for these four traits, respectively (Figure 1, Table 1). Among them, the QTL qSY1/qPH2/qPY2 on chromosome 6 was found for the three traits seed yield, plant height and protein yield, and explained the highest proportion of genotypic variance within families, ranging from 21.21% to 59.21% for seed yield, 31.29% to 75.98% for plant height and 21.76% to 60.04% for protein yield. We identified another two QTL underlying these three traits, qSY3/qPH4/qPY5 on chromosome 18 and qSY4/qPH5/qPY6 on chromosome 19, with a relatively high proportion of explained genotypic variance for each trait. Interestingly, the estimated effects of the three QTL showed differences between families. There was also a high detection frequency for the QTL on chromosome 6 for all three traits and for the QTL on chromosomes 18 and 19 for plant height. Moreover the QTL qSY6/qTSW8 on chromosome 20 was found to control both seed yield and its component thousand-seed weight.

We also adopted multi-family QTL mapping as complementary strategy to compare with the results from single-family mapping (Figure 1, Table S1). We identified 12 QTL for seed yield with a global proportion of explained phenotypic variance ($R^2$) of 50.88%, 12 QTL for plant height with global $R^2$ of 68.39%, nine QTL for protein yield with global $R^2$ of 44.66% and 23 QTL for thousand-seed weight with global $R^2$ of 60.86%. Four, four, five, and six of the QTL for each of the four traits, respectively, were also identified by single-family mapping (Table 1). The QTL on chromosome 6 explained the largest $R^2$ with 17.11% for seed yield, followed by the QTL on chromosome 19 with $R^2$ of 10.59% and the QTL on chromosome 18 with 4.36% explained variance. The same three QTL were also the major QTL for plant height and protein yield, with the QTL on chromosome 6 always explaining the largest proportion of phenotypic variance. QTL underlying protein yield were all also identified for seed yield and the QTL on chromosome 20 also for protein content (Zhu et al., 2020). The number of identified QTL was highest for thousand-seed weight, with 23 QTL on 12 chromosomes, and the QTL on chromosome 20 explained the highest $R^2$ with 7.28%. In addition to the two linkage mapping approaches, we performed genome-wide association mapping. The positions of all of the identified QTL were either in the same regions or very close to the results from single-family mapping and multi-family mapping (Table S2–5, Figure S2).

Fine characterization was performed for the most interesting QTL on chromosomes 6, 18, 19 and 20 by plotting the results from single marker regression in individual families against the physical positions of the markers on the reference genome (Gmax Wm82. a2v1; Schmutz et al., 2010) (Figure 2, Figure S3). Regarding the three co-located QTL for plant height and seed yield, qSY1/qPH2 on chromosome 6 had a relatively broad peak in the genomic region around 18–24 Mb that coincides with the position of the maturity gene $E_1$. qSY3 and qPH4 are located in the same genomic region between 54–56 Mb on chromosome 18, with the peak at around 56 Mb, near the stem growth habit gene $DT_2$. qSY4 and qPH5 on chromosome 19 are located between 47 and 50 Mb, the chromosomal region encompassing the maturity gene $E_3$. As $E_1$, $E_3$ and $DT_2$ are known to affect maturity (Ping et al., 2014; Watanabe et al., 2009; Xia et al., 2012), we calculated the frequency of the seed yield-increasing allele of these three QTL in each of the three field trials (Figure S4). qSY1/qPH2, that likely corresponds to $E_1$, showed a clear trend with increasing frequency of the seed yield-increasing allele from early maturity genotypes (Trial 1) to latest maturity genotypes (Trial 3). Note, that early and late do not refer to the maturity group here, but to the classification of this population into three groups. For qSY3/qPH4 and qSY4/qPH5, by contrast, no such pattern was discernible. For the QTL pair qSY6/qTSW8 of seed yield and thousand-seed weight, the peak region on chromosome 20 was approximately between 27 and 32 Mb (Figure S3).
To obtain a more complete picture of QTL co-localization, we combined the QTL for the four traits from the present study with QTL for protein content and oil content identified previously in the same population (Zhu et al., 2020). The upset diagram visualizes that two QTL were found to control four traits, another two QTL affect three traits, and three QTL have effects on two of the traits (Figure 3). Most of the QTL were specific for one trait, with the largest number for thousand-seed weight.

Owing to the co-localization of the QTL on chromosome 6, 18, 19 and 20 between investigated traits, we chose the segregating marker nearest the QTL position in each of the eight families to investigate the allelic effect for the respective traits (Figure 4, Figure S5).
| Trait  | QTL  | Multi-family mapping       | Family                       | Chr. | Pos. in cM [CI] | pg [%] | Effect  | CV  |
|--------|------|---------------------------|------------------------------|------|----------------|--------|---------|-----|
| SY1    | Y    | Gallec × Sigalia [P1 × P5]| 6  123 [122–124]            | 21.21| 1.98           | 0.95   |
| SY2    | N    | Sultana × Sigalia [P4 × P5]| 7  91 [86–98]               | 16.30| −1.59          | 0.27   |
| SY3    | Y    | Gallec × Protina [P1 × P3]| 18 161 [159–165]            | 29.81| 1.89           | 0.48   |
| SY4    | Y    | Gallec × Sigalia [P1 × P5]| 19 120 [117–121]            | 31.69| −1.77          | 0.64   |
| SY5    | N    | Gallec × Sigalia [P1 × P5]| 19 130 [128–159]            | 31.64| −1.97          | 0.79   |
| SY6    | Y    | Gallec × Primus [P1 × P2]| 20  40 [38–42]              | 27.01| −1.33          | 0.93   |
| PH1    | N    | Primus × Protina [P2 × P3]| 5  104 [102–105]            | 20.49| 5.11           | 0.40   |
| PH2    | Y    | Gallec × Sigalia [P1 × P5]| 6  123 [121–123]            | 41.58| 15.04          | 1.00   |
| PH3    | Y    | Gallec × Protina [P1 × P3]| 16  15 [13–18]              | 18.58| −5.82          | 0.04   |
| PH4    | Y    | Gallec × Primus [P1 × P2]| 18 168 [166–169]            | 40.07| 7.66           | 0.79   |
| PH5    | Y    | Gallec × Protina [P1 × P3]| 19 121 [119–122]            | 21.60| −6.43          | 0.11   |
| PY1    | N    | Primus × Protina [P2 × P3]| 5  114 [111–116]            | 20.82| 0.40           | 0.27   |
| PY2    | Y    | Gallec × Sigalia [P1 × P5]| 6  123 [122–124]            | 21.76| 0.69           | 0.87   |
| PY3    | Y    | Gallec × Primus [P1 × P2]| 8  209 [207–210]            | 16.64| 0.37           | 0.37   |
| PY4    | Y    | Gallec × Protina [P1 × P3]| 16  8 [7–11]                | 22.39| 0.60           | 0.03   |
| PY5    | Y    | Gallec × Protina [P1 × P3]| 18 155 [153–159]            | 33.61| 0.85           | 0.52   |
| PY6    | Y    | Gallec × Protina [P1 × P3]| 19 120 [119–122]            | 31.33| −0.78          | 0.61   |
| PY7    | N    | Primus × Sultana [P2 × P4]| 20  91 [88–94]              | 16.82| 0.39           | 0.41   |
| TSW1   | N    | Primus × Protina [P2 × P3]| 6  32 [27–37]               | 23.42| −9.52          | 0.26   |
| TSW2   | N    | Protina × Sigalia [P3 × P5]| 6  97 [94–98]               | 16.97| −8.51          | 0.18   |
| TSW3   | Y    | Gallec × Sigalia [P1 × P5]| 9  63 [61–67]               | 17.09| −7.74          | 0.85   |
| TSW4   | Y    | Gallec × Sigalia [P1 × P5]| 13 110 [108–111]            | 18.51| 8.33           | 0.74   |
| TSW5   | Y    | Sultana × Sigalia [P4 × P5]| 13 127 [126–128]            | 25.60| 10.83          | 0.51   |
| TSW6   | Y    | Primus × Sigalia [P2 × P5]| 15  85 [84–87]              | 27.06| −8.82          | 0.69   |
| TSW7   | Y    | Gallec × Protina [P1 × P3]| 17 148 [146–155]            | 27.32| −10.22         | 0.76   |

(Continues)
and qPH2 on chromosome 6, presumed to be E1, were identified in three (P1 × P5, P2 × P5, P3 × P5) and four families (P1 × P5, P2 × P5, P3 × P5, P4 × P5), respectively. In family P2 × P5 the difference between the two QTL genotypic classes was 0.7 Mg/ha and 39.0 cm, and in family P3 × P5 0.6 Mg/ha and 41.6 cm. In family P1 × P5, these differences were less pronounced with 0.4 Mg/ha and 28.7 cm. In addition, a significant effect was observed in family P4 × P5, for which the QTL was not identified, but can be expected to segregate (Figure S6). As for qSY3/qPH4 on chromosome 18, the differences between genotypic groups within families in which this QTL was identified were 0.3 Mg/ha (P1 × P3) and 0.2 Mg/ha (P2 × P3) for seed yield and 15.7 cm (P1 × P2), 21.9 cm (P1 × P3) and 15.0 cm (P2 × P4) for plant height. Regarding qSY4 and qPH5 on chromosome 19, the genotypic differences were 0.3 Mg/ha (P1 × P5) for seed yield and 11.0 cm (P1 × P3) and 12.8 cm (P1 × P5) for plant height. These result further confirmed that the effect of the QTL on chromosome 6 was larger than that of the QTL on chromosomes 18 and 19, and in addition, that it showed the strongest variation among families. In general, for the identified QTL pairs the allele that increases seed yield also increases plant height. The QTL pair qSY6/qTSW8 was identified in two families (P1 × P2, P2 × P4) with comparable differences for both traits. This QTL has also recently been identified for protein content and oil content in these two families (Zhu et al., 2020). Further analyses revealed that the allele that increases seed yield and thousand-seed weight also increases oil content, but reduces protein content (Figure S5).

We further investigated the effect of QTL combinations on seed yield and plant height by classifying the whole population into eight groups based on the genotypic state at the three co-located QTL, qSY1/qPH2, qSY3/qPH4 and qSY4/qPH5 (Figure 5). The upward trend with an increasing number of the beneficial alleles for seed yield was observed for both traits. Plants carrying all three favourable alleles had the highest average seed yield, but were also amongst the tallest ones. The largest differences between groups were 0.9 Mg/ha for seed yield and 50.0 cm for plant height.

4 | DISCUSSION

The acreage of soybean cultivation in Europe is constantly increasing due to the high demand for soybean for human consumption and animal feed. In order to bridge the protein gap in Europe and for soybean to become economically competitive or even superior to other crops, soybean yield needs to be improved further. It is well known, that seed yield is a complex trait, controlled by many genes with small effect. An alternative is, therefore, the selection on yield component traits or other correlated traits in an indirect selection, as these often show a higher heritability and a less complex genetic architecture. In this study, we performed QTL mapping for the four traits seed yield, plant height, protein yield and thousand-seed weight, in order to understand the genetic control underlying these traits in early-maturing soybean germplasm, towards a marker-assisted breeding.

4.1 | Identification of QTL for the investigated traits

Single-family mapping identified QTL for all four traits, with most of them also being detected by multi-family mapping (Table 1, Table S1). Interestingly, the proportion of explained genotypic variance was rather high for some QTL, reaching up to 50% or higher in some families. Some QTL for the different traits are located in very close (no more than 1 cM distance) or overlapping QTL regions and can thus be regarded as the same or closely linked QTL. The most interesting examples of co-located QTL are qSY1/qPH2/qPY2 on chromosome 6, qSY3/qPH4/qPY5 on chromosome 18, and qSY4/qSY5/qPH5/qPY6 located on chromosome 19. Another QTL pair, qSY6/qTSW8, is located on chromosome 20 and was recently also identified as QTL for protein content and oil content (Zhu et al., 2020). While comparisons across studies using different population sizes, QTL confidence intervals and marker systems are difficult, some of the identified QTL might be identical to QTL reported in previous studies, for example qPH1 (Sun et al., 2006), qSY1, qPH2, qPY2 (Cao et al., 2019; Du et al., 2009; Liu et al., 2011; Specht et al., 2001; Wang et al., 2004), qSY3, qPH4, qPY5 (Cao et al., 2019; Kabelka et al., 2004; Kim et al., 2012; Sun et al., 2006), qSY4, qPH5 (Du et al., 2009; Orf, Chase, Jarvik, et al., 1999), qSY6 (Chung et al., 2003), qTSW1 (Mansur et al., 1996; Orf, Chase, Adler, et al., 1999; Specht et al., 2001), qTSW3 (Copley et al., 2018; Hyten et al., 2004; Kim et al., 2010), qTSW4 (Eskandari et al., 2013; Lee et al., 2001; Rossi et al., 2013; Teng et al., 2009), qTSW5 (Teng et al., 2004; Kim et al., 2010), qTSW6 (Rossi et al., 2013).
et al., 2009; Yan et al., 2014), qTSW6 (Kato et al., 2014; Li et al., 2008; Orf, Chase, Jarvik, et al., 1999), qTSW7 (Kato et al., 2014) and qTSW8 (Sebolt et al., 2000). Taken together, QTL mapping resulted in the identification of QTL for all four investigated traits, including seed yield. Some of these QTL had medium to large effects, which warrants further characterization.

### 4.2 | Fine characterization of the major QTL on chromosomes 6, 18, and 19

In this study, we found the maturity gene $E_1$ to be located in the genomic region of the major QTL on chromosome 6. $E_1$ has been cloned (Xia et al., 2012) and in addition to regulating photoperiodic...
response and thus maturity, it has been shown to affect plant height and node number of the main stem (Liu et al., 2011). Further support for E1 to underlie this QTL came from the analysis of the allele frequencies in the three trials, where the seed yield-increasing allele increased in frequency from the early to the late genotypes (Figure S4). This illustrates that it is not only a seed yield or plant height QTL, but related to maturity. In addition, the analysis of E1 for the five parents revealed allelic differences, all carrying photoperiod insensitive alleles as expected for these early maturity varieties, but P5 being e1-as, while P1-P4 are e1-nl (Kurasch, Hahn, Leiser, Starck, et al., 2017). e1-nl is reported as dysfunctional allele with a deletion encompassing the entire E1 gene while e1-as is presumed to be a weaker allele caused by a missense mutation that reduces the repression of E1 on flowering-inducing GmFT genes (Tsubokura et al., 2014; Xu et al., 2013). Consistently, this QTL for plant height occurred in all four families involving P5, but not in any of the other crosses (Figure S6). Overall, E1 is a very likely candidate gene for this major QTL on chromosome 6, affecting maturity, plant height, seed yield and protein yield.

For the QTL on chromosome 18, the growth habit gene Dt2 is a strong candidate (Figure 2). In soybean, the two genes Dt1 and Dt2 regulate stem growth habit in an epistatic interaction (Bernard, 1972; Lockhart, 2014; Ping et al., 2014). In the presence of wild-type Dt1, that maintains indeterminate growth, the dominant Dt2 allele results in semi-determinate plants. A potential advantage of indeterminate growth types in regions of higher latitude is the overlap of vegetative and generative growth stages, which allows higher yields under short growing seasons. Semi-determinate types are useful, as they produce fewer nodes compared to indeterminate types, making them less susceptible to lodging. As evident by the segregation of this locus, both alleles and thus growth types appear to be prevalent in early-maturing European soybean germplasm. Generally, the growth habit and flowering time with maturity can directly determine plant morphology like plant height, and thereby lead to a change in yield (Cao et al., 2017; Cober & Morrison, 2010; Xia et al., 2012; Zhang et al., 2015). Indeed, several studies reported higher yield to occur with larger plant height and later maturity (Cober & Morrison, 2010; Kabelka et al., 2004; Kim et al., 2012), and also in this population seed yield was highest in the latest maturing trial and significantly positively correlated with plant height. Interestingly, we identified this presumed Dt2 QTL in three families (P1 × P2, P1 × P3 and P2 × P4), indicating that parents P1 and P4 carry one allele and parents P2 and P3 the other allele (Figure S6). Consequently, the QTL can be expected to segregate in two out of the four families involving parent P5. However, the QTL was not detected in any of the four families with P5, that, as mentioned, is the only one to carry the e1-as allele. Only in families homozygous for e1-nl, did we identify this QTL. This indicates a dependency of the effect of Dt2 on the allelic state at E1. Dt2 encodes a MADS-box transcription factor and has recently been shown to affect a range of target genes that are involved in the regulation of different traits including flowering, maturity and water-use efficiency (Zhang et al., 2019). Particularly, it has also been shown to function as a direct activator of floral identity genes GmSOC1, GmAP1, and GmFUL, which are likely involved in flowering control. As photoperiod-insensitive alleles are often used in early-maturing soybean, this interplay between Dt1, Dt2 and E1 appears interesting and requires further research.

Another soybean maturity locus, E3 (Watanabe et al., 2009), was also found close to the peak marker of the major QTL on chromosome 19 (Figure 2). The stem growth habit gene E3 is also located on chromosome 19 at 45.18 Mb and thus a bit further apart from the QTL peak than E3. However, this QTL was also identified in families with identical Dt1 sequence, making it an unlikely candidate. In contrast to the presumed E1 QTL on chromosome 6, the frequency of the alleles at this QTL showed no change between the three maturity trials (Figure S4). Notably, this does not rule out E3 as the gene
underlying this QTL and may be caused by a more complex interplay between the different maturity loci and a stronger effect of the E1 alleles in this germplasm and these environments.

In conclusion, the known phenology genes E1, E3 and Dt2 are likely candidates to underlie the major QTL identified here, illustrating their strong pleiotropic effects on a range of agronomically important traits, including seed yield.

4.3 Characterization of pleiotropic QTL

Our previous phenotypic analysis of this population revealed a strong positive correlation between seed yield, plant height and protein yield (Kurasch, Hahn, Leiser, Starck, et al., 2017). In line with this, we also found the abovementioned three medium- to major-effect QTL to affect all three traits, illustrating an at least in part shared genetic
control. In total, seven QTL with effects on more than one trait were identified (Figure 3). Such co-located QTL can be due to pleiotropic effects of a single QTL, as illustrated by the example of the maturity gene E1, or by closely linked QTL. From a breeding perspective, the latter are not necessarily different from the former, as depending on the distance of the QTL and the recombination rate in that chromosomal region, linked QTL can be close to impossible to separate.

Regarding the utilization of marker-assisted selection, it is important to know the effect of the different alleles on each of the target traits. Is the one allele the favourable allele for both or all target traits, or is one allele positive for one trait but negative for another trait. We therefore assessed the effects of alleles of co-located QTL on the target traits in all eight families (Figure 4, Figure S5).

$q_{SY1}/q_{PH2}$, the major QTL on chromosome 6, resulted in a difference of 0.3–0.7 Mg/ha for seed yield and 28.7–41.6 cm for plant height between genotypic groups in four families, illustrating a certain dependency of this QTL on the genetic background. Thus, the allele carried by parent P5 (presumably $e_{1-as}$), resulted in taller plants with a higher yield under these environmental conditions, compared to the other allele (presumed $e_{1-nl}$). The QTL $q_{SY3}/q_{PH4}$ and $q_{SY4}/q_{PH5}$ appear to be more stable across genetic backgrounds, and also have positive effects on both seed yield and plant height. It has been reported that plant height is positively associated with yield (Liu et al., 2011; Lü et al., 2017), but some studies also reported contrasting results (Contreras-Soto et al., 2017). Taller plants, however, are more prone to lodging, which should also be considered in breeding programmes.

Another interesting example is the QTL on chromosome 20, affecting seed yield and thousand-seed weight, as well as protein content and oil content. We have recently reported that this QTL was identified in four families, with an inverse effect on protein content and oil content. The allele that increases protein content by 1% reduces oil content by approximately 0.4% in each family (Zhu et al., 2020). Here, we show that the effects of this QTL on seed yield, plant height and oil content are positively correlated (Figure S5). This illustrates that selection for the higher seed yield allele at this QTL will come at the expense of a lower protein content. This may be compensated by a higher protein yield and in addition, could be counteracted by selection for other protein content-increasing QTL. However, it would be disadvantageous if a high protein content in the seed is required for specialty soybean processing. In conclusion, our results illustrate the complex interplay between agronomically important traits, which is also reflected by pleiotropically acting QTL. In breeding programmes, index selection maybe a good strategy to simultaneously improve several target traits.

4.4 | QTL stacking for seed yield and effects on plant height

Regarding the quantitative nature of seed yield, stacking of favourable seed yield QTL alleles might be an effective way in marker-assisted soybean breeding programmes. The three pleiotropic QTL on chromosomes 6, 18 and 19 were selected to observe their combined effects on seed yield and plant height in the whole population. Owing to the similar sign of the effects on both traits, both showed an increase with increasing number of favorable seed yield alleles (Figure 5). Thus, marker-assisted selection for higher seed yield inevitably results in the selection of taller and later-maturing plants. These effects of selection on pleiotropically acting QTL may be balanced by the already mentioned index selection.

Based on the successful experience from rice and maize, Liu et al. (2020) purposed that an ideal plant architecture was essential for a ‘green revolution’ of soybean. This not only includes an appropriate plant height, but also other characteristics such as shorter internode lengths, more internodes, fewer branches and others. Thus, a consequent next step is to dissect the genetic control of such plant architectural traits, which may be aided by soybean functional
genomics. Collectively, our results and the characterization of identified QTL illustrate the potential of marker-assisted selection, but also its limitations, highlighting that selection must always consider all target traits in order to identify superior genotypes as our future varieties.

ACKNOWLEDGEMENTS

This work was funded by the German Federal Ministry of Food and Agriculture (BMEL, grants numbers 2814500110, 2814EPS011) and by the Federal Ministry of Education and Research of Germany (BMBF, grant numbers 031B033A9, 031B0339B).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTION

TW and VH designed the study. VH and WLL collected phenotypic and genotypic data. XZ performed the analyses. XZ, VH, WLL and TW wrote the paper.

ORCID

Xintian Zhu https://orcid.org/0000-0001-8597-2762
Willmar L. Leiser https://orcid.org/0000-0001-5964-5610
Volker Hahn https://orcid.org/0000-0003-4837-1460
Tobias Würschum https://orcid.org/0000-0002-7397-7731

REFERENCES

Baierl, A., Bogdan, M., Frommlet, F., & Futschik, A. (2006). On locating multiple interacting quantitative trait loci in intercross designs. Genetics 173: 1693–1703. Genetics, 173, 1693–1703. https://doi.org/10.1534/genetics.105.048108
Bernard, R. L. (1972). Two Genes Affecting Stem Termination in Soybeans. Crop Science, 12(2), 235–239. https://doi.org/10.2135/cropsci1972.0011183x00120000208x
Blanc, G., Charcosset, A., Mangin, B., Gallais, A., & Moreau, L. (2006). Connected populations for detecting quantitative trait loci and testing for epistasis: An application in maize. Theoretical and Applied Genetics, 113(2), 206–224. https://doi.org/10.1007/s00122-006-0287-1
Broman, K. W., Wu, H., Sen, Š., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics, 19(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112
Burris, J. S., Edje, O. T., & Wahab, A. H. (1973). Effects of seed size on seedling performance in soybeans: II. seedling growth and photosynthesis and field performance. Crop Science, 13(2), 207–210. https://doi.org/10.2135/cropsci1973.0011183x001300022017x
Butler, D. G., Cullis, B. R., Gilmour, A. R., & Gogel, B. J. (2009). ASReml-R reference manual. The State of Queensland, Department of Primary Industries and Fisheries. https://asreml.kb.vsnl.co.in/wp-content/uploads/sites/3/2018/02/ASReml-R-3-Reference-Manual.pdf
Cao, Y., Li, S., Chen, G., Wang, Y., Bhat, J. A., Karikari, B., Kong, J., Gai, J., & Zhao, T. (2019). Deciphering the genetic architecture of plant height in soybean using two RIL populations sharing a common M8206 parent. Plants, 8(10), 1–22. https://doi.org/10.3390/plants8100373
Cao, Y., Li, S., He, X., Chang, F., Kong, J., Gai, J., & Zhao, T. (2017). Mapping QTLs for plant height and flowering time in a Chinese summer planting soybean RIL population. Euphytica, 213(2), 39. https://doi.org/10.1007/s10681-016-1934-8
Chang, F., Guo, C., Sun, F., Zhang, J., Wang, Z., Kong, J., He, Q., Sharmin, R. A., & Zhao, T. (2018). Genome-wide association studies for dynamic plant height and number of nodes on the main stem in summer sowing soybeans. Frontiers in Plant Science, 9, 1–13. https://doi.org/10.3389/fpls.2018.01184
Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., Shoemaker, R. C., & Specht, J. E. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. Crop Science, 43(3), 1053–1067. https://doi.org/10.2135/cropsci2003.1053
Cober, E. R., & Morrison, M. J. (2010). Regulation of seed yield and agronomic characters by photoperiod sensitivity and growth habit genes in soybean. Theoretical and Applied Genetics, 120(5), 1005–1012. https://doi.org/10.1007/s00122-009-1228-6
Contreras-Soto, R. I., Mora, F., de Oliveira, M. A. R., Higashi, W., Scapim, C. A., & Schuster, I. (2017). A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype type analysis. PLoS One, 12(2), e0171105. https://doi.org/10.1371/journal.pone.0171105
Copley, T. R., Duceppe, M. O., & O’Donnoghue, L. S. (2018). Identification of novel loci associated with maturity and yield traits in early maturity soybean plant introduction lines. BMC Genomics, 19(1), 1–12. https://doi.org/10.1186/s12864-018-4558-4
Darvasi, A., & Soller, M. (1997). A simple method to calculate resolving power and confidence interval of QTL map location. Behavior Genetics, 27(2), 125–132. https://doi.org/10.1023/A:1025685324830
Du, W., Wang, M., Fu, S., & Yu, D. (2009). Mapping QTLs for seed yield and drought susceptibility index in soybean (Glycine max L.) across different environments. Journal of Genetics and Genomics, 36(12), 721–731. https://doi.org/10.1016/S1673-8527(08)60165-4
Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One, 6(5), 1–10. https://doi.org/10.1371/journal.pone.0019379
Eskandari, M., Cober, E. R., & Rajcan, I. (2013). Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration with- out decreasing protein or with increased seed yield. Theoretical and Applied Genetics, 126(6), 1677–1687. https://doi.org/10.1007/s00122-013-2083-z
FAOSTAT (2020). FAOSTAT statistical database. Nations, Food and Agriculture Organization of the United website. https://search.library.wisc.edu/catalog/999890171702121
Fasoula, V. A., Harris, D. K., & Boerma, H. R. (2004). Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Science, 44(12), 1218–1225. https://doi.org/10.2135/cropsci2004.1218
Fox, C. M., Cary, T., Nelson, R., & Diers, B. W. (2015). Confirmation of a seed yield QTL in soybean. Crop Science, 55(3), 992–998. https://doi.org/10.2135/cropsci2014.10.0688
Gai, J., Wang, Y., Wu, X., & Chen, S. (2007). A comparative study on segregation analysis and QTL mapping of quantitative traits in plants— with a case in soybean. Frontiers in Agriculture in China, 1(1), 1–7. https://doi.org/10.1007/s11703-007-0001-3
Garin, V., Wimmer, V., Mezmouk, S., Malosetti, M., & van Eeuwijk, F. (2017). How do the type of QTL effect and the form of the residual term influence QTL detection in multi-parent populations? A case study in the maize EU-NAM population. Theoretical and Applied Genetics, 130(8), 1753–1764. https://doi.org/10.1007/s00122-017-2923-3
Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. (2009). SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research, 38(suppl_1), D843–D846. https://doi.org/10.1093/nar/gkp798
Guzman, P. S., Diers, B. W., Neece, D. J., St. Martin, S. K., LeRoy, A. R., Grau, C. R., Hughes, T. J., & Nelson, R. L. (2007). QTL associated with yield in three backcross-derived populations of soybean. Crop Science, 47(1), 111–122. https://doi.org/10.2135/cropsci2006.01.0003
Hahn, V., & Würschum, T. (2014). Molecular genetic characterization of 17 European soybean breeding germplasm. Plant Breeding, 133(6), 748–755. https://doi.org/10.1111/pbr.12212

Han, Y., Li, D., Zhu, D., Li, H., Li, X., Teng, W., & Li, W. (2012). QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. Theoretical and Applied Genetics, 125(4), 671–683. https://doi.org/10.1007/s00122-012-1859-x

Hirata, K., Masuda, R., Tsukbura, Y., Yasui, T., Yamada, T., Takahashi, K., Nagaya, T., Sayama, T., Ishimoto, M., & Hajiaka, M. (2014). Identification of quantitative trait loci associated with boiled seed hardness in soybean. Breeding Science, 64(4), 362–370. https://doi.org/10.1270/jsbbs.64.362

Hnetkovsky, N., Chang, S. J. C., Doublter, T. W., Gibson, P. T., & Lightfoot, D. A. (1996). Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS). Crop Science, 36(2), 393–400. https://doi.org/10.2138/cropsci.1996.0011183X0036000020030x

Hyten, D. L., Pantalone, V. R., Sams, C. A., Saxton, A. M., Landau- Ellis, D., Hnetkovsky, N., Chang, S. J. C., Doubler, T. W., Gibson, P. T., & Lightfoot, D. A. (1996). Phenotypic analysis of major agronomic traits in 1008 RILs from a diallel of early European soybean varieties. Plant, Cell & Environment, 19(2), 276–287. https://doi.org/10.1111/jepe.12581

Kurash, A. K., Hahn, V., Leiser, W. L., Starck, N., & Würschum, T. (2017). Phenotypic analysis of major agronomic traits in 1008 RILs from a diadel of early European soybean varieties. Crop Science, 57(2), 726–738. https://doi.org/10.2135/cropsci2016.05.0318

Kurash, A. K., Hahn, V., Leiser, W. L., Vollmann, J., Schori, A., Bétrix, C.-A., & Würschum, T. (2017). Identification of mega-environments in Europe and effect of allelic variation at maturity E loci on adaptation of European soybean. Plant, Cell & Environment, 40(5), 765–778. https://doi.org/10.1111/pce.12896

Kurash, A. K., Hahn, V., Miersch, M., Bachteler, K., & Würschum, T. (2018). Analysis of tofu-related traits by a bench-scale tofu production method and their relationship with agronomic traits in European soybean. Plant Breeding, 137(3), 271–282. https://doi.org/10.1111/pbr.12581

Kurash, A. K., Leiser, W. L., Bachteler, K., Miersch, M., Hahn, V., & Würschum, T. (2018). Evaluation of the genetic architecture of tofu traits in soybean towards genomics-assisted breeding. Plant Breeding, 137(6), 873–882. https://doi.org/10.1111/pbr.12651

Lee, S. H., Park, K. Y., Lee, H. S., Park, E. H., & Boerma, H. R. (2001). Genetic mapping of QTLs conditioning soybean sprout yield and quality. Theoretical and Applied Genetics, 103(5), 702–709. https://doi.org/10.1007/s001220000595

Li, R., Jiang, H., Zhang, Z., Zhao, Y., Xie, J., Wang, Q., Zheng, H., Hou, L., Xiong, X., Xin, D., Hu, Z., Liu, C., Wu, X., & Chen, Q. (2020). Combined linkage mapping and bsa to identify qtl and candidate genes for plant height and the number of nodes on the main stem in soybean. International Journal of Molecular Sciences, 21(1), 42. https://doi.org/10.3390/ijms21010042

Li, W., Zheng, D., Van, K., & Lee, S. (2008). QTL mapping for major agronomic traits across two years in soybean (Glycine max L. Merr.). Journal of Crop Science and Biotechnology, 11(3), 171–190.

Liu, S., Zhang, M., Feng, F., & Tian, Z. (2020). Toward a ‘Green Revolution’ for Soybean. Molecular Plant, 13(5), 688–697. https://doi.org/10.1016/j.molp.2020.03.002

Liu, W., Kim, M. Y., Van, K., Lee, Y.-H., Li, H., Liu, X., & Lee, S.-H. (2011). QTL identification of yield-related traits and their association with flowering and maturity in soybean. Journal of Crop Science and Biotechnology, 14(1), 65–70. https://doi.org/10.1007/s12892-010-0115-7

Lockhart, J. (2014). Finding D2t, the dominant gene that specifies the semideterminate growth habit in soybean. The Plant Cell, 26(7), 2725. https://doi.org/10.1105/tpc.114.129056

Lü, H. Y., Li, H. W., Fan, R., Li, H. Y., Yin, J. Y., Zhang, J. J., & Zhang, D. (2017). Genome-wide association study of dynamic developmental plant height in soybean. Canadian Journal of Plant Science, 97(2), 308–315. https://doi.org/10.1139/cjps-2016-0152

Lu, X., Xiong, Q., Cheng, T., Li, Q.-T., Liu, X.-L., Bi, Y.-D., Li, W., Zhang, W.-K., Ma, B., Lai, Y.-C., Du, W.-G., Man, W.-Q., Chen, S.-Y., & Zhang, J.-S. (2017). A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. Molecular Plant, 10(5), 670–684. https://doi.org/10.1016/j.molp.2017.03.006

Mansur, L. M., Orf, J. H., Chase, K., Jarvik, T., Cregan, P. B., & Lark, K. G. (1996). Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Science, 36(5), 1327–1336. https://doi.org/10.2138/cropsci1996.0011183X0036000050042x

Mian, M. A. R., Bailey, M. A., Tamulonis, J. P., Shiue, E. R., Carter, T. E., Parrott, W. A., Ashley, D. A., Hussey, R. S., & Boerma, H. R. (1996). Molecular markers associated with seed weight in two soybean populations. Theoretical and Applied Genetics, 93(7), 1011–1016. https://doi.org/10.1007/BF00230118

Orf, J. H., Chase, K., Adler, F. R., Mansur, L. M., & Lark, K. G. (1999). Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. Crop Science, 39(6), 1652–1657. https://doi.org/10.2138/cropsci1999.3961652x

Orf, J. H., Chase, K., Jarvik, T., Mansur, L. M., Cregan, P. B., Adler, F. R., & Lark, K. G. (1999). Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Science, 39(6), 1642–1651. https://doi.org/10.2138/cropsci1999.3961642x

Piepho, H. P., & Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. Genetics, 177, 1881–1888. https://doi.org/10.1534/genetics.107.074229

Ping, J., Liu, Y., Sun, L., Zhao, M., Li, Y., She, M., & Ma, J. (2014). D2t2 is a gain-of-function MADS-domain factor gene that specifies semideterminate in soybean. The Plant Cell, 26(7), 2831–2842. https://doi.org/10.1105/tpc.114.126938

Rosi, M. E., Orf, J. H., Liu, L.-J., Dong, Z., & Rajcan, I. (2013). Genetic basis of soybean adaptation to North American vs. Asian mega-environments in two independent populations from Canadian × Chinese crosses. Theoretical and Applied Genetics, 126(7), 1809–1823. https://doi.org/10.1007/s00122-013-2094-9

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hestlen, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharya, M. K., Sandhu, D., Valliyodan, B., ... Jackson, S. A. (2010). Genome sequence
