THE CROP JOURNAL XX (2017) XXX–XXX

Available online at www.sciencedirect.com

ScienceDirect

# Development and validation of InDel markers for identification of QTL underlying flowering time in soybean☆

*Jialin Wang*[a,1], *Lingping Kong*[a,e,1], *Kanchao Yu*[c,f,1], *Fengge Zhang*[a,e,1], *Xinyi Shi*[a,1], *Yanping Wang*[d], *Haiyang Nan*[a], *Xiaohui Zhao*[a,b], *Sijia Lu*[a,b], *Dong Cao*[a], *Xiaoming Li*[a,e], *Chao Fang*[a,e], *Feifei Wang*[a,e], *Tong Su*[a,e], *Shichen Li*[a,e], *Xiaohui Yuan*[a,*], *Baohui Liu*[a,b,**], *Fanjiang Kong*[a,b,**]

[a]*The Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin 150081, Heilongjiang, China*
[b]*School of Life Sciences, Guangzhou University, Guangzhou 510006, Guangdong, China*
[c]*Qiqihar Branch of Heilongjiang Academy of Agricultural Sciences, Qiqihar 161006, Heilongjiang, China*
[d]*Mudanjiang Branch of Heilongjiang Academy of Agricultural Sciences, Mudanjiang 157041, Heilongjiang, China*
[e]*University of Chinese Academy of Sciences, Beijing 100049, China*
[f]*College of Agriculture, Northeast Agricultural University, Harbin 150030, Heilongjiang, China*

## ARTICLE INFO

*Article history:*
Received 31 March 2017
Received in revised form 10 July 2017
Accepted 16 August 2017
Available online xxxx

*Keywords:*
Soybean
Resequencing data
InDel markers
Genetic map
QTL analysis

## ABSTRACT

Soybean [*Glycine* max (L.) Merrill] is a major plant source of protein and oil. An accurate and well-saturated molecular linkage map is a prerequisite for forward genetic studies of gene function and for modern breeding for many useful agronomic traits. Next-generation sequence data available in public databases provides valuable information and offers new insights for rapid and efficient development of molecular markers. In this study, we attempted to show the feasibility and facility of using genomic resequencing data as raw material for identifying putative InDel markers. First, we identified 17,613 InDel sites among 56 soybean accessions and obtained 12,619 primer pairs. Second, we constructed a genetic map with a random subset of 2841 primer pairs and aligned 300 polymorphic markers with the 20 consensus linkage groups (LG). The total genetic distance was 2347.3 cM and the number of mapped markers per LG ranged from 10 to 23 with an average of 15 markers. The largest and smallest genetic distances between adjacent markers were 52.3 cM and 0.1 cM, respectively. Finally, we validated the genetic map constructed by newly developed InDel markers by QTL analysis of days to flowering (DTF) under different environments. One major QTL (*qDTF4*) and four minor QTL (*qDTF20*, *qDTF13*, *qDTF12*, and *qDTF11*) on 5 LGs were detected. These results demonstrate the utility of the InDel markers developed in this work for map-based cloning and molecular breeding in soybean.



☆ Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

* *Corresponding author.*
** *Corresponding authors at: The Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin 150081, Heilongjiang, China.*
*E-mail addresses:* yuanxh@iga.ac.cn (X. Yuan), liubh@iga.ac.cn (B. Liu), kongfj@iga.ac.cn (F. Kong).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cj.2017.08.001


Please cite this article as: J. Wang, et al., Development and validation of InDel markers for identification of QTL underlying flowering time in soybean, The Crop Journal (2017), https://doi.org/10.1016/j.cj.2017.08.001

## 1. Introduction

Soybean [*Glycine* max (L.) Merrill] is a globally important crop that provides a steady source of high-quality vegetable protein and oil for food products and industrial materials. Accordingly, many useful agronomic trait loci associated with growth, product quality, tolerance to biotic and abiotic stresses, and other characteristics have been identified in the past few decades. Accurate and well-saturated genetic linkage maps have become a valuable tool for genetics and plant breeding, as have genome assembly, QTL analysis, gene tagging, and marker-assisted selection (MAS). Since the first genetic map of soybean was constructed with phenotypic traits [1], various types of molecular markers including restriction fragment length polymorphism (RFLP), random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeat (SSR) markers have been used to construct linkage maps [2–5]. Although the current BARCSOYSSR_1.0 database has a total of 33,065 SSR primer sets and the average density of SSR loci in the whole genome is one SSR marker per 0.072 cM [6], many genomic intervals contain no SSR markers, and they are not sufficient for positional cloning and fine mapping in all possible parental crosses.

A functional gene can be identified via forward and reverse genetics strategies [7,8]. Positional cloning is widely used as a forward genetics approach to isolate genes in different organisms [9], and its utility can be fully exploited in modern molecular plant breeding systems, such as corn and soybean, when markers linked to genes of interest are discovered [10]. The principle of positional cloning is to systematically narrow down the genetic interval containing a causal mutation by sequentially excluding all other regions in the genome [11]. All rely on the development of highly dense genetic markers that are polymorphic between the accessions used for generating the mapping population(s) to provide adequate mapping resolution. This dependence is a major limiting factor for the rate of mapping progress.

With the decreasing cost of next-generation sequencing, there have been several proposals to exploit single-nucleotide polymorphisms (SNPs) and Insertion/Deletions (InDels) for genetic mapping with high-density markers. In contrast to SNPs, InDel polymorphisms, another form of natural genetic variation, have received relatively little attention. Mechanisms such as transposable elements, slippage in simple sequence replication, and unequal crossover events can result in the formation of InDels [12]. They can be converted to a user-friendly marker type, show high variation and codominant inheritance, and are relatively abundant and uniformly distributed throughout the genome [13,14]. InDel markers are PCR-based and readily genotyped by fragment length polymorphism with minimal laboratory equipment. Recently InDel markers have been widely applied for genotyping, genetic diversity analysis, QTL mapping, map-based cloning, and even marker-assisted selection in *Arabidopsis*, rice, wheat, turnip, sunflower, pepper, sesame, cotton, and citrus [14–27]. However, InDel markers have seldom been identified and used in soybean. A recent study used 73,327 InDels in six soybean cultivars to build a soybean barcode system for comparing data from different sources [28]. In another study, 165 validated InDel markers were used to develop an InDel-based linkage map for a mapping population between Hedou 12 and Williams 82 [29]. By exploiting the reference genome sequence of soybean and the large amount of intensive resequencing data available in public databases [30–35], it is now possible to detect genome-wide InDel polymorphisms amongst different accessions using whole-genome resequencing to guide rapid and efficient development of InDel markers for high-resolution genetic analysis.

In this study, we attempted to develop InDel markers using genomic resequencing data using a series of bioinformatic approaches. In total, these methods yielded 12,619 new markers that were variously polymorphic amongst 56 soybean accessions. An InDel-based genetic map of soybean was constructed with 300 polymorphic InDel markers. QTL analysis was performed to identify genomic regions associated with flowering time. One major QTL (*qDTF4*) was identified in 2015 and confirmed in 2016. The InDel markers, genetic map, and QTL identified in this study will lay a foundation for the genetic/QTL analysis and isolation of genes underlying variation in flowering time and provide useful information for MAS breeding in soybean.

## 2. Materials and methods

### 2.1. Plant materials and trait evaluation

The $F_{7:8}$ seeds for the mapping populations were grown in walk-in plant growth chambers at 22 °C, 65% relative humidity, and long-day (LD) photoperiod (16 h light/8 h dark) in October 2015 and in the field in Harbin (45°43′ N, 126°45′ E) and Mudanjiang (44°36′ N, 129°35′ E), China in May 2016.

Days to flowering were recorded at the R1 stage (days from emergence to first open flower appearing on 50% of plants). For chamber experiments, seeds from each line were sown in pots. After germination, the seedlings were thinned until each pot contained five uniform plants. Populations were sown in the field with a single seed every 20 cM in 5-m rows spaced 60 cM apart and 25 seeds per line. All trials received standard cultural practices to control insects and weeds.

### 2.2. Mapping populations and sequence data sets

The BA population, derived from a cross between Mufu12-604 × HB-2 and consisting of 156 $F_2$ genotypes, was used to test the newly developed markers and construct a high-density InDel linkage map. The DW population (144 RILs), derived from a cross between Dongnong 50 (early-flowering in LD photoperiod) and Williams 82 (late-flowering in LD photoperiod), was used to evaluate the InDel markers for QTL mapping.

Fifty six accessions, including 29 from three recent research papers and 27 from this study, were used for InDel polymorphism validation (Table 1). Young leaves from 27 accessions were collected three weeks after planting in

Please cite this article as: J. Wang, et al., Development and validation of InDel markers for identification of QTL underlying flowering time in soybean, The Crop Journal (2017), https://doi.org/10.1016/j.cj.2017.08.001