# Identification of a Key Major-effect QTL Associated with Pre-harvest Sprouting in Cucumber (*Cucumis sativus* L.) Using the QTL-Seq Method

**Mingming Cao[1], Shuju Li[1*], Qiang Deng[1], Huizhe Wang[2], Ruihuan Yang[2]**

[1]Tianjin Kernel Cucumber Research Institute, Tianjin 300192, China

[2]Institute of Cucumber Research, Tianjin Academy of Agricultural Sciences, Tianjin 300192, China

*** Correspondence:**

Shuju Li

lishuju1964@126.com

Email address :

Mingming Cao: caoming2013@126.com

Shuju Li: lishuju1964@126.com

Qiang Deng: dengqiang022@126.com

Huizhe Wang: wanghuizhe@126.com

Ruihuan Yang: yruihuan@126.com

## ABSTRACT

**Background:** Cucumber (*Cucumis sativus* L.) is cultivated worldwide, and it is essential to produce enough high-quality seeds to meet demand. Pre-harvest sprouting (PHS) in cucumber is a critical problem and causes serious damage to seed production and quality. Nevertheless, the genetic basis and molecular mechanisms underlying cucumber PHS remain unclear. QTL-seq is an efficient approach for rapid quantitative trait loci (QTL) identification that simultaneously takes advantage of bulked-segregant analysis (BSA) and whole-genome resequencing. In the present research, QTL-seq analysis was performed to identify QTLs associated with PHS in cucumber using an $F_2$ segregating population.

**Results:** Two QTLs that spanned 7.3 Mb on Chromosome 4 and 0.15 Mb on Chromosome 5 were identified by QTL-seq and named *qPHS4.1* and *qPHS5.1*, respectively. Subsequently, SNP and InDel markers selected from the candidate regions were used to refine the intervals using the extended $F_2$ populations grown in the 2016 and 2017 seasons. Finally, *qPHS4.1* was narrowed to 0.53 Mb on chromosome 4 flanked by the markers SNP-16 and SNP-24 and was found to explain 19-22% of the phenotypic variation in cucumber PHS. These results reveal that *qPHS4.1* is the key major-effect QTL associated with PHS in cucumber. Based on gene annotations and qRT-PCR expression analyses, *Csa4G622760* and *Csa4G622800* were proposed as the candidate genes.

**Conclusions:** These results provide novel insights into the genetic mechanism controlling PHS in cucumber and highlight the potential for marker-assisted selection of PHS resistance breeding.

**Keywords:** Cucumber; Pre-harvest Sprouting; QTL-seq; *qPHS4.1*

## INTRODUCTION

Cucumber (*Cucumis sativus* L.) is an economically important vegetable globally. In 2018, cucumber was grown on 1,984,518 hectares worldwide, and the cultivated area in China accounted for 52.72% of this area (www.fao.org/faostat/en). It is necessary to produce enough excellent-quality cucumber seeds to meet demand, especially in China. However, pre-harvest sprouting (PHS), also known as vivipary, a critical trait describing the untimely germination of seeds inside maternal fruits under certain conditions, severely decreases seed yields and quality. Breeding for resistance to PHS in cucumber is necessary.

In agriculture, it is widely accepted that PHS is a complex agronomic trait controlled by multiple genes or quantitative trait loci (QTLs) [1,2] and tightly connected with seed dormancy, which is characterized as the prevention of physiologically mature seeds from germinating under unfavorable environmental conditions [3,4]. Low levels of seed dormancy lead to PHS [5], while excessive seed dormancy usually gives rise to PHS resistance but unfortunately causes undesirable results, such as nonuniform seedling establishment after sowing [6,7]. Therefore, maintenance of the balance between seed dormancy and germination is critical.

Regarding the genetic and molecular basis of seed dormancy and PHS resistance, extensive QTLs or genes for this trait have been identified in cereal crops and other vegetables, such as rice (*Oryza sativa*), wheat (*Triticum aestivum*), maize (*Zea mays*), barley (*Hordeum vulgare*) and tomato (*Solanum lycopersicum*). To date, in rice, more than 165 QTLs associated with seed dormancy or PHS resistance and located on different chromosomes have been identified [8,9]. Similar to rice, QTLs responsible for PHS identified in wheat, which has a much more complicated genome, were distributed on almost all of the chromosomes [10]. Among them, the

major QTLs were detected mainly on chromosome 2B [11], 3AS [12], and 7B [13], while minor QTLs were detected on chromosomes 3B and 5A [12]. In barley, several QTLs associated with seed dormancy have been identified [14,15,16]. Among the QTLs, two QTLs, *SD1* and *SD2* on chromosome 5H, contributed the major effects on seed dormancy [17]. *SD1* was a major regulator of dormancy [18], and *SD2* was identified to prevent PHS [16]. Abscisic acid (ABA) plays a critical role in seed dormancy and germination. The expression of genes involved in ABA biosynthesis or signaling regulates seed dormancy and PHS. Among these genes, *ABSCISIC ACID-INSENSITIVE3* (*ABI3*) is an important gene in the ABA signaling pathway [19]. For maize, a series of *viviparous* genes, e.g., *Vp1*, *Vp5*, *Vp7*, *Vp10/Vp13*, *Vp14* and *Vp15*, have been cloned [10]. The *VP1* gene is an ortholog of *ABI3* in Arabidopsis and is associated with seed dormancy and PHS in maize. *Vp5* and *Vp7* were also found to be necessary for ABA biosynthesis. Mutants of *Vp5* and *Vp7* decreased endogenous ABA biosynthesis and promoted PHS in maize [20]. In tomato, the *SlDET1* (*METHYLTRANSFERASE1*) gene participates in regulating the expression of ABA biosynthesis/response genes. Overexpression of *SlNCED* increases the level of ABA and inhibits seed germination, while silencing of *SlNCED* causes ABA deficiency and leads to vivipary in tomato [21]. However, to date, QTL genetic mapping for PHS in cucumber has not been reported.

Traditional QTL mapping requires a segregating population originating from two parents with extreme opposite traits and polymorphic markers linked to target genes. It is extremely time-consuming and labor-intensive to screen DNA markers and genotype individuals in the segregating populations [22]. Bulked-segregant analysis (BSA) is an effective method to rapidly identify polymorphic markers linked to traits of interest [23]. QTL-seq [24], a powerful new approach combining BSA and next-generation sequencing, is used for the rapid identification of QTLs. Recently, QTL-seq has been widely used in the detection of QTLs for many traits in various plants, including 100-seed weight trait in chickpea [25], branch angle in oilseed rape [26], fruit length in cucumber [27], stalk rot in maize [28], heat-tolerance and high-temperature stress response in tomato [29], and cooked grain elongation [30] and salt tolerance [31] in rice. Therefore, QTL-seq provides a convenient method for identifying key loci controlling PHS in cucumber.

Our previous studies have revealed the inheritance of PHS in cucumber, but genetic mapping and QTL location have not been performed. In this paper, we performed QTL-Seq analysis using an $F_2$ population derived from Q12 and P60, which were typically resistant and susceptible to PHS in cucumber, respectively. SNP and InDel markers generated from QTL-seq were developed to genotype all the individuals in the $F_2$ population grown in two years. The major QTLs were refined, and annotated genes located in the associated regions were analyzed by quantitative RT-PCR. This study may have the potential for cucumber breeding of PHS resistance by marker-assisted selection (MAS) and gene cloning analysis.

## RESULTS

### Phenotypic Evaluation of PHS in cucumber

Phenotypic data of the PHS rate were collected from Q12, P60, and their $F_1$, $F_2$ populations in the greenhouse (Additional file 1: Table S1). The mean PHS rates of the resistant parent Q12, susceptible parent P60 and $F_1$ progeny were 0%, 64.97% and 13.88%, respectively (Table 1). Q12 showed a significantly ($P < 0.01$) lower PHS rate than P60. The PHS rates of the segregating

mapping population of 328 $F_2$ individuals grown in 2016 covered the full range from 0 to 100% (20.77% on average) and showed a skewed normal distribution (Fig. 1). The PHS rates of the 298 $F_2$ individuals grown in 2017 showed a similar distribution. This phenotypic variation in the $F_2$ population indicated that PHS is a quantitative trait controlled by a key major-effect QTL.

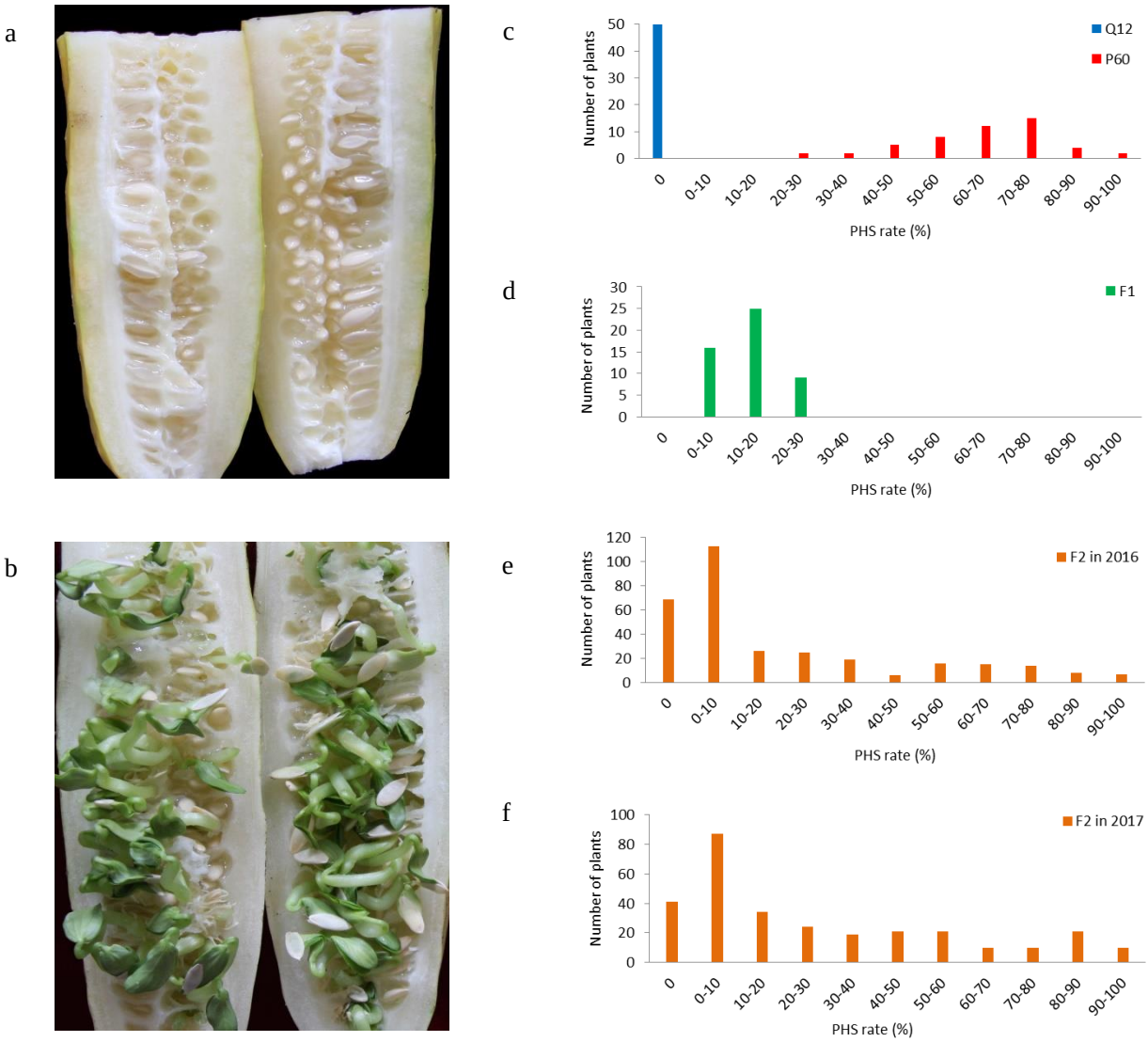


**Fig. 1 Pre-harvest sprouting (PHS) and its frequency distribution in the parental lines, $F_1$ and $F_2$ populations**

a: Phenotype of Q12, resistant to PHS; b: Phenotype of P60, susceptible to PHS; c: Frequency distribution of PHS in the parental lines, Q12 and P60; d: Frequency distribution of PHS in the $F_1$ generation grown in 2016; e: Frequency distribution of $F_2$ population grown in 2016; f: Frequency distribution of $F_2$ population grown in 2017.

Table 1 Frequency of pre-harvest sprouting rate in different populations

| opulation | Rate of PHS/% | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 0~10 | 10~20 | 20~30 | 30~40 | 40~50 | 50~60 | 60~70 | 70~80 | 80~90 | 90~100 | Average |
| $P_1$ | 50 | | | | | | | | | | | 0 |
| $P_2$ | | | | 2 | 2 | 5 | 8 | 12 | 15 | 4 | 2 | 64.97 |
| $F_1$ | | 16 | 25 | 9 | | | | | | | | 13.88 |
| $F_2$ in 2016 | 71 | 112 | 26 | 25 | 19 | 6 | 16 | 15 | 14 | 8 | 6 | 20.77 |
| $F_2$ in 2017 | 41 | 87 | 34 | 24 | 19 | 21 | 21 | 10 | 10 | 21 | 10 | 28.07 |

## Pool Construction and QTL-seq

Based on the phenotyping data of $F_2$ individuals (Additional file 1: Table S1), 30 extremely resistant and 30 extremely susceptible individuals were selected from the $F_2$ population grown in 2017 for the construction of the R- and S-pool, respectively. The PHS rate of each extreme $F_2$ individual in the R-pool was 0%, and the PHS rate of extreme individuals in the S-bulk ranged from 80% to 100%. Each DNA pool, along with the R-parent (Q12) and S-parent (P60), were subjected to whole-genome resequencing (WGRS) using the Illumina HiSeq4000 platform, and 36.83 Gb raw data was generated. The clean data were mapped to the cucumber reference genome [32], and 36.55 Gb remained after trimming and adapter removal. A total of 10.92 Gb clean data (18.59X coverage) for Q12, 10.37 Gb (22.30X coverage) for P60, 8.43 Gb for the R-pool (28.06X coverage) and 6.83 Gb (30.54X coverage) S-pool was generated. Detailed information is listed in Table 2.

Using GATK software, a total of 62,504 SNPs and 18,646 InDel variants were detected between the two parents. The Δ(SNP/InDel index) of the polymorphic loci between the R-pool and S-pool was calculated with a statistical confidence of $P< 0.05$ based on the SNP/InDel index to locate the QTL. The sliding window approach was used, and the average values of Δ(SNP/InDel index) in each window were plotted in a graph (Fig. 2). Only two regions that were significantly different from 0 regions at the 95% confidence value were identified; one region spanned 7.3 Mb (region 13,778,717-21,079,500) on chromosome 4, and the other spanned 0.15 Mb (region 6,688,410-6,840,412) on chromosome 5. We named these two predicted regions that were putatively associated with PHS in cucumber *qPHS4.1* and *qPHS5*, respectively. These two regions contained 443 SNPs and 124 InDels (Table 3), of which 272 SNPs and 82 InDels were found to be intergenic, 70 SNPs and 19 InDels intronic, 4 SNPs synonymous, 6 SNPs nonsynonymous, 39 SNPs and 11 InDels in 3′UTRs, and 47 SNPs and 10 InDels in 5′ UTRs (Table 3). Based on the gene annotation, genes containing stop loss, stop gain or nonsynonymous mutations were preferentially selected as candidate genes (Additional file 2: Table S2) from the associated regions.

Table 2 Resequencing summary of the parental lines, R-pool and S-pool

| Sample | Clean bases (Gb) | Total reads | Mapped reads | Rate of mapped reads (%) | Sequencing depth (X) | Genome coverage |
|---|---|---|---|---|---|---|
| Q12 | 10.92 | 45 507 344 | 38 958 916 | 85.61 | 18.59 | 98.75 |
| P60 | 10.37 | 56 203 612 | 47 443 606 | 84.41 | 22.30 | 98.82 |
| R-pool | 8.43 | 69 130 804 | 61 513 704 | 88.98 | 28.06 | 98.99 |
| S-pool | 6.83 | 72 805 348 | 63 923 453 | 87.80 | 30.54 | 98.99 |

Table 3 Categorization of Detected Variations

| Category | | SNPs | InDels |
|---|---|---|---|
| Exonic | Synonymous | 4 | 0 |
| | Non-Synonymous | 6 | 0 |
| | Non-Frameshift Insertion | - | 1 |
| Intronic | | 70 | 19 |
| 3′UTR | | 39 | 11 |
| 5′UTR | | 47 | 10 |
| 3′UTR/5′UTR | | 5 | 1 |
| Intergenic | | 272 | 82 |
| TS | | 277 | - |
| TV | | 166 | - |
| Insertion | | - | 62 |
| Deletion | | - | 62 |
| Total | | 443 | 124 |

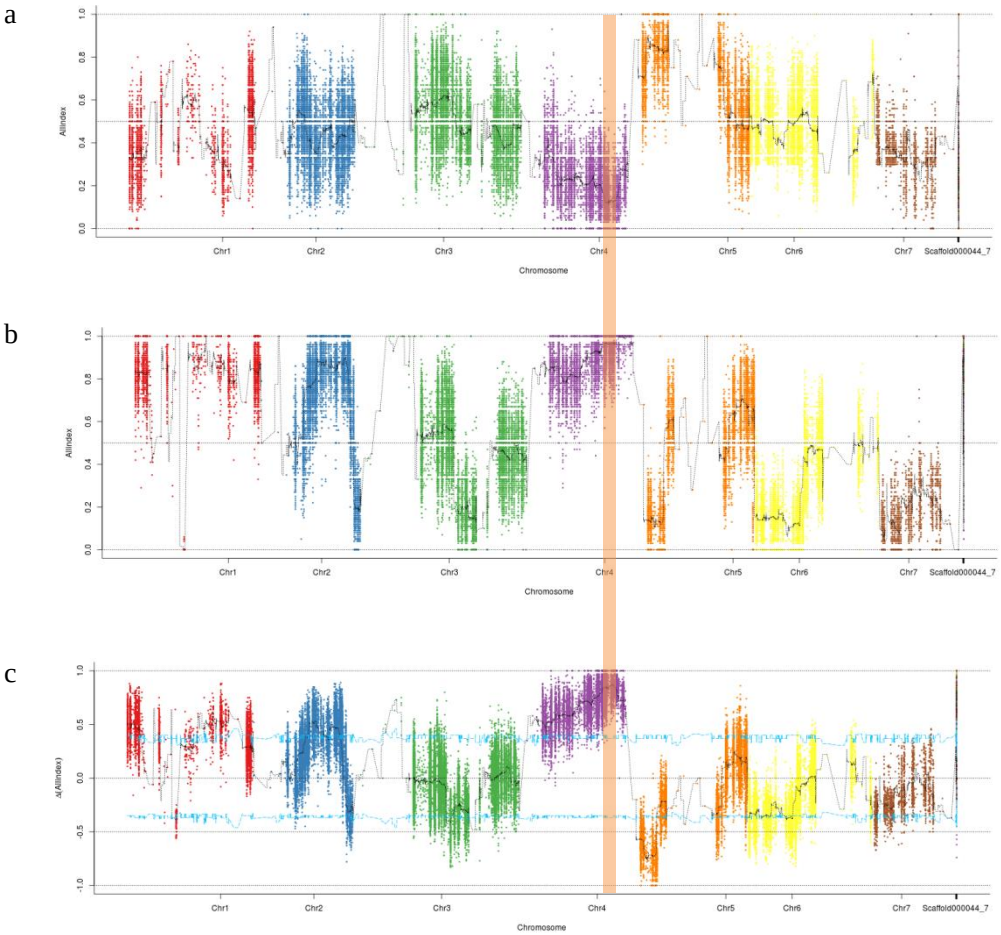


**Fig. 2 SNP/InDel-index Manhattan graphs of R-pool, S-pool and Δ (SNP/InDel-index) from QTL-seq approach for mapping the genomic regions controlling pre-harvest sprouting in cucumber.**

a: SNP/InDel-index plot of R-pool; b: SNP/InDel-index plot of S-pool; c: the Δ (SNP/InDel-index) plot of all chromosomes with the statistical confidence interval under the null hypothesis of no QTLs (blue line $P < 0.05$). The significant genomic region on Chromosome 4 is highlighted in shaded color (13.78–21.08 Mb).

## Validation and Narrowing Down the Associated Region

To verify the results detected by QTL-seq and narrow down the candidate intervals, a traditional QTL mapping method was used. We genotyped all $F_2$ individuals grown in 2016 and 2017 for 62 SNP and/or InDel markers selected from the *qPHS4.1* and *qPHS5.1* intervals, respectively. Finally, twenty-nine markers in *qPHS4.1* were accurately genotyped and applied to construct the local genetic linkage maps by JoinMap 4.0 software. Two InDel markers on Chromosome 5 were unmapped. After calculation by MapQTL version 6 software, two loci with LOD scores over the threshold, SNP-16 and SNP-23, were found by using the 2016 $F_2$ population. As shown in Table 4, the peak LOD scores of SNP-16 and SNP-23 were 15.07 and 15.28, respectively. This interval explained 19.6-19.8% of the phenotypic variation in PHS. In the 2017 $F_2$ population, two peak SNP loci, SNP-17 (LOD=13.89) and SNP-24 (LOD=16.06), were detected (Table 4, Additional file 3: Table S3). The interval explained 19.3-22.0% of the phenotypic variation in PHS. By taking the overlapping regions into account, these results reduced the candidate genomic interval associated with *qPHS4.1* from 7.3 Mb to the 0.53 Mb flanked by the markers SNP-17 to SNP-23 on chromosome 4 in cucumber (Fig. 3).

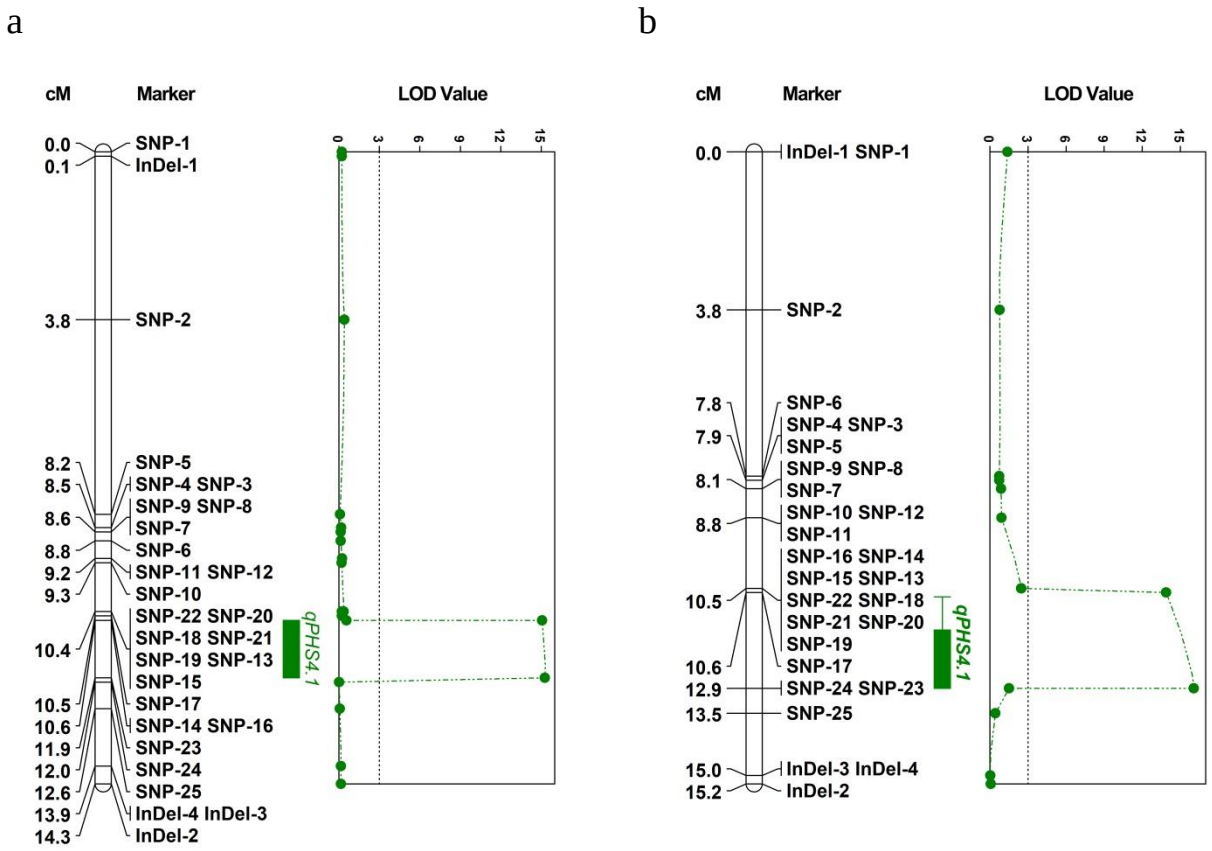


**Fig. 3 Fine mapping of the major-effect QTL *qPHS4.1* in cucumber using $F_2$ populations grown in 2016 (a) and 2017 (b).**
SNP and InDel markers in candidate regions generated by QTL-seq were selected and genotyped in the 318 $F_2$ individuals grown in 2016 and 298 $F_2$ individuals grown in 2017. One major-effect QTL in the overlapping region was identified. The interval of *qPHS4.1* was narrowed down to 0.53 Mb on Chromosome 4.

Table 4 LOD Values, Additive Effects, and Variance Explained for the Significant Loci Associated with pre-harvest sprouting in Cucumber

| Year | The SNP markers | Physical position on Chromosome 4 (bp) | Interval (Mb) | LOD[a] | Additive effect[b] | Dominance | Variance explained(%)[c] |
|---|---|---|---|---|---|---|---|
| 2016 | SNP-16 | 19973741 | 0.53 | 15.07 | -0.136391 | -0.0594552 | 19.6 |
| | SNP-23 | 20505510 | | 15.28 | -0.141843 | -0.0345112 | 19.8 |
| 2017 | SNP-17 | 19973782 | 0.55 | 13.89 | -0.159806 | -0.0408643 | 19.3 |
| | SNP-24 | 20521004 | | 16.06 | -0.182258 | 0.0193450 | 22.0 |

[a]Peak LOD score of the QTL. [b]Additive or dominant effect of the SNPs. [c]Percentage of variance explained by the QTL peak.

## Gene Annotation and Expression Analysis of Candidate Genes

On the basis of the gene annotations, within the *qPHS4.1* region, *Csa4G622760*, *Csa4G622800* and *Csa4M628930.1* (Table 5), in which nonsynonymous or 3'UTR mutations occurred, were selected as candidate genes for further analysis. The relative expression levels of the candidate genes in seed cavity flesh tissues were examined by Real-time Quantitative PCR (qRT-PCR), as shown in Fig. 4. The expression level of *Csa4G622760*, which is predicted to encode a chalcone isomerase-like protein, was 1.9-fold higher in Q12 than in P60 at the 34 DAP stage. However, its expression level was 5.4-fold lower in Q12 than in P60 at 40 DAP. This indicated that the expression level of the *Csa4G622760* gene significantly decreased, by approximately 20-fold, from 34 DAP to 40 DAP in Q12 but was only 2-fold down-regulated in

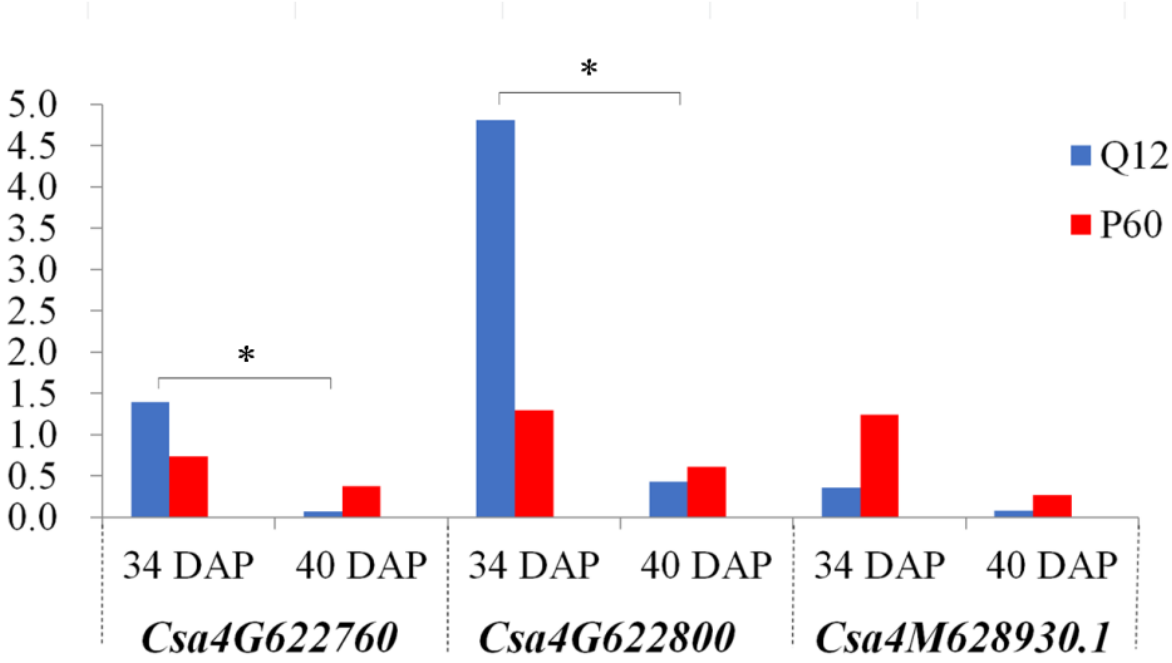


**Fig. 4 The relative quantitative expression analysis of the predicted genes in cucumber cavity flesh tissue of Q12 and P60**

The blue bars represent Q12, the red bars represent P60. 34 DAP indicates the levels in the cucumber cavity flesh tissue sampled from cucumber fruits at 34 days after pollination (DAP), at which point the seeds had not germinated in the cucumber cavities. 40 DAP indicates the levels in the cucumber fruits at 40 days after pollination, at which point the seeds had germinated in those cucumbers that were susceptible to pre-harvest sprouting. * $P < 0.05$.

P60. The *Csa4G622800* gene is annotated as a peptide methionine sulfoxide reductase msrB. Its expression level was 3.7-fold higher in Q12 than in P60 at the 34 DAP stage. At the 40 DAP stage, the expression level was down-regulated 11.2-fold in Q12 and 2.1-fold in P60. Gene expression of *Csa4G622800* also decreased significantly. *Csa4M628930.1* is a putative ERI1 exoribonuclease 3 protein. At 34 DAP, the expression in P60 was 3.4-fold higher than that in Q12. From 34 DAP to 40 DAP, gene expression decreased approximately 4.6-fold in both parental lines.

Table 5 Candidate Genes Underlying *qPHS4.1* Control of Preharvest Sprouting in Cucumber

| Gene ID | SNP location | SNP locus | Physical position (bp) | Mutation | | Functional prediction |
|---|---|---|---|---|---|---|
| | | | | Q12 | P60 | |
| *Csa4G622760* | upstream | SNP-14 | 19973692 | G | T | Chalcone isomerase-like protein |
| | upstream | SNP-15 | 19973724 | C | A | |
| | upstream | SNP-16 | 19973741 | T | C | |
| | upstream | SNP-17 | 19973782 | A | T | |
| *Csa4G622800* | upstream | SNP-18 | 19995077 | A | G | Peptide methionine sulfoxide reductase msrB |
| | upstream | SNP-19 | 19995107 | G | A | |
| | upstream | SNP-20 | 19995109 | C | G | |
| | upstream | SNP-21 | 19995123 | A | C | |
| | upstream | SNP-22 | 19995137 | C | A | |
| *Csa4M628930.1* | nonsynonymous | SNP-23 | 20505510 | T | C | ERI1 exoribonuclease 3 |

Taken together, these data show that the expression levels of the three genes were both down-regulated in Q12 and P60 with increasing ripeness of cucumber fruits. The *Csa4M628930.1* gene showed a different expression pattern from that of *Csa4G622760* and *Csa4G622800*. *Csa4G622760* and *Csa4G622800* gene expression significantly decreased in Q12 but decreased slightly in P60 from the 34 DAP stage (PHS not occurred) to 40 DAP stage (PHS occurred). These results suggested that *Csa4G622760* and *Csa4G622800* gene expression levels were higher in resistant cucumbers than in susceptible cucumbers before PHS occurred. Subsequently, accompanying the occurrence of PHS, its gene expression levels decreased significantly in resistant cucumbers compared to susceptible cucumbers. Therefore, we hypothesized that *Csa4G622760* and *Csa4G622800* are possible candidate genes involved in PHS in cucumber, but further functional analysis of these genes needs to be conducted.

## DISCUSSION

In cucumber and other seed-bearing crops, pre-harvest sprouting (PHS) is a critical problem that causes devastating losses to seed yields and quality and widely limits seed dispersal. To promote the process of cucumber PHS resistance breeding, it is greatly important to identify key loci controlling PHS resistance and develop molecular markers for marker-assisted selection (MAS). In cereal crops, including wheat, rice, maize and barley, PHS is a very popular research topic, and the investigation of genetic mapping and molecular mechanisms underlying PHS is extensive and intensive. However, unfortunately, few published studies have focused on the PHS trait in cucumber [33]. In this study, we identified QTLs associated with PHS by a QTL-seq approach in the $F_2$ population derived from the two parents Q12 and P60, which showed opposite extremes of PHS phenotypes. Q12 is a typical resistant line in which PHS never occurs in favorable environments, while PHS occurs in the P60 line (Fig. 1). The frequency distribution of

PHS in P60 was normal. Subsequently, in the $F_2$ population, the frequency distribution was skewed normal rather than normal (Fig. 1), suggesting that PHS in cucumber was controlled by a major-effect QTL. This is consistent with our previous research on the inheritance of PHS.

The application of high-throughput next-generation sequencing technology promotes the development of rapid molecular marker discovery and physical map construction. QTL-seq is a new method that combines next-generation sequencing with BSA for the rapid detection of QTLs and links molecular markers associated with traits of interest. It was first developed by Takagi et al and applied in rice [24]. Since that time, QTL-seq has been successfully used in many species, such as chickpea, oilseed rape, maize, tomato, and cucumber [25-31]. However, the candidate regions generated from QTL-Seq are often too rough or too broad, and additional QTL analysis performed by traditional methods is necessary to refine gene locations and narrow chromosomal intervals. In the present study, a QTL-seq approach was performed in the $F_2$ population grown in 2016. Two QTLs associated with PHS, *qPHS4.1* and *qPHS5.1*, were initially identified, which spanned 7.3 Mb on chromosome 4 (13.78 Mb-21.08 Mb) and 0.15 Mb on chromosome 5 (6.69 Mb to 6.84 Mb), respectively. Traditional QTL mapping methods were also conducted. The phenotype identification and QTL mapping using the $F_2$ population grown in 2016 was consistent with the findings from the 2017 season, which indicated that the experimental results were reliable and accurate. Subsequently, the *qPHS4.1* regions generated from the two seasons were overlapped. Therefore, *qPHS4.1* was refined and narrowed down to 0.53 Mb on Chromosome 4. However, *qPHS5.1* remained unmapped. Therefore, this result demonstrated that *qPHS4.1* was the key major-effect QTL controlling PHS in cucumber.

In total, 39 genes underlie the refined *qPHS4.1* interval. Based on gene annotation by ANNOVER software and gene expression analysis by qRT-PCR, two genes, *Csa4G622760* and *Csa4G622800*, containing upstream polymorphic SNPs, were considered candidate PHS regulating genes in cucumber (Table 5). The *Csa4G622760* gene is predicted to encode a chalcone isomerase-like protein that catalyzes the biosynthesis of flavonoids and secondary metabolism in plants [34]. Flavonoids are important secondary metabolites found in various plant tissues, such as leaves, flowers, fruits and seeds. In Arabidopsis, overexpression of the *chalcone isomerase-like* gene increased the accumulation of proanthocyanidin and flavonol, which are flavonoids, while loss of function of the *chalcone isomerase-like* gene led to a strong reduction in proanthocyanidin and flavonol levels and influenced the seed phenotype [35]. However, the correlation between flavonoids and PHS in cucumber is unclear. The *Csa4G622800* gene is predicted to encode the peptide methionine sulfoxide reductase msrB. In plant seeds, methionine sulfoxide reductase plays a decisive role in the establishment and preservation of seed longevity. Higher activity of this enzyme leads to better preservation of the seeds and higher germination capacity [36]. However, further experiments need to be performed to test the functionality of the two candidate genes in the genetic mechanisms of cucumber PHS.

In some cereal crops, e.g., wheat, rice, barley, etc., extensive QTLs and numerous genes associated with seed dormancy and PHS have been reported. However, only one key major-effect QTL in cucumber was identified in this study. In contrast to cereal crops, cucumber seeds are surrounded by flesh tissues in seed cavities, in which the water content is higher than 95%. Therefore, PHS in cucumber cannot be influenced by the humidity of the environment. Cucumber PHS is a very specific and interesting trait, and the molecular mechanisms underlying PHS need further study.

## Conclusions

In this study, two QTLs associated with PHS in cucumber were detected using QTL-seq approach. The key major-effect QTL *qPHS4.1* was refined to 0.53 Mb on chromosome 4. Based on the gene annotation and qRT-PCR analysis, two genes located in *qPHS4.1* were proposed to be the candidate genes associated with cucumber PHS. To our knowledge, this is the first report on the identification of QTLs associated with PHS trait in cucumber. This study provides novel insights into the genetic mechanism controlling PHS in cucumber and highlights the potential for PHS resistance MAS breeding.

## MATERIALS AND METHODS

### Plant materials and Phenotypic Evaluation

The high-generation inbred cucumber lines Q12 (North China fresh market cucumber, PHS resistant, $P_1$) and P60 (North China fresh market cucumber, PHS susceptible, $P_2$) were crossed to obtain $F_1$. $F_1$ plants were self-crossed to generate an $F_2$ segregating population. The $F_2$ population was evaluated for PHS in the experimental farm of the Tianjin Kernel Cucumber Research Institute (Tianjin, China) in the 2016 (328 plants) and 2017 (299 plants) seasons. The seeds of the populations were sown directly into soil in a greenhouse on April 15 each year. For plant management, two female flowers were self-pollinated, and all the other female flowers and lateral branches were removed from each plant. The pollination date was recorded on labels hung on the handle of the fruits. All the plants were grown in greenhouse conditions under whole-day light exposure. The day/night temperature in the greenhouse was controlled at 28-35℃/15-26℃. The seeds in the cucumber fruits were harvested at 45 days after pollination (DAP), and the numbers of germinated seeds and total seeds were counted immediately. The PHS rate (%) was calculated as (germinated seeds/total seeds in fruit) × 100%. The average PHS rates of two cucumber fruits grown on the same plant were used for QTL analysis.

### Pool Construction and Whole-Genome Re-Sequencing

The genomic DNA of Q12, P60 and $F_2$ individuals was extracted from seedling leaves using a Quick Prep Plant Genome DNA Kit (HUALIKEXI, Tianjin, China). Q12 and P60 genomic DNA were used to construct the $P_1$ pool and $P_2$ pool. Based on the phenotype data of $F_2$ individuals grown in the 2017 season (Additional file 1: Table S1), 30 extreme resistant plants and 30 extreme susceptible plants were selected to construct a resistant pool (R-pool) and susceptible pool (S-pool), respectively. Equal amounts of DNA from the selected individuals were mixed and subsequently processed to generate sequencing libraries using the TruSeq Nano DNA HT Sample preparation Kit (Illumina USA). These libraries were resequenced by the Illumina HiSeq4000 platform.

### QTL-seq

The raw sequencing data were filtered to obtain clean data through a series of quality control procedures. The clean reads obtained from four pools were aligned to the cucumber reference genome sequence [32] using the BWA (Burrows-Wheeler Aligner) tool [32,37]. Variant calling was performed for the samples by using the Unified Genotyper function in GATK software [38]. To determine the genomic regions associated with PHS, we calculated the SNP/InDel-index and Δ(SNP/InDel-index) to locate the QTLs. The SNP/InDel-index refers to the proportion of reads carrying a SNP/InDel different from the reference reads of either parent. The Δ(SNP/InDel-index)

of each locus was determined based on the difference in the SNP/InDel-index between the R-pool and S-pool. To eliminate background interference, we filtered out all loci with an SNP/InDel-index of less than 0.3. Using the slicing window method with a 1 Mb window size and 1 kb increment, the average SNP/InDel-index of loci in a given genomic interval was calculated. The Δ(SNP/InDel-index) of the R-pool and S-pool and the corresponding SNP/InDel-indexes in the slicing window were plotted in a graph to generate SNP/InDel-index plots. We calculated statistical confidence intervals of Δ(SNP/InDel-index) for all SNP and InDel loci with a given read depth and obtained 95% and 99% confidence intervals. By examining the Δ(SNP/InDel-index), the plot peak regions above the confidence value were defined as predicted regions for association with PHS. ANNOVAR software was used to annotate the candidate genes in the regions.

## Genotyping, Regional Linkage Mapping and QTL analysis

To verify the candidate SNP and InDel markers and narrow down the regions identified by QTL-seq, significant SNPs and InDels in the candidate regions were first selected and validated in the two parents and their $F_1$ plants. Then, polymorphic SNP and InDel markers were used to genotype the extended $F_2$ individuals sown in the 2016 season and 2017 season. This validation was performed on the high-throughput HI-SNP genotyping platform. The specific multiplex PCR primers of the markers were designed by Primer 3 online software (http://frodo.wi.mit.edu/, Version 0.4.0) based on the cucumber reference genome [32] (listed in Additional file 4: Table S4). The final SNP/InDel information was obtained by bioinformatics analysis of the sequence data.

Based on the genotypes of significant SNPs and InDels in candidate regions of $F_2$ individuals sown in 2016 and 2017, regional linkage maps were constructed using JoinMap 4.0 software [39] with the maximum likelihood mapping algorithm and Kosambi mapping function [40], respectively. According to the phenotyping datasets of the $F_2$ individuals, QTL analysis was performed by the software MapQTL version 6 [41]. The "MQM mapping" algorithm with an LOD threshold score of > 3.0 was used to perform the calculation. The output logarithm of odds (LOD) scores were plotted along the genetic distances of the markers analyzed.

## Candidate Gene Annotation

According to the further narrowed region of the QTLs, effective SNPs or InDels associated with PHS were identified. Based on the Cucurbit Genomics Database (http://www.icugi.org/cgi-bin/ICuGI/index.cgi), the functions of candidate PHS-associated genes that contained non-synonymous or upstream/downstream variations were predicted.

## Gene Expression Analysis by qRT-PCR

We used qRT-PCR to investigate the relative expression levels of the candidate genes between the two parents. The cucumber cavity flesh tissues surrounding the seeds at 34 DAP (PHS not occurred) and 40 DAP (PHS occurred) were sampled, and RNA was extracted using TRNzol Universal Reagent following the manufacturer's protocol (TIANGEN, Beijing, China). The RNA quality was evaluated by agarose gel electrophoresis. cDNA was synthesized using a RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific Inc., MA, USA). Primers for candidate genes were designed by Primer 3 and synthesized by Sangon Biotech (Shanghai) Co., Ltd. (Sangon, Shanghai, China). Details of the primer sequences are presented in Table 6. qRT-PCR was conducted by using TB Green Premix Ex Taq II (Takara Bio Inc., Dalian, China). The qRT-PCR conditions were set as follows: 95℃ for 30 s; followed by 40 cycles of 95℃ for 5 s and then 60℃ for 30 s; and then denaturation at 95℃ for 15 s, 60℃ for 60 s, a temperature increase of 0.3℃ per 15 s, and finally 95℃ for 15 s. The *tubulin* gene (GenBank ID: AF044573.1) was used

as the reference gene for normalization of the relative expression of the candidate genes. The relative expression levels of the target genes were calculated using the $2^{-\Delta\Delta Ct}$ method [41]. The experiments were conducted with three biological replicates.

Table 6 qRT-PCR primers for candidate genes and reference gene (*Tubulin*)

| Gene ID | F | R | Size ( bp ) |
|---|---|---|---|
| *Csa4G622760* | TAACTCTGCCAGGCTGCTCAAC | GTCTCGGACAAGAACAATCTGTAAAG | 242 |
| *Csa4G622800* | GCATCAAAGAGGCTGGCACC | TATCCCTGCCGTTTTGGTGTTC | 152 |
| *Csa4M628930.1* | GGAGTTGACCGTGTCTGGC | TGATGTGGGACCTGAGTTT | 173 |
| *Tubulin* | GCAAGGAAGATGCTGCCAATA | TCCATAGTCAACAGACAAACGCTC | 203 |

## References


1. Ogbonnaya FC, Imtiaz M, Ye G, Hearnden PR, Hernandez E, Eastwood RF, van Ginkel M, Shorter SC, Winchester JM. Genetic and QTL analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. Theor Appl Genet. 2008;116(7):891-902.
2. Vetch JM, Stougaard RN, Martin JM, Giroux MJ. Review: Revealing the genetic mechanisms of pre-harvest sprouting in hexaploid wheat (*Triticum aestivum* L.). Plant Sci. 2019;281:180-185.
3. Finch-Savage WE, Leubner-Metzger G. Seed dormancy and the control of germination. New Phytol. 2006;171(3):501-23.
4. Bentsink L, Koornneef M. Seed dormancy and germination. Arabidopsis Book. 2008;6:e0119.
5. Rodr ́guez M, Barrero J, Corbineau F, Gubler F, Benech-Arnold R. Dormancy in cereals (not too much, not so little): About the mechanisms behind this trait. Seed Science Research. 2015;25(2):99-119.
6. Takahashi N. Effect of environmental factors during seed formation on pre-harvest sprouting. Cereal Res Commun. 1980;8(1):175-183.
7. Nakamura S, Pourkheirandish M, Morishige H, Sameri M, Sato K, Komatsuda T. Quantitative Trait Loci and Maternal Effects Affecting the Strong Grain Dormancy of Wild Barley (*Hordeum vulgare* ssp. *spontaneum*). Front Plant Sci. 2017;8:1840.
8. Magwa RA, Zhao H, Xing Y. Genome-wide association mapping revealed a diverse genetic basis of seed dormancy across subpopulations in rice (*Oryza sativa* L.). BMC Genet. 2016;17:28.
9. Mizuno Y, Yamanouchi U, Hoshino T, Nonoue Y, Nagata K, Fukuoka S, Ando T, Yano M, Sugimoto K. Genetic dissection of pre-harvest sprouting resistance in an upland rice cultivar. Breed Sci. 2018;68(2):200-209.
10. Shu K, Meng YJ, Shuai HW, Liu WG, Du JB, Liu J, Yang WY. Dormancy and germination: How does the crop seed decide? Plant Biol (Stuttg). 2015;17(6):1104-12.
11. Somyong S, Ishikawa G, Munkvold JD, Tanaka J, Benscher D, Cho YG, Sorrells ME. Fine mapping of a preharvest sprouting QTL interval on chromosome 2B in white wheat. Theor Appl Genet. 2014;127(8):1843-55.

12. Shao M, Bai G, Rife TW, Poland J, Lin M, Liu S, Chen H, Kumssa T, Fritz A, Trick H, Li Y, Zhang G. QTL mapping of pre-harvest sprouting resistance in a white wheat cultivar Danby. Theor Appl Genet. 2018;131(8):1683-1697.
13. Cao L, Hayashi K, Tokui M, Mori M, Miura H, Onishi K. Detection of QTLs for traits associated with pre-harvest sprouting resistance in bread wheat (*Triticum aestivum* L.). Breed Sci. 2016;66(2):260-70.
14. Hickey LT, Lawson W, Arief VN, Fox G, Franckowiak J, Dieters M J. Grain dormancy QTL identified in a doubled haploid barley population derived from two non-dormant parents. Euphytica. 2012;188:113-122.
15. Vanhala T K , Stam P . Quantitative Trait Loci for Seed Dormancy in Wild Barley (*Hordeum spontaneum* C. Koch). Genet Resour Crop Evol. 2006;53(5):1013-1019.
16. Nakamura S, Pourkheirandish M, Morishige H, Kubo Y, Nakamura M, Ichimura K, Seo S, Kanamori H, Wu J, Ando T, Hensel G, Sameri M, Stein N, Sato K, Matsumoto T, Yano M, Komatsuda T. Mitogen-Activated Protein Kinase Kinase 3 Regulates Seed Dormancy in Barley. Curr Biol. 2016;26(6):775-81.
17. Hori K, Sato K, Takeda K. Detection of seed dormancy QTL in multiple mapping populations derived from crosses involving novel barley germplasm. Theor Appl Genet. 2007 Oct;115(6):869-76.
18. Sato K, Yamane M, Yamaji N, Kanamori H, Tagiri A, Schwerdt JG, Fincher GB, Matsumoto T, Takeda K, Komatsuda T. Alanine aminotransferase controls seed dormancy in barley. Nat Commun. 2016;7:11625.
19. Finkelstein R, Reeves W, Ariizumi T, Steber C. Molecular aspects of seed dormancy. Annu Rev Plant Biol. 2008;59:387-415.
20. Singh M, Lewis PE, Hardeman K, Bai L, Rose JK, Mazourek M, Chomet P, Brutnell TP. *Activator* mutagenesis of the pink *scutellum1*/*viviparous7* locus of maize. Plant Cell. 2003;15(4):874-84.
21. Yao M, Chen W, Kong J, Zhang X, Shi N, Zhong S, Ma P, Gallusci P, Jackson S, Liu Y, Hong Y. *METHYLTRANSFERASE1* and Ripening Modulate Vivipary during Tomato Fruit Development. Plant Physiol. 2020;183(4):1883-1897.
22. Salvi S, Tuberosa R. To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci. 2005;10(6):297-304.
23. Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci U S A. 1991;88(21):9828-32.
24. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S, Innan H, Cano LM, Kamoun S, Terauchi R. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 2013 Apr;74(1):174-83.
25. Singh VK, Khan AW, Jaganathan D, Thudi M, Roorkiwal M, Takagi H, Garg V, Kumar V, Chitikineni A, Gaur PM, Sutton T, Terauchi R, Varshney RK. QTL-seq for rapid identification of candidate genes for 100-seed weight and root/total plant dry weight ratio under rainfed conditions in chickpea. Plant Biotechnol J. 2016;14(11):2110-2119.
26. Wang H, Cheng H, Wang W, Liu J, Hao M, Mei D, Zhou R, Fu L, Hu Q. Identification of *BnaYUCCA6* as a candidate gene for branch angle in *Brassica napus* by QTL-seq. Sci Rep.

2016;6:38493.
27. Wei QZ, Fu WY, Wang YZ, Qin XD, Wang J, Li J, Lou QF, Chen JF. Rapid identification of fruit length loci in cucumber (*Cucumis sativus* L.) using next-generation sequencing (NGS)-based QTL analysis. Sci Rep. 2016;6:27496.
28. Chen Q, Song J, Du WP, Xu LY, Jiang Y, Zhang J, Xiang XL, Yu GR. Identification, Mapping, and Molecular Marker Development for Rgsr8.1: A New Quantitative Trait Locus Conferring Resistance to Gibberella Stalk Rot in Maize (*Zea mays* L.). Front Plant Sci. 2017;8:1355.
29. Wen J, Jiang F, Weng Y, Sun M, Shi X, Zhou Y, Yu L, Wu Z. Identification of heat-tolerance QTLs and high-temperature stress-responsive genes through conventional QTL mapping, QTL-seq and RNA-seq in tomato. BMC Plant Biol. 2019;19(1):398.
30. Arikit S, Wanchana S, Khanthong S, Saensuk C, Thianthavon T, Vanavichit A, Toojinda T. QTL-seq identifies cooked grain elongation QTLs near soluble starch synthase and starch branching enzymes in rice (*Oryza sativa* L.). Sci Rep. 2019;9(1):8328.
31. Lei L, Zheng H, Bi Y, Yang L, Liu H, Wang J, Sun J, Zhao H, Li X, Li J, Lai Y, Zou D. Identification of a Major QTL and Candidate Gene Analysis of Salt Tolerance at the Bud Burst Stage in Rice (*Oryza sativa* L.) Using QTL-Seq and RNA-Seq. Rice (N Y). 2020;13(1):55.
32. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan, Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S. The genome of the cucumber, *Cucumis sativus* L. Nat Genet. 2009;41(12):1275-81.
33. Mu Y, Liu Y, Bai L, Li S, He C, Yan Y, Yu X, Li Y. Cucumber CsBPCs Regulate the Expression of CsABI3 during Seed Germination. Front Plant Sci. 2017;8:459.
34. Lin LM, Guo HY, Song X, Zhang DD, Long YH, Xing ZB. Adaptive Evolution of Chalcone Isomerase Superfamily in Fagaceae. Biochem Genet. 2020;11:1-15.
35. Jiang W, Yin Q, Wu R, Zheng G, Liu J, Dixon RA, Pang Y. Role of a chalcone isomerase-like protein in flavonoid biosynthesis in Arabidopsis thaliana. J Exp Bot. 2015;66(22):7165-79.
36. Ch âtelain E, Satour P, Laugier E, Ly Vu B, Payet N, Rey P, Montrichard F. Evidence for participation of the methionine sulfoxide reductase repair system in plant seed longevity. Proc Natl Acad Sci U S A. 2013;110(9):3633-8.
37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.
38. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.
39. Van Ooijen JW. JointMap 4. *Software for the calculation of genetic linkage maps in experimental populations*. Wageningen: Kyazma BV; 2006.

40. Kosambi DD. The Estimation of Map Distances from Recombination Values. In: Ramaswamy R, editor. DD Kosambi: Selected Works in Mathematics and Statistics. New Delhi: Springer India; 2016;pp.125-130.
41. Van Ooijen JW. MapQTL 6.0, software for the mapping of quantitative trait loci in experimental populations of dihaploid species. The Netherlands: Kyazma BV, Wageningen; 2009.
42. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.

## Declarations:

**Ethics approval and consent to participate:** Not applicable.

**Consent for publication:** Not applicable.

**Competing interests:** The authors declare that they have no competing interests.

**Availability of data and materials:** All data generated or analyzed during this study are included in this published article and its supplementary information files.

**Funding:** This research was financially supported by Natural Science Foundation of Tianjin City (17JCYBJC29400).

**Authors' contributions:** CM performed the linkage analysis, genotyping work, mapping the QTLs and wrote the manuscript. LS aided research design, assisted data analysis and revised the manuscript. DQ ang WH did the PCR and electrophoresis detection. YR did the target trait testing. All authors read and approved the final manuscript.

**Acknowledgements:**
We are grateful to Yu Ning (Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences) and Qingzhen Wei (Institute of Vegetable Research, Zhejiang Academy of Agricultural Sciences) for their excellent suggestions and revision on our manuscript.