

SCIENTIFIC REPORTS



OPEN

Identification of Quantitative Trait Loci and Candidate Genes for Maize Starch Granule Size through Association Mapping

Na Liu^{1,2}, Zhanhui Zhang¹, Yadong Xue¹, Shujun Meng¹, Yubi Huang³, Weihua Li¹, Jihong Huang² & Jihua Tang^{1,4}

Starch is an important nutrient component of maize kernels, and starch granule size largely determines kernel waxiness, viscosity, and other physicochemical and processing properties. To explore the genetic basis of maize starch granule size, 266 tropical, subtropical, and temperate inbred lines were subjected to genome-wide association analyses with an array of 56,110 random single nucleotide polymorphisms (SNPs). In the present panel, the kernel starch granule size ranged from 7–15.8 μm long and 6.8–14.3 μm wide. Fourteen significant SNPs were identified as being associated with the length of starch granules and 9 with their width. One linkage disequilibrium block flanking both sides of a significant SNP was defined as a quantitative trait locus (QTL) interval, and seven QTLs were mapped for both granule length and width. A total of 79 and 88 candidate genes associated with starch length and width, respectively, were identified as being distributed on QTL genomic regions. Among these candidate genes, six with high scores were predicted to be associated with maize starch granule size. A candidate gene association analysis identified significant SNPs within genes *GRMZM2G419655* and *GRMZM2G511067*, which could be used as functional markers in screening starch granule size for different commercial uses.

Starch is a widely and naturally occurring biopolymer. It is composed of D-glucose units that form two types of polymers: amylose and amylopectin. Many cereals and tuber crops produce different types of starch. For instance, maize is an important crop worldwide¹, and starch is the main component of maize kernels, comprising ~70% of the total weight². Maize starch is not only an important food source but is also one of the most industrially used resources³. Its uses include the preparation of soups, sauces, baked goods, dairy, confectionery, snacks, pasta, coatings, and meat-containing products^{4,5}, as well as adhesives, paper, and textiles⁶.

Starch formation in cereal grains involves the synthesis of ADP-glucose (Glc) by ADP-Glc pyrophosphorylase and the incorporation of ADP-Glc into starch by ADP-Glc starch synthase⁷. Developing seeds synthesise storage compounds from imported sucrose during their maturation phase⁸. The assimilation of sucrose, which is imported through the phloem, by the endosperm may involve sucrose synthase to form ADP-Glc or UDP-Glc and fructose or, alternatively, involve invertase to form free hexose. Many studies have focused on the physicochemical properties of maize starch and their influencing factors^{9–12}. Starch type¹³ and amylose content¹⁴ have important effects on starch properties. Other influencing factors include granular structure (shape, size, and porosity), molecular structure (organisation of growth rings and degree of crystallinity), and the presence of non-starch materials¹⁵.

The size of the starch granules, which depends on the plant species, is an important factor affecting starch characteristics¹⁶ and ultimately determines the industrial application¹⁷. Small starch granules can be used to replace fat in food applications because of their fat mimetic properties¹⁸. In food production, granule size affects

¹Key Laboratory of Wheat and Maize Crops Science, Collaborative Innovation Center of Henan Grain Crops, College of Agronomy, Henan Agricultural University, Zhengzhou, 450002, China. ²College of Biological engineering, Henan University of Technology, Zhengzhou, 450001, China. ³College of Agronomy, Sichuan Agricultural University, Wenjiang, 611130, China. ⁴Hubei Collaborative Innovation Center for Grain Industry, Yangtze University, Jingzhou, 434025, China. Correspondence and requests for materials should be addressed to J.H. (email: huangjih1216@126.com) or J.T. (email: tangjihua1@163.com)

the pasting properties of starch, with smaller granules showing lower peaks, troughs, and final viscosities than larger granules¹⁹. The starch granule size may also influence gelatinisation temperature²⁰, viscosity¹⁹, and enzymatic susceptibility^{21–23}. Additionally, it determines the grain milling yield in hard wheat²⁴. In maize, the size of the starch granules varies according to chemical composition^{25,26}.

Quantitative trait loci (QTLs) of starch granule sizes in *Triticeae* crops have been identified, including a major QTL related to the A:B ratio of wheat starch granules on chromosome 4S²⁷ and a QTL on barley chromosome 2²⁸. Recently, genome-wide association studies (GWASs) have been proven to be useful tools for the identification of candidate loci associated with traits in animal and plant species²⁹. For example, an analysis of maize oil biosynthesis identified 74 loci significantly associated with kernel oil concentration and fatty acid composition in a GWAS using 1 million single nucleotide polymorphisms (SNPs) characterised in 368 inbred maize lines³⁰. Furthermore, a GWAS and QTL mapping were found to be complementary, overcoming each other's limitations, in *Arabidopsis*³¹.

Compared with starches having a bimodal size distribution^{17,32–34}, few studies have investigated the unimodal starches, particularly that of maize³⁵. Although the sizes of maize starch granules are highly linked to the end-use quality of the products, many studies on maize starch have focused on its processing and nutritional properties³⁵, with little attention paid to the study of granule size³⁴. Here, we used a set of associated populations to identify significant SNP markers for starch granule size with the aim of predicting associated candidate genes.

Results

Phenotypic analysis of maize starch granule size. A total of 266 maize lines were used for association mapping. Although the starch granules of these inbred maize lines varied largely in size, more than 75% of the granules were 10–13.5 µm long and 9.7–11.8 µm wide (Table S1). The inbred line CIMBL30 had the smallest granule size (7 µm long × 6.8 µm wide; Fig. 1a), while the inbred line CML470 had the largest granule size (15.8 µm long × 14.3 µm wide; Fig. 1b). The starch granules of most inbred lines had a smooth surface (Fig. 1a–c), although some were rough or porous/cracked (Fig. 1d). The shapes varied, including rounded, spherical (Fig. 1e), or irregular (Fig. 1f). Thus, the sizes and shapes of the starch granules varied among different inbred lines (Fig. 2, Table 1), which may affect starch processing characteristics and seed unit weight.

Evaluation of starch pasting viscosity characteristics. The rapid visco analyser (Newport Scientific, Australia) profile revealed the paste viscosity characteristics of maize starch (Table S2). Seven parameters showed that large starch granules (such as in 'CIMBL12' and 'Zheng58') have smaller final viscosity levels than smaller starch granules.

Association analysis. The average data from different replicates of each inbred line were used for association analysis (Figs 3 and 4). For starch granule size, 14 significant SNPs were identified ($p < 2.25 \times 10^{-4}$; Fig. 4, Table 2), with 1, 2, and 11 SNPs distributed on chromosomes 6, 3, and 7, respectively. Seven QTLs, distributed over 79 candidate genes, were identified for starch granule length (Table 2).

Nine significant SNPs were identified to be associated with starch granule width, as well as seven QTLs and 88 candidate genes (Fig. 5, Table 2). Seven SNPs were identified for both starch granule length and width: one (*PZE-103182712*) on chromosome 3, one (*PZE-106103012*) on chromosome 6, and five (*PZE-107043911*, *PZE-107044857*, *PZE-107044898*, *PZE-107044943*, and *PZE-107045024*) on chromosome 7.

Gene ontology analysis. The QTLs analysis led to the identification of 108 candidate genes that were either associated with granule length or width. Among these, six genes with higher scores were located close to associated SNPs (Table 3). *GRMZM2G180104* was located between 75,849,776–75,850,502 bp on chromosome 7 and 3,503 bp upstream of significant SNP30343, while *GRMZM2G419655* and *GRMZM2G419660*, also on chromosome 7, were identified as being associated with starch granule size.

The Blast2Go program was used to predict the functions of these candidate genes (Tables 3 and S3, S4). The candidate gene on chromosome 6, *GRMZM2G167673*, was predicted to be involved in gibberellin synthesis and electron transport as a p450 cytochrome. *GRMZM2G419655* and *GRMZM2G419660* on chromosome 7 were predicted to encode phytoalkyltransferase precursors. *GRMZM2G511067* and *GRMZM2G663759* on chromosome 3 were predicted to encode a zinc finger CCCH domain-containing protein that binds metal ions and may repress the inhibitor of the phytoalkyltransferase receptor protein kinase.

The results of the GWAS analysis revealed that maize kernel starch granule size is a typical quantitative trait determined by multiple genes.

Association between candidate genes and maize starch granule size. An association analysis between three candidate genes and maize starch granule size revealed that the SNP at 352 bp of the *GRMZM2G419655* genomic sequence and the SNP at 58 bp of the *GRMZM2G511067* genomic sequence were significantly associated with maize starch granule length and width (Fig. 6). No significant SNP was identified after an association analysis between *GRMZM2G419660* and maize starch granule size. All of the sequences of the three candidate genes are shown in Supplementary file S5.

Expression levels of candidate genes in maize lines with different size starch granules. To verify the predicted candidate genes, 10 of them were chosen to study the differences in their expression levels within different starch granule size groups using reverse transcription and fluorescence quantitative PCR. The result are shown in Table 4 and Fig. 7. Six in the 10 selected genes, including *GRMZM2G134597*, *GRMZM2G167673*, *GRMZM2G419660*, *GRMZM2G511067*, *GRMZM2G352959* and *GRMZM2G419655*, showed significant difference at 20 d after pollination.

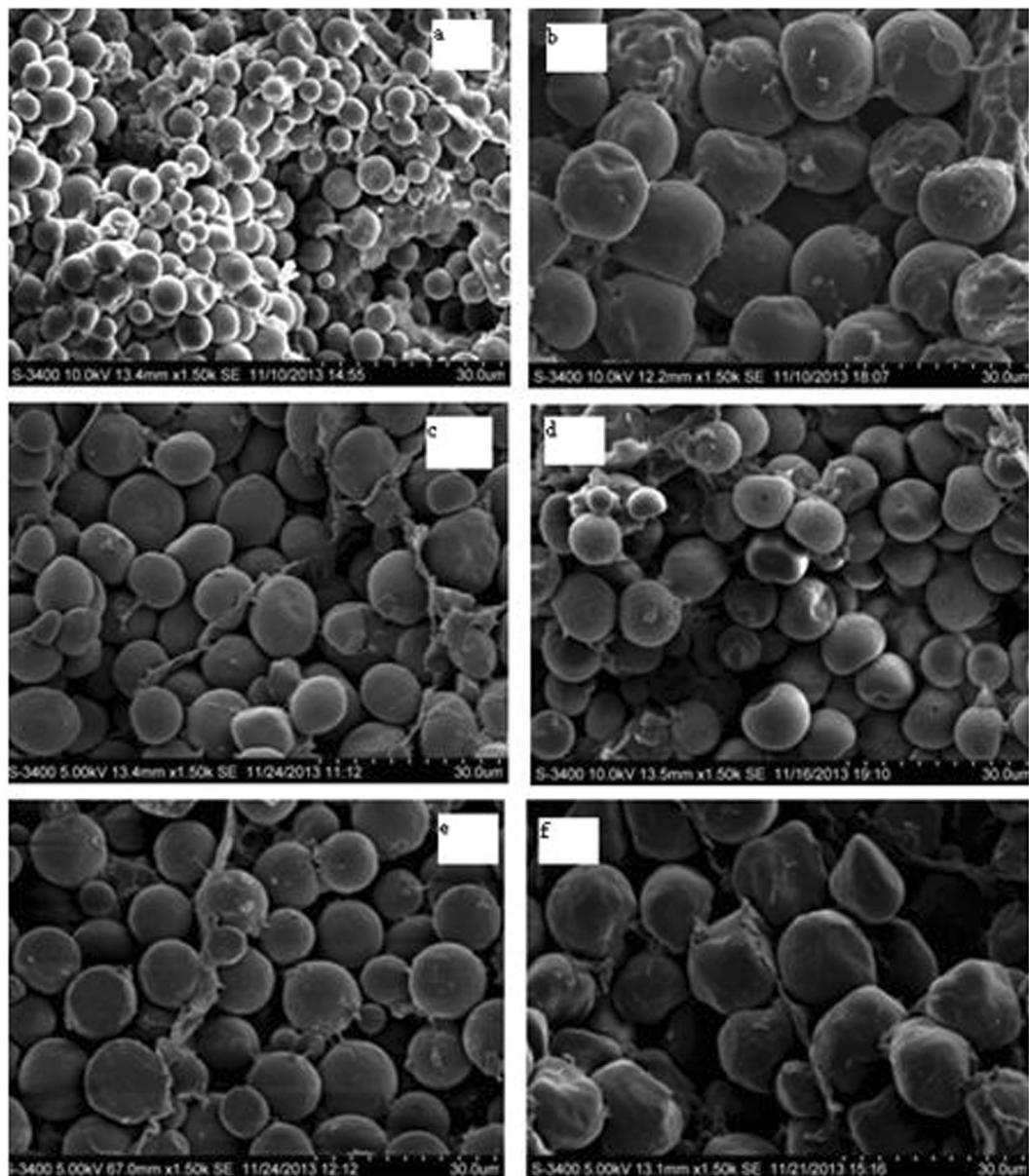


Figure 1. Scanning electron micrographs of starch granules in kernels of inbred maize lines. (a) CIMBL30; (b) CML470; (c) GEMS52; (d) 7884-4Ht; (e) 526018; (f) GEMS65.

Discussion

Analysis of the maize starch granule size. Starch is the major storage carbohydrate in cereal seeds, and the size of the starch granules is strongly associated with its end use. However, it is difficult to accurately determine the size of starch granules. To date, two techniques have been developed to analyse granule sizes: laser light scattering (LDS)³⁶ and digital image analysis (IA)³⁷. LDS for particle size analysis is simple to perform; however, the starch granule's oblate shape can cause the laser to diffract from the flat surface or narrow edge, or at obtuse angles to these surfaces, leading to system errors. Comparing LDS with IA, Wilson *et al.*³⁸ reported that LDS underestimated A-type granule diameters by ~40% and B-type granule diameters by ~50% in wheat. Edwards *et al.*²⁴ revealed that LDS measurements underestimated C-, B-, and A-type granules by maximum averages of 0.83, 3, and 23 mm, respectively. Additionally, LDS requires a prior starch extraction, which may cause artefacts to develop during extraction or precipitation³⁹. Therefore, LDS is more suitable for the analysis of totally spherical granules. Thus, LDS has often been used in the study of wheat starch but rarely in the study of maize starch.

Starch granule size has been measured by other direct methods, such as light microscopy and scanning electron microscopy (SEM). Chen *et al.*⁴⁰ analysed the size of starch granules in *Brachypodium distachyon* by SEM, while Zhang *et al.*²⁶ used a light microscope with the Zeiss software AxioVision to observe the starch granule size in potato. Compared with LDS, IA coupled with light microscopy or SEM is more direct and more readily distinguishes among individual granules, agglomerated granules, and non-starch particles. It can also simultaneously

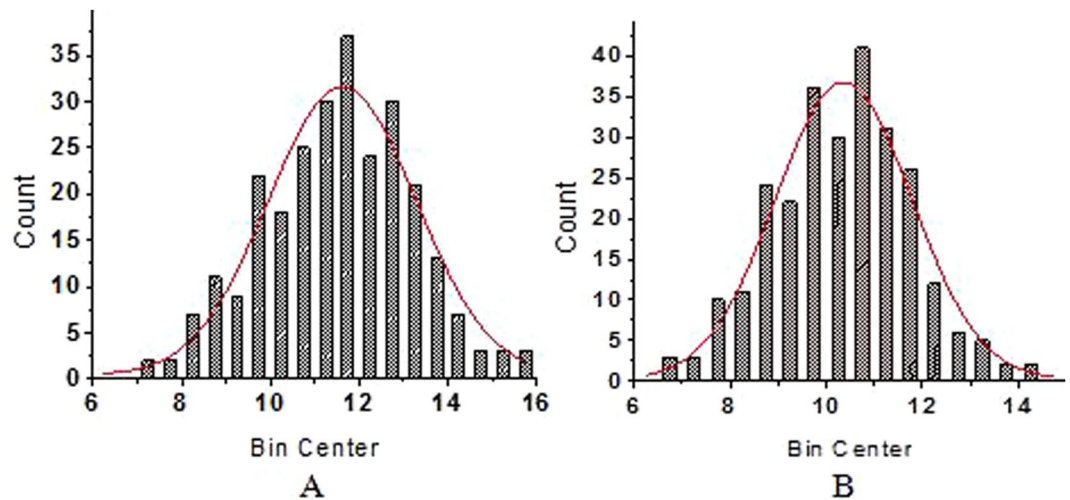


Figure 2. Frequency map of starch granule sizes. (a) Frequency of granule lengths; (b) Frequency of granule widths.

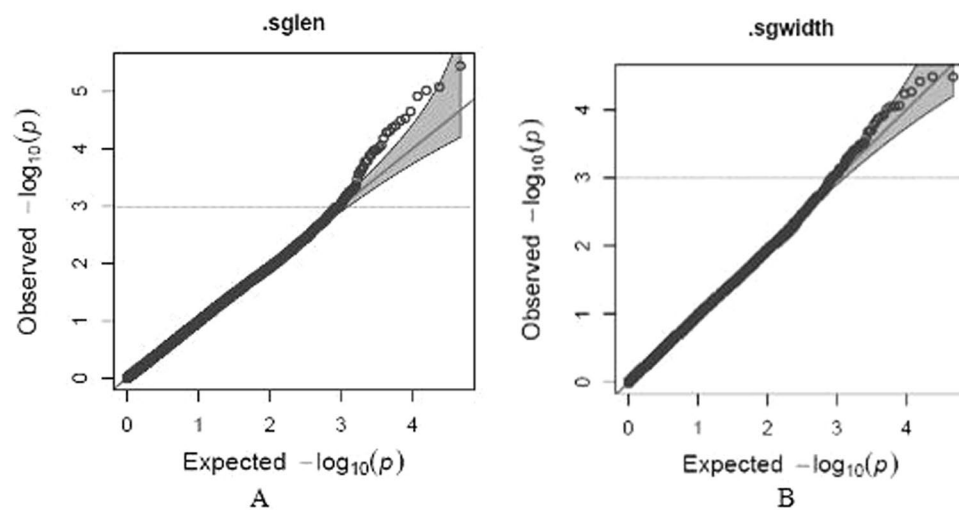


Figure 3. Quantile–quantile plot of associations with starch granule size. The p -values are shown on a –a rapid visco analyser according log10 scale, and the dashed line indicates a Bonferroni-corrected threshold of $0.1/N$. (a) Starch granule length; (b) Starch granule width.

| Trait | | SS | DF | MF | F | P |
|-----------------------|----------------|----------|-----|-------|---------|---|
| Starch granule length | Between Groups | 2211.491 | 266 | 8.314 | 122.387 | 0 |
| | Within Group | 36.071 | 531 | 0.068 | | |
| | Total | 2247.562 | 797 | | | |
| Starch granule width | Between Groups | 1554.199 | 266 | 5.843 | 104.063 | 0 |
| | Within Group | 29.814 | 531 | 0.056 | | |
| | Total | 2247.562 | 797 | | | |

Table 1. ANOVA of starch granule length and width in inbred maize lines.

record the surface features of individual particles. Considering the shape diversity of maize starch granules, IA with SEM was chosen as a more accurate method of obtaining direct data in this study.

In cereal, the size of the starch granules is an important property affecting the appropriate industrial use¹⁷. Variations in starch granule size mainly exist among inbred lines of maize, which allows for the selection of different commercial hybrids having the required granule size, resulting in improved industrial use.

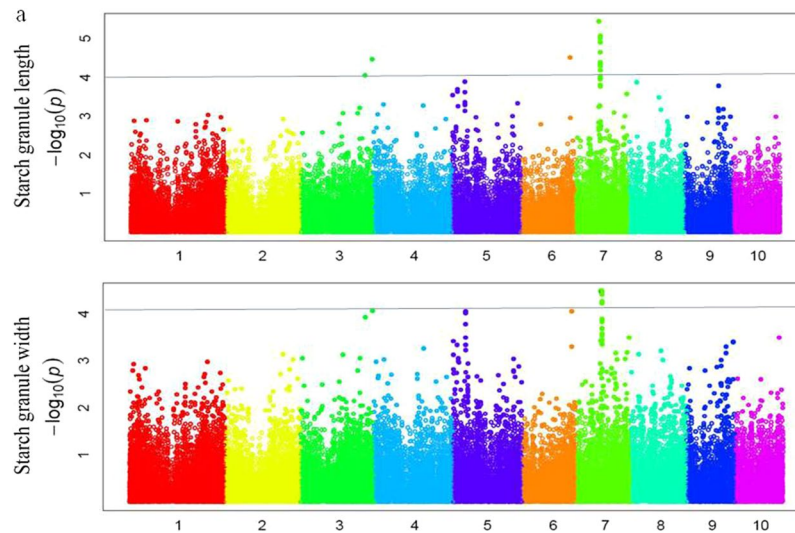


Figure 4. Manhattan plot of starch granule size.

Association analysis of maize starch granule size. Association analyses are effective tools in finding candidate genes and putative functional markers for simple and complex plant traits^{41,42}. In the current study, 14 and 9 SNPs were significantly associated with maize starch granule length and width, respectively. Among them, seven significant SNPs and five QTLs were shared between starch granule length and width (Table 2).

In a candidate gene analysis for starch granule size, *GRMZM2G167673* was predicted to encode cytochrome P450 (CYP) 714D (CYP714D). In plants, CYP is involved in several cellular processes. A homologous gene in rice encodes the CYP protein, which regulates the embryo to endosperm ratio and increases the proportion of kernel endosperm⁴³. More recently, the insertion of a 247-bp transposable element into the 3'-untranslated region of *ZmGIANT EMBRYO 2* (*ZmGE2*, encodes a CYP protein) was associated with an increased embryo to endosperm ratio⁴⁴. In the present study, the gene ontology analysis revealed that *GRMZM2G167673* has monooxygenase activity. Subfamily members of CYP, including CYP78A in maize also have monooxygenase activities, with CYP78A5, CYP78A7, and CYP78A9 regulating organ size by generating mobile growth signals that stimulate cell proliferation^{45,46}. Thus, *GRMZM2G167673* may regulate endosperm growth, determining the size of the starch granules.

Another candidate gene associated with starch granule size, *GRMZM2G419660*, encodes a protein with serine/threonine kinase activity. Serine/threonine kinases are a subfamily of calcium-dependent protein kinases (CDPKs) in plants⁴⁷; moreover, overexpressing and silencing the CDPK gene *OsCPK31* indicated that it regulates grain filling and early maturation in the Taipei 309 rice cultivar. A SEM examination showed that the starch granules increased in size when *OsCPK31* was overexpressed compared with in non-transformed controls⁴⁸. Thus, *GRMZM2G419660* in maize may also be an essential factor for the phosphorylation of sucrose synthase, which is a major enzyme involved in the starch biosynthetic pathway, similar to protein kinases in rice.

The size of the starch granules is an important factor influencing the industrial applications of starch. In the present study, inbred maize lines with different starch granules sizes were evaluated. Maize lines with different sized starch granules can be used in different industries. Moreover, SNPs or candidate genes identified in this study could be used as molecular markers to accelerate the breeding and production of plants with starch granules appropriate for different commercial purposes.

Methods

Plant materials. The investigation was based on a set of 266 inbred maize lines, containing a wide range of temperate, subtropical, and tropical germplasm^{49,50}. Because some tropical germplasm cannot mature in temperate zones, affecting the starch content and granule size, all of the inbred lines were cultivated in Sanya (Hainan Province, PR China; E 18°37', N 18°09') during the winters of 2012 and 2013. The field experiment followed a randomised complete block design with two replications. Plots were 4 m × 0.67 m and comprised 16 plants at a density of 65,250 plants per hectare. During the growing seasons, plants were irrigated and underwent common field management practices to avoid any stress.

Evaluation of starch granule size and starch paste viscosity characteristics. All of the inbred lines were self-pollinated by hand in the field, harvested when physiologically mature, and dried under natural conditions; those ears that showed abnormal development were subsequently discarded. Kernels in the middle part of each ear were then hand-threshed for starch granule size evaluation. Ten representative matured and dried kernels were selected (five kernels from each ear) and affixed to aluminium specimen stubs using double-sided adhesive tape. The samples were then sprayed with gold powder and screened using SEM (Hitachi S-3400, Tokyo, Japan) at the Centre of Biotechnology, Henan Agricultural University, Zhengzhou, China. The sizes of 20

| QTL | Position of QTL | SNP | Position of SNP | Candidate gene |
|----------------|---------------------|--|--|--|
| <i>qSGL7-1</i> | 75753279–75953279 | SNP30343 | 75853279 | GRMZM2G180104, GRMZM5G866141, GRMZM2G544148, GRMZM2G020156, GRMZM2G172345, GRMZM2G172307 |
| <i>qSGL7-2</i> | 80321800–80573170 | SNP30418 SNP30419 SNP30423 SNP30425 | 80421800 80421839 80493311 80573170 | GRMZM2G324991, AC206845.3_FG004 |
| <i>qSGL7-3</i> | 80645735–80882295 | SNP43662 SNP30432 SNP30430 | 80745735 80782295 80748880 | AC209906.3_FG001, GRMZM2G061010, GRMZM2G061014, GRMZM2G003165, GRMZM2G486223, GRMZM2G306371, GRMZM2G486201 |
| <i>qSGL7-4</i> | 80935113–81135283 | SNP30436 SNP30438 SNP30437 | 81035068 81035283 81035113 | GRMZM2G419655, GRMZM2G419660, AC196428.3_FG003, GRMZM2G177900, GRMZM2G542753 |
| <i>qSGL3-1</i> | 202581272–202781272 | SNP1148 | 202681272 | GRMZM2G511067, GRMZM2G048290, GRMZM2G347659, GRMZM2G347675, GRMZM2G054340, GRMZM2G054351, GRMZM2G054221, GRMZM5G858429, GRMZM2G054115, GRMZM5G857098, GRMZM2G591200, GRMZM2G181135, GRMZM2G702146, GRMZM2G017308, GRMZM2G017310 |
| <i>qSGL3-2</i> | 225599972–225799972 | SNP17473 | 225699972 | GRMZM2G12277, GRMZM2G033304, AC200626.3_FG008, GRMZM2G303999, GRMZM2G168049, GRMZM5G810209, GRMZM2G702310, GRMZM2G003254, GRMZM2G076245, GRMZM2G070775, GRMZM2G529436, GRMZM5G837511, GRMZM2G369956, AC209364.3_FG009, GRMZM2G369931, GRMZM2G070405, AC209364.3_FG007, GRMZM2G524240, GRMZM2G369839 |
| <i>qSGL6</i> | 154943436–155143436 | SNP28873 | 155043436 | GRMZM2G522194, GRMZM2G067073, GRMZM2G522185, GRMZM2G501825, GRMZM2G037343, GRMZM5G897015, GRMZM2G037286, GRMZM2G501821, GRMZM2G037111, GRMZM2G129031, GRMZM2G167673, GRMZM2G429714, GRMZM2G129064, GRMZM5G851528, GRMZM2G559994, GRMZM2G429720, GRMZM2G129166, AC206988.3_FG005, GRMZM2G129230, GRMZM5G834657, GRMZM2G310552, GRMZM5G896682, GRMZM2G010764, GRMZM2G010357, GRMZM2G488711 |
| <i>qSGW7-1</i> | 75753279–75953279 | SNP30343 | 75853279 | same as <i>qSGL7-1</i> |
| <i>qSGW7-2</i> | 80321800–80673170 | SNP30418 SNP30423 SNP30425 | 80421800 80493311 80573170 | same as <i>qSGL7-2</i> |
| <i>qSGW7-3</i> | 80935068–81135068 | SNP30436 | 81035068 | same as <i>qSGL7-3</i> |
| <i>qSGW3</i> | 225599972–225799972 | SNP17473 | 225699972 | same as <i>qSGL3-2</i> |
| <i>qSGW6</i> | 154943436–155143436 | SNP28873 | 155043436 | same as <i>qSGL6</i> |
| <i>qSGW5-1</i> | 39834864–40034864 | SNP22884 | 39934864 | GRMZM2G012167, GRMZM5G861448, GRMZM2G012209, AC206260.3_FG004, GRMZM2G012926, GRMZM2G012933, GRMZM2G012958, GRMZM2G013196, AC203800.3_FG005, GRMZM2G134623, GRMZM2G563606, GRMZM2G134629, GRMZM2G134597, GRMZM2G000005, AC203800.3_FG001, GRMZM2G524940, GRMZM2G435497, GRMZM2G000005, GRMZM2G492258, AC203800.3_FG006, GRMZM2G552085, GRMZM2G000007, GRMZM2G482454, GRMZM2G000009, GRMZM2G482456, GRMZM2G482457 |
| <i>qSGW5-2</i> | 39979303–40179301 | SNP22889 | 40079301 | GRMZM2G320731, GRMZM2G494899, GRMZM2G014180 |

Table 2. Quantitative trait loci, single nucleotide polymorphisms, and candidate genes for maize starch granule size. Note: SGL, starch granule length; SGW, starch granule width.

randomly selected maize starch granules were evaluated for length and width. Data were analysed using the analysis of variance method with SPSS software (IBM Corp., Armonk, NY, USA). A frequency map was constructed by Origin 8.0 software (OriginLab Corporation, Northampton, MA, USA). Sample means were used as phenotypic data for an association mapping analysis.

Four maize lines were selected to extract starch, and the pasting properties of the starch were measured using a rapid visco analyser according to Hao *et al.*⁵¹.

Genotyping and association analysis. All statistical analyses were performed using the R statistical environment (www.r-project.org). Frequency plots were also constructed by the plot function in R. Averaged data for each inbred line were used in the association analysis.

Selected inbred lines were genotyped using two genotyping platforms (RNA-sequencing and SNP array) containing 56,110 SNPs according to the method described by Yang *et al.*⁵². SNP data are available from <http://www.maizego.org/Resources.html>. SNPs with more than 12% missing data and a minor allele frequency <5% were excluded, resulting in 47,237 SNPs for further analyses. The linkage disequilibrium (LD) between SNPs on each chromosome was estimated with r^2 using TASSEL 5.0⁵³. A mixed linear model with the obtained SNPs, principal components, kinships, and the mean starch granule sizes was used for the GWAS. The relative distribution of $-\log_{10} p$ -values was observed for each SNP association and compared individually with the expected distribution using a quantile–quantile plot. The adjusted p -value threshold of significance in each trait was corrected. SNP loci

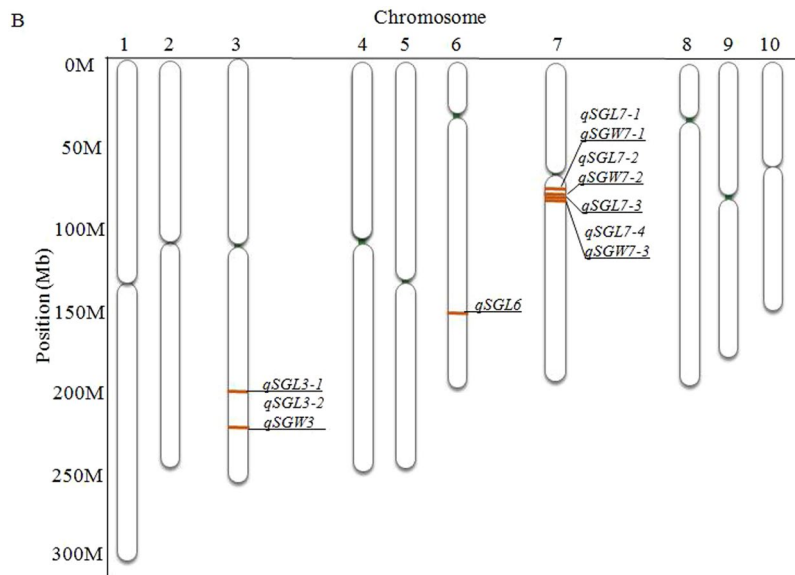


Figure 5. The chromosomal locations of the identified QTL for maize kernel starch granule.

| Score | Gene ID | Chr. | SNP pos(bp) | Distance to SNP (bp) |
|-------|---------------|------|-------------|----------------------|
| 5.44 | GRMZM2G180104 | 7 | 75853279 | 3503 |
| 5.07 | GRMZM2G419655 | 7 | 81035068 | 3127 |
| 5.07 | GRMZM2G419660 | 7 | 81035068 | 3097 |
| 4.52 | GRMZM2G167673 | 6 | 155043436 | 8547 |
| 4.47 | GRMZM6G663759 | 3 | 225699972 | 7418 |
| 4.06 | GRMZM2G511067 | 3 | 202681272 | 2118 |

Table 3. Candidate genes for maize starch granule size.

in significant LD regions were identified by revealing significant contributions to the phenotypic variations of the agronomic traits with the highest r^2 values (magnitude of marker–trait association) and lowest adjusted p -values (threshold $p < 1 \times 10^{-4}$).

The overall LD decay across the genome of this panel was 100 kb⁵⁴, thus a 100-kb region flanking the left and right sides of a SNP was defined as a QTL. If several SNPs were located closely within one LD block, the middle coordinate was chosen.

Analysis of candidate genes. The available maize genome sequence (B73) was used as the reference genome for candidate gene identification. SNP probe sequences of ~120 bp (Illumina Inc., San Diego, CA, USA) were used as queries in a BLAST algorithm-based search against the reference genome sequence in MaizeGDB (<http://www.maizegdb.org/gbrowse>). Based on the LD decay, a 200-kb window for the significant SNPs (100-kb upstream and downstream of the lead SNP) was selected to identify candidate genes. Genes within the region were identified according to the position of the closest flanking significant SNP ($p < 1 \times 10^{-4}$). The Blast2Go program was used to predict the functions of corresponding genes (<http://www.geneontology.org/>).

Sequencing and candidate gene association analysis. Three candidate genes, *GRMZM2G419655*, *GRMZM2G419660*, and *GRMZM2G511067*, were selected for sequencing based on the GWAS. DNA was extracted from seedlings of 26 maize lines with the largest starch granule sizes and 21 maize lines with the smallest starch granule sizes⁵⁵. PCR reaction mixes (20 μ l) contained 1 μ l of NEB (New England BioLabs Inc., Ipswich, MA, USA) Taq DNA Polymerase, 4 μ l of 5 \times NEB PCR Buffer, 0.5 μ l of dNTP mixture, 0.5 μ l each of the two primers, and 1 μ l of template DNA. The PCR reaction was carried out in a Bio-Rad Thermal cycler (Bio-Rad Laboratories, Inc., Hercules, CA, USA) with an initial denaturation at 94 °C for 3 min followed by 34 cycles of denaturation for 10 s at 94 °C, annealing for 1 min at 64 °C and extension for 1 min at 68 °C, with a final extension for 10 min at 68 °C. SNPs within the three candidate gene sequences were selected for the association analysis. Primers for amplification of the three genes are listed in Table 5.

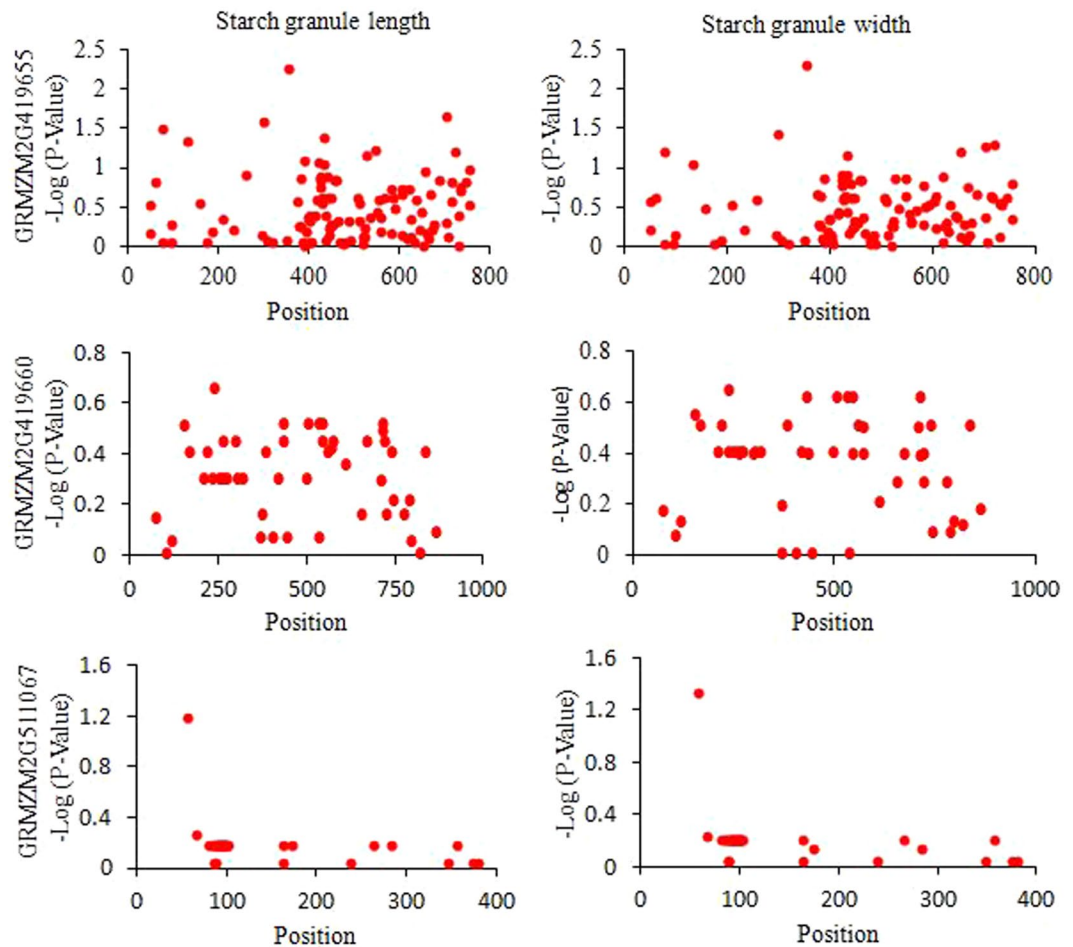


Figure 6. Candidate gene association analysis for three candidate genes and maize granule size.

| Gene ID | Mean _{SSG} | Mean _{MSG} | Mean _{LSG} | t-value _{SSG/MSG} | t-value _{SSG/LSG} | t-value _{MSG/LSG} |
|---------------|---------------------|---------------------|---------------------|----------------------------|----------------------------|----------------------------|
| GRMZM2G134597 | 1.54 ± 0.02 | 1.47 ± 0.05 | 1.38 ± 0.05 | 1.36 | 5.54** | -0.42 |
| GRMZM2G167673 | 1.46 ± 0.04 | 1.42 ± 0.04 | 1.37 ± 0.04 | 1.55 | 5.26** | 2.43* |
| GRMZM2G369956 | 1.20 ± 0.03 | 1.19 ± 0.02 | 1.18 ± 0.02 | 1.48 | 2.92* | 1.31 |
| GRMZM2G419660 | 1.37 ± 0.05 | 1.35 ± 0.08 | 1.29 ± 0.02 | 1.43 | 7.96** | 2.30* |
| GRMZM2G511067 | 1.46 ± 0.05 | 1.37 ± 0.05 | 1.32 ± 0.05 | 3.92** | 4.34** | 1.71 |
| GRMZM2G003165 | 1.48 ± 0.06 | 1.51 ± 0.05 | 1.51 ± 0.07 | -0.69 | -0.37 | 0.45 |
| GRMZM2G014180 | 1.15 ± 0.02 | 1.14 ± 0.02 | 1.14 ± 0.02 | 1.20 | 0.79 | -0.52 |
| GRMZM2G352959 | 1.17 ± 0.07 | 1.11 ± 0.05 | 1.14 ± 0.04 | 2.89** | 1.30 | -2.74* |
| GRMZM2G419655 | 1.35 ± 0.05 | 1.40 ± 0.06 | 1.28 ± 0.03 | -2.52* | 6.75** | 5.55** |
| GRMZM2G542753 | 1.23 ± 0.07 | 1.23 ± 0.05 | 1.21 ± 0.05 | 0.17 | 1.26 | 1.20 |

Table 4. Expression levels of different candidate genes for maize starch granule size. Note: SSG, gene expression of the small starch granule inbred line group; MSG, gene expression of the medium starch granule inbred line group; LSG, gene expression of the large starch granule inbred line group; t-value_{SSG/MSG}, t-value of the t-test between SSG and MSG; t-value_{SSG/LSG}, t-value of the t-test between SSG and LSG; t-value_{MSG/LSG}, t-value of the t-test between MSG and LSG; *, 0.05 significance level; **, 0.01 significance level.

Expression levels of candidate genes in maize lines having different starch granule sizes. In total, 30 maize lines with different starch granule sizes (small starch granule group: CIMBL140, CIMBL157, IRF314, CIMBL153, GEMS41, GEMS15, By804, GEMS55, BS16, and By813; medium starch granule group: 526018, Dong237, DH3732, GEMS23, TY5, CIMBL139, CIMBL102, Tie7922, GEMS17, and B113; large starch granule group: Dan360, CML325, GEMS54, CIMBL87, GEMS51, Zheng29, 835b, Ye8001, K22, and CIMBL10) were selected from the association panel to study the expression levels of candidate genes. Maize kernels (20 d after pollination) were used for RNA extraction according to the manufacturer's user manual (Transgene Biotech,

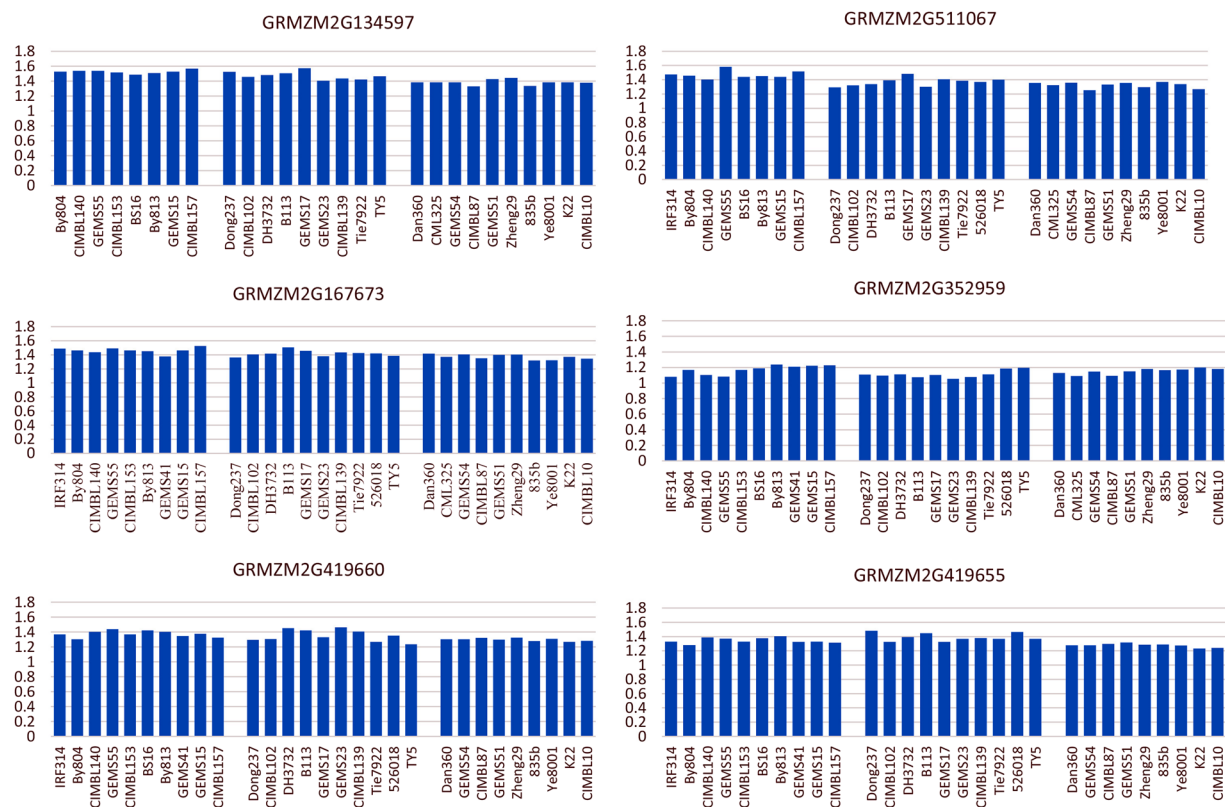


Figure 7. Different expression level of candidate genes.

| Candidate gene | | PCR Primer |
|----------------|---------|----------------------------|
| GRMZM2G511067 | Froward | 5'-AAACTTGTTCGATTTGGT-3' |
| | Reverse | 5'-CACATATAAAGGATCATAA-3' |
| GRMZM2G419655 | Froward | 5'-AATCAGTTCCTGGCAACC-3' |
| | Reverse | 5'-TCGCTTGCTCGAGCCATTAT-3' |
| GRMZM2G419660 | Froward | 5'-CTGAGAGAGGTTACCAAG-3' |
| | Reverse | 5'-CTTGTAACCTCTCTCAG-3' |

Table 5. PCR primers for candidate maize genes.

Beijing, China). The primers for the selected 10 candidate genes are shown in Table S6. A reverse transcription and fluorescence quantitative PCR analysis was conducted according to the user manual for the qPCR Master Mix (Vazyme Biotech, Nanjing, China). The actin gene was used as a reference, and all samples were analysed three times. The mean value of every sample was used for analysis.

Data Availability

<https://pan.baidu.com/s/1bpFjeLD#list/path=%2F>.

References

- Liu, J. *et al.* Functional, physicochemical properties and structure of cross-linked oxidized maize starch. *Food Hydrocolloids* **36**, 45–52 (2014).
- Jiang, L. *et al.* Multigene engineering of starch biosynthesis in maize endosperm increases the total starch content and the proportion of amylose. *Transgenic Research* **22**, 1133–1142 (2013).
- Singh, N., Singh, J., Kaur, L., Sodhi, N. & Gill, B. Morphological, thermal and rheological properties of starches from different botanical sources. *Food Chemistry* **81**, 219–231 (2003).
- Santos, M. *et al.* Changes in maize starch water sorption isotherms caused by high pressure. *International Journal of Food Science and Technology* **49**, 51–57 (2014).
- Liu, Y., Selontulyo, V. & Zhou, W. Effect of high pressure on some physicochemical properties of several native starches. *Journal of Food Engineering* **88**, 126–136 (2008).
- Paraginskia, R. *et al.* Characteristics of starch isolated from maize as a function of grain storage temperature. *Carbohydrate Polymers* **102**, 88–94 (2014).
- Preiss, J. Regulation of the biosynthesis and degradation of starch. *Annu Rev Plant Physiol* **33**, 431–454 (1982).
- Weschke, W. *et al.* Sucrose Transport into Barley Seeds: Molecular Characterization of Two Transporters and Implications for Seed Development and Starch Accumulation. *The Plant Journal* **21**, 455–467 (2000).

9. Malumba, P. *et al.* Comparative study of the effect of drying temperatures and heat-moisture treatment on the physicochemical and functional properties of corn starch. *Carbohydrate Polymers* **79**, 633–641 (2010).
10. Setiawan, S., Widjaja, H., Rakphongphairoj, V. & Jane, J. Effects of dry-ing conditions of corn kernels and storage at an elevated humidity on starch structures and properties. *Journal of Agricultural and Food Chemistry* **58**, 12260–12267 (2010).
11. Bauer, B. & Knorr, D. The impact of pressure, temperature and treatment time on starches: pressure-induced starch gelatinisation as pressure time temperature indicator for high hydrostatic pressure processing. *Journal of Food Engineering* **68**, 329–334 (2005).
12. Vittadini, E., Carini, E., Chiavaro, E., Rovere, P. & Barbanti, D. High pressure-induced tapioca starch gels: physico-chemical characterization and stability. *European Food Research and Technology* **226**, 889–896 (2008).
13. BeMiller, J. Pasting, paste, and gel properties of starch-hydrocolloid combinations. *Carbohydrate Polymers* **86**, 386–423 (2011).
14. Man, J. *et al.* Structural Changes of High-Amylose Rice Starch Residues following *In Vitro* and *In Vivo* Digestion. *Journal of Agricultural and Food Chemistry* **60**, 9332–9341 (2012).
15. Svihus, B., Uhlen, A. & Harstad, O. Effect of starch granule structure, associated components and processing on nutritive value of cereal starch: A review. *Animal Feed Science and Technology* **122**, 303–320 (2005).
16. Matsushima, R. *et al.* Amyloplast-localized SUBSTANDARD STARCH GRAIN4 protein influences the size of starch grains in rice endosperm. *Plant Physiol* **162**, 623–636 (2014).
17. Lindeboom, N., Chang, P. & Tyler, R. Analytical, biochemical and physicochemical aspects of starch granule size, with emphasis on small granule starches: a review. *Starch-Stärke* **56** (2004).
18. Malinski, E., Daniel, J., Zhang, X. & Whistler, R. Isolation of small starch granules and determination of their fat mimic characteristics. *Cereal Chem* **80**, 1–4 (2003).
19. Kaur, L., Singh, J., McCarthy, O. & Singh, H. Physico-chemical, rheological and structural properties of fractionated potato starches. *J. Food Eng* **82**, 383–394 (2007).
20. Singh, N. & Kaur, L. Morphological, thermal, rheological and retrogradation properties of potato starch fractions varying in granule size. *Journal of the Science of Food & Agriculture* **84**, 1241–1252 (2004).
21. Al-Rabadi, G. J. S., Gilbert, R. G. & Gidley, M. J. Effect of particle size on kinetics of starch digestion in milled barley and sorghum grains by porcine alpha-amylase. *Journal of Cereal Science* **50**, 198–204 (2009).
22. Mahasukhonthachat, K., Sopade, P. & Gidley, M. Kinetics of starch digestion in sorghum as affected by particle size. *Journal of Food Engineering* **96**, 18–28 (2010).
23. Dhital, S., Shrestha, A. & Gidley, M. Relationship between granule size and *in vitro* digestibility of aize and potato starches. *Carbohydrate Polymers* **82**, 480–488 (2010).
24. Edwards, M., Osborne, B. & Henry, R. Effect of endosperm starch granule size distribution on milling yield in hard wheat. *Journal of Cereal Science* **48**, 180–192 (2008).
25. Dhital, S., Shrestha, A., Hasjim, J. & Gidley, M. Physicochemical and structural properties of maize and potato starches as a function of granule size. *Journal of agricultural and food chemistry* **59**, 10151–10161 (2011).
26. Zhang, J. *et al.* Starch granule size variation and relationship with tuber dry matter content in heritage potato varieties. *Scientia Horticulturae* **130**, 503–509 (2011).
27. Batey, I. *et al.* Genetic mapping of commercially significant starch characteristics in wheat crosses. *Crop & Pasture Science* **52**, 1287–1296 (2001).
28. Borém, A., Mather, D., Rasmusson, D., Fulcher, R. & Hayes, P. Mapping quantitative trait loci for starch granule traits in barley. *Journal of Cereal Science* **29**, 153–160 (1999).
29. Korte, A. & Farlow, A. The advantages and limitations of trait analysis with GWAS: a review *Plant Methods* **9** (2013).
30. Li, H. *et al.* Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet* **45**, 43–50 (2013).
31. Brachi, B. *et al.* Linkage and association mapping of Arabidopsis thaliana flowering time in nature29. *PLoS Genetics* **6**, e1000940 (2010).
32. Stoddard, F. Survey of Starch Particle-Size Distribution in Wheat and Related Species. *Cereal Chem* **76**, 145–149 (1999).
33. Muñoz, L. A., Pedreschi, F., Leiva, A. & Aguilera, J. M. Loss of birefringence and swelling behavior in native starch granules: Microstructural and thermal properties. *Journal of Food Engineering* **152**, 65–71 (2015).
34. Feng, N., He, Z., Zhang, Y., Xia, X. & Zhang, Y. QTL mapping of starch granule size in common wheat using recombinant inbred lines derived from a PH82-2/Neixiang 188 cross. *The Crop Journal*, 166–171 (2013).
35. Tester, R., Karkalas, J. & Qi, X. Starch structure and digestibility enzyme–substrate relationship. *World's Poultry Science Journal* **60**, 186–195 (2004).
36. Borch, J., Sarko, A. & Marchessault, R. Light scattering analysis of starch granules. *Journal of Colloid and Interface Science* **41**, 574–587 (1972).
37. Zayas, I., Bechtel, D., Wilson, J. & Dempster, R. *Digital image analysis of starch granules for recognizing hard red and soft red winter wheats* (1993).
38. Wilson, J., Bechtel, D., Todd, T. & Seib, P. Measurement of wheat starch granule size distribution using image analysis and laser diffraction technology. *Cereal Chemistry* **83**, 259–268 (2006).
39. Li, X. *et al.* Effects of sampling methods on starch granule size measurement of potato tubers under a light microscope. *International journal of plant biology* **2**, 118–122 (2011).
40. Chen, G. *et al.* Dynamic development of starch granules and the regulation of starch biosynthesis in Brachypodium distachyon: comparison with common wheat and Aegilops peregrine. *BMC Plant Biology* **14**, 198 (2014).
41. Wen, Z. *et al.* Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genomics* **15**, 809–820 (2014).
42. Sun, F. *et al.* Identification of stable QTLs for seed oil content by combined linkage and association mapping in Brassica napus. *Plant Science* **252**, 388–399 (2016).
43. Park, D. *et al.* Molecular characterization and physico-chemical analysis of a new giant embryo mutant allele (*ge'*) in rice (*Oryza sativa* L.). *Genes & Genomics* **31**, 277–282 (2009).
44. Zhang, P. *et al.* A transposable element insertion within ZmGE2 gene is associated with increase in embryo to endosperm ratio in maize. *Theoretical and Applied Genetics* **125**, 1463–1471 (2012).
45. Anastasiou, E. *et al.* Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. *Developmental Cell* **13**, 843–856 (2007).
46. Adamski, N. M., Anastasiou, E., Eriksson, S., O'Neill, C. M. & Lenhard, M. Local maternal control of seed size by KLUH/CYP78A5-dependent growth signaling. *Proceedings of the National Academy of Science* **106**, 20115–20120 (2009).
47. Asano, T. *et al.* Rice SPK, a calmodulin-like domain protein kinase, is required for storage product accumulation during seed development: phosphorylation of sucrose synthase is a possible factor. *Plant Cell* **14**, 619 (2002).
48. Manimaran, P., Mangrauthia, S., Sundaram, R. & Balachandran, S. Constitutive expression and silencing of a novel seed specific calcium dependent protein kinase gene in rice reveals its role in grain filling. *Journal of Plant Physiology* **174**, 41–48 (2015).
49. Setter, T. *et al.* Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. *Journal of Experimental Botany* **62**, 701–716 (2010).
50. Liu, N., Xue, Y., Guo, Z., Li, W. & Tang, J. Genome-Wide Association Study Identifies Candidate Genes for Starch Content Regulation in Maize Kernels. *Frontiers in Plant Science* **7**, 1–8 (2016).

51. Hao, D. *et al.* Genetic dissection of starch paste viscosity characteristics in waxy maize revealed by high-density SNPs in a recombinant inbred line population. *Molecular Breeding* **37**, 50 (2017).
52. Yang, N. *et al.* Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genetics* **10**, e1004573 (2014).
53. Bradbury, P. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
54. Xue, Y. *et al.* Genome-wide association analysis for nine agronomic traits in maize under well-watered and water-stressed conditions. *Theor Appl Genet* **126**, 2587–2596 (2013).
55. Healey, A., Furtado, A., Cooper, T. & Henry, R. J. Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **10**, 21–28 (2014).

Acknowledgements

We would like to thank Prof. Yan Jianbing for materials and SNP data support. This research was supported by the State Key Basic Research and Development Plan of China (2014CB138203), the National Natural Science Foundation of China (91335205), the Fundamental Research Funds for the Henan Provincial Colleges (2015RCJH05) and the Major Science and Technology Projects in Henan Province (161100110500). We also thank Lesley Benyon (PhD, from Liwen Bianji, Edanz Group China, www.liwenbianji.cn/ac) and Cankui Zhang (Associate professor, Purdue University), for editing the English text of a draft of this manuscript.

Author Contributions

This study was conceived by J.T. and N.L. The GWAS and G.O. analyses were conducted by Y.X., Z.Z., and Y.H. Maize collection was performed by W.L. and J.H. qRT-PCR was conducted by S.M. N.L. wrote the manuscript. All of the authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-31863-y>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018