Sequence polymorphism of the *waxy* gene in waxy maize accessions and characterization of a new *waxy* allele

Meijie Luo\(^1,2\), Yaxing Shi\(^1,2\), Yang Yang\(^1,2\), Yanxin Zhao\(^1\), Yunxia Zhang\(^1\), Yamin Shi\(^1\), Mengsi Kong\(^1\), Chunhui Li\(^1\), Zhen Feng\(^1\), Yanli Fan\(^1\), Li Xu\(^1\), Shengli Xi\(^1\), Baishan Lu\(^1,2\) & Jiuran Zhao\(^1,2\)

Waxy maize has many excellent characteristics in terms of its nutritional and economic value. In recent decades, the waxy maize germplasm has increased dramatically as a result of different selection methods. We collected 200 waxy maize inbred accessions from different origins to study their genetic diversity and phylogenetic relationships, and to identify new *waxy* mutations. A simple sequence repeat (SSR) analysis revealed wide genetic diversity among the 200 waxy maize accessions. The maize accessions were clustered into three groups. We sequenced the *waxy* gene from the first to the 14th exon. Nucleotide variation analysis of 167 waxy maize and 14 flint maize lines revealed some nucleotide differences in the *waxy* gene among different waxy maize groups, and much narrower nucleotide diversity in waxy maize than in flint maize. In a phylogenetic analysis, waxy maize carrying the same mutation allele clustered together, and waxy maize carrying different mutation alleles distributed in different groups; waxy maize was intermixed with flint maize in each branch, and *wx-D7* waxy maize separated significantly from waxy maize lines carrying *wx-D10*, *wx-124* and *wx-hAT* mutant alleles. The *wx-hAT* was a new *waxy* mutation identified in this study. It consisted of a 2286-bp transposon inserted into the middle of exon three of the *waxy* gene. A PCR marker specific for the *wx-hAT* allele was developed. These results will be useful for the utilization and preservation of the waxy maize germplasm, and the PCR marker has potential uses in waxy maize breeding programs.

Waxy maize, also known as sticky maize, has high economic, nutritional, and processing value\(^1,2\). The starch in the endosperm of waxy maize is nearly 100% amylopectin, which confers the sticky quality to maize grains. It is mainly consumed as food in Asia, and is also an important ingredient in the textile and paper industries. Waxy maize was first discovered in China in 1908, and was later reported in other locations of Asia\(^3\). The results of several studies suggest that the southwest region of China, particularly Yunnan Province and its surrounding areas, is the central origin of Chinese waxy maize\(^3\). In recent decades, many adapted maize lines have been developed for hybrid seed production by different selection methods, but the relationships among these waxy maize lines are unclear.

The glutinous genotypes of waxy maize are null mutations of the *waxy* gene, which encodes the granule bound starch synthase (GBSSI) that is necessary for amylose synthesis\(^5\). The wild-type *waxy* gene in maize is 3.93 kb long, located on chromosome 9, and composed of 14 exons. Insertions and deletions in the DNA sequence are the main types of *waxy* mutations in maize\(^6\). Globally, more than 50 mutations in the *waxy* gene have been characterized at the molecular level\(^7\). In Chinese waxy maize, the two deletion mutations *wx-D7* and *wx-D10* (a 30-bp deletion in the seventh exon and a 15-bp deletion in the tenth exon, respectively) are the main *waxy* alleles\(^8,9\). Transposable element insertions are another type of *waxy* gene mutation\(^8\). Transposons can be divided into DNA transposons and RNA transposons according to their transposition mechanism. Members of the DNA-transposon family can jump from one gene to another in a cut-and-paste fashion to produce unstable alleles.
mutations\textsuperscript{11,12}. Among the identified mutations in the maize \textit{waxy} gene, \textit{wx-m9}, \textit{wx-m5}, \textit{wx-B3}, \textit{wx-m1}, \textit{wx-B4}, \textit{wx-m6}, and \textit{wx-m7} are insertion mutations of Ac/Ds transposable elements, \textit{wx-m8} is an insertion mutation of the dSpm element, and \textit{wx-844} is a mutation caused by the En/Spm element\textsuperscript{13,15}. The Ac/Ds, dSpm, and En/Spm elements are all members of the DNA-transposon family\textsuperscript{14} and their \textit{waxy} alleles are important resources for transposon research. RNA transposons move within the genome using a copy-and-paste mechanism via reverse transcription of an RNA intermediate\textsuperscript{1}. Among the known mutant alleles of the \textit{waxy} gene in maize, \textit{wx-stonor}, \textit{wx-B5}, \textit{wx-G}, \textit{wx-M}, \textit{wx-I}, \textit{wx-K}, \textit{wx-Cin4} and \textit{wx-Rina} are RNA transposon insertion mutations\textsuperscript{3,8}. RNA transposons result in stable mutations and can be used as markers to identify \textit{waxy} mutation loci. However, there are still many \textit{waxy} maize lines with unknown \textit{waxy} alleles.

There is wide genetic diversity at the \textit{waxy} locus among \textit{waxy} maize accessions. Previous studies on \textit{waxy} diversity have shown that different \textit{waxy} mutants have independent origins\textsuperscript{10}. Compared with non-glutinous maize, Chinese \textit{waxy} maize shows much narrower nucleotide diversity at the \textit{waxy} locus, suggesting that there has been strong selection in the \textit{waxy} genomic region during \textit{waxy} maize breeding\textsuperscript{11,14}. An association analysis between allelic variations of \textit{waxy} and starch physicochemical properties showed that \textit{waxy} allelic variation affects the gel consistency, gelatinization temperature, and pasting viscosity properties of rice starch, implying that certain \textit{waxy} alleles have been favored for grain quality improvement\textsuperscript{15,16}. Therefore, comparison of \textit{waxy} sequence variation among maize germplasm accessions not only provides insights into selection at the \textit{waxy} locus during domestication, but also can highlight mutations in the \textit{waxy} gene that will be useful for the germplasm utilization and quality breeding of \textit{waxy} maize\textsuperscript{16,17}.

In this study, we collected \textit{waxy} maize breeding accessions from Jilin province, Shanxi province, and Beijing in China, as well as from Korea. First, we used simple sequence repeat (SSR) markers to study the genetic diversity of the \textit{waxy} maize inbred lines. Then we sequenced the \textit{waxy} genes from the first to the fourteenth exon to analyze sequence variations and the phylogenetic relationships among \textit{waxy} maize lines. Additionally, a newly identified allele of the \textit{waxy} gene in \textit{waxy} maize was characterized.

\section*{Results}

\subsection*{Genetic analyses of \textit{waxy} maize.}

The genetic diversity of the 200 \textit{waxy} maize inbred lines (Supplementary Table S2) was evaluated using SSR markers. In total, 458 alleles were found at the 40 SSR loci, with a range of 2 to 25 alleles per marker. The average number of alleles per marker locus across genotypes was 11.45, about double that obtained by Zheng et al.\textsuperscript{14} using 20 SSR markers and 165 accessions. The polymorphism information content (PIC) values for the 40 SSR loci ranged from 0.17 to 0.89 (average, 0.7). The genetic similarity coefficient was analyzed using the SSR data. The genetic similarity coefficient of 200 \textit{waxy} maize inbred lines ranged from 0.03 to 0.95, with an average of 0.31, and 86.79\% of them were less than 0.45 (Fig. 1a). On the basis of similarities of SSR data, a cluster analysis of the 200 \textit{waxy} maize inbred lines was performed using the neighbor-joining method\textsuperscript{18}. The cluster analysis grouped the 200 \textit{waxy} inbred lines into three main groups: group A included 63 accessions, group B included 59 accessions, and group C included 78 accessions (Fig. 1b). These results show that there was wide genetic diversity among the tested \textit{waxy} maize accessions.

The genetic distances of \textit{waxy} maize inbred lines in group A, B and C ranged from 0.11 to 0.97, 0.05 to 0.94, 0.06 to 0.96, with the average values of 0.68, 0.66 and 0.67, respectively. All the three groups contained \textit{waxy} maize inbred lines originated from Beijing and Jilin province of China, which indicated that \textit{waxy} maize with different genetic background had been widely used in China, which was consistent with the rapid development of \textit{waxy} maize breeding and industry in China in recent years. It is worth mentioning that the \textit{waxy} maize inbred lines originated from Shanxi Province of China and Korea only existed in group C, which indicated that the \textit{waxy} maize inbred lines from Shanxi Province of China had similar genetic background, and those from Korea had similar genetic background (Fig. 1b). As expected, \textit{waxy} maize inbred lines with similar pedigree were clustered in the same group. For example, \textit{waxy} maize inbred lines JYN3, JYN4, JYN5, JYN6, JYN7 and JYN9 all had similar pedigree and were clustered in group A; the \textit{waxy} maize inbred lines DN1, DN2, DN3 and DN4 had similar pedigree and were clustered in group C (Supplementary Table S2 and Fig. 1b).

\subsection*{Nucleotide variation at \textit{waxy} locus in \textit{waxy} maize.}

Nucleotide sequences from the first to the 14\textsuperscript{th} exon of the \textit{waxy} genes were determined using four pairs of primers (Supplementary Fig. S1 and Table S1). We examined the DNA nucleotide variations in an approximately 3523-bp region of \textit{waxy} loci in 167 \textit{waxy} and 14 flint maize accessions (Supplementary Data File S1). In this region, the variable nucleotide sites of 169 \textit{waxy} maize and 14 flint maize were 57 and 87, respectively (Table 1). The genetic variation in \textit{waxy} gene was compared among the three different \textit{waxy} maize groups and the flint maize. All three \textit{waxy} maize groups and all \textit{waxy} maize in three groups showed a minimum level of genetic variation at the \textit{waxy} locus. The estimate of nucleotide diversity for \textit{waxy} maize accessions in group A, group B, group C and all \textit{waxy} maize in three groups was 0.00173, 0.00097, 0.00228 and 0.00150, respectively, indicating that the genetic diversity of the \textit{waxy} locus differs among \textit{waxy} maize populations with different genetic backgrounds. Apparent nucleotide diversity was observed at the \textit{waxy} locus in flint maize (8.4 fold) than in \textit{waxy} maize. Consistently, the values of $S$ (number of polymorphic sites), and $K$ (average number of pairwise nucleotide differences) were lower in \textit{waxy} maize accessions than in non-glutinous maize accessions (Table 1). The significant reduction in diversity at the \textit{waxy} locus in \textit{waxy} maize suggests that modern \textit{waxy} maize has experienced a genetic bottleneck during its domestication.

The Tajima's $D$ and Li & Fu's $D^*$ and $F^*$ values were calculated to test the deviation from the neutral equilibrium model. All three tests identified negative selection at the \textit{waxy} locus in \textit{waxy} maize populations, but not in flint maize, suggesting that there has been strong selection acting on \textit{waxy} maize accessions (Table 1).
Figure 1. Genetic similarity coefficient and cluster analysis of 200 maize accessions using SSR data. (a) Genetic similarity coefficient frequency distribution. The genetic similarity coefficient was calculated by SSRAnalyzer V1.0 (Software copyright registration number: 2018SR003610). (b) Cluster analysis of waxy lines. Cluster analysis based on allele identity was carried out using PowerMarker V3.25 with the neighbor-joining method. Different colors of taxon names represent different mutant alleles in waxy gene. Different colors of subtree markers represent different origin regions of waxy maize. Different colors of branch lines represent different waxy maize groups. wx-hAT: waxy maize with wx-hAT mutant allele; wx-D7: waxy maize with wx-D7 mutant allele; wx-D10: waxy maize with wx-D10 mutant allele; wx-D14: waxy maize with wx-D14 mutant allele; Other: waxy maize had other mutation in waxy gene, which was different from wx-hAT, wx-D7, wx-D10 and wx-D14; Not analyzed: the sequence of these waxy maize lines were not obtained; Beijing, China: waxy maize originated from Beijing city in China; Jilin, China: waxy maize originated from Jilin province of China; Shanxi, China: waxy maize originated from Shanxi province of China; Korea: waxy maize originated from Korea; Group A: waxy maize classified into group A; Group B: waxy maize classified into group B; Group C: waxy maize classified into group C.

Table 1. Summary of nucleotide diversity for waxy gene within maize taxa. N: number of sequences, Site: number of sites (excluding sites with gaps/missing data), S: variable (polymorphic) sites, h: number of haplotypes, Pi: nucleotide diversity, K: average number of pairwise nucleotide differences, D: Tajima’s D, D*: Fu and Li’s D* test, F*: Fu and Li’s F* test, Rm: Minimum number of recombination events. *P < 0.05, **P < 0.01.
The haplotypes of \textit{waxy} gene in waxy maize and flint maize were calculated using DnaSP\textsuperscript{19} software and the results were shown in Table 2. Thirty haplotypes were detected in 167 waxy maize accessions, among which 117 lines shared haplotype Hap_19. Eight haplotypes were detected in 14 flint maize accessions.

Phylogenetic analysis based on sequence polymorphisms. We obtained the sequences of the 3523-bp region at the \textit{waxy} loci in 167 waxy and 14 flint maize. \textit{Waxy} sequence data for eight wild relatives of maize and seven landraces from Southwestern China were downloaded from the GenBank database. Based on these sequences, a phylogenetic tree including waxy maize inbred lines, flint maize, waxy maize landraces and their relatives was constructed by the neighbor-joining method\textsuperscript{20}. According to the tree, five wild relatives of maize formed a branch that was basal to five separate branches. All five branches contained waxy maize and flint maize, indicating that a number of glutinous maize accessions may have been developed from domesticated non-glutinous maize. Obviously, maize with the same mutation allele clustered together, and waxy maize in different branches carried different mutation alleles. Five waxy maize inbred lines and seven waxy maize landraces

| Haplotype | Variation position | Frequency |
|-----------|--------------------|-----------|
| **Waxy maize** |                   |           |
| Hap_1     | CCA GAGCTACCTCGCTTGAGATAGGAAAAACCAATGTACCAATTCGCCGAAAGTATTTG | 2          |
| Hap_2     | ...............TGTCGG........ | 3          |
| Hap_3     | ...............C............ | 1          |
| Hap_4     | ...........A.A.G.G........... | 1          |
| Hap_5     | ...........A.A.G.G........... | 1          |
| Hap_6     | ...........A.A.C....TGA........GCT..CC.CA....... | 5          |
| Hap_7     | ...........A.A.C....GTG...CTC...CT....TT........ | 1          |
| Hap_8     | ...........A.A.......TA............ | 1          |
| Hap_9     | G.C...........T............G.A........ | 1          |
| Hap_10    | G...........T............G.A........ | 1          |
| Hap_11    | G...........T............G.A........ | 1          |
| Hap_12    | G...........T............G.A........ | 1          |
| Hap_13    | G...........T............G.A........ | 1          |
| Hap_14    | G...........T............G.A........ | 1          |
| Hap_15    | G...........T............G.A........ | 1          |
| Hap_16    | G...........T............G.A........ | 1          |
| Hap_17    | G...........T............G.A........ | 1          |
| Hap_18    | G...........T............G.A........ | 1          |
| Hap_19    | G...........T............G.A........ | 1          |
| Hap_20    | G...........T............G.A........ | 1          |
| Hap_21    | G...........T............G.A........ | 1          |
| Hap_22    | G...........T............G.A........ | 1          |
| Hap_23    | G...........T............G.A........ | 1          |
| Hap_24    | G...........T............G.A........ | 1          |
| Hap_25    | G...........T............G.A........ | 1          |
| Hap_26    | G...........T............G.A........ | 1          |
| Hap_27    | G...........T............G.A........ | 1          |
| Hap_28    | G...........T............G.A........ | 1          |
| Hap_29    | G...........T............G.A........ | 1          |
| Hap_30    | G...........T............G.A........ | 1          |
| **Flint maize** |                   |           |
| Hap_1     | TCGTCTCCGCTCGGTGTTGACTTATCGTATCTGGGCCCGGCTCGCTGACCACTCCTACACTACTACGACTCAACCCCTGCGGCGCGC | 1          |
| Hap_2     | CTAATTCGCGGATAGTTCCTACACGTGGCCTTTGGCAGAGTTTTGCGAGGCCATTTTTGCGACGACGGTACGCTGGACGGCCAAACTGA | 3          |
| Hap_3     | CTAATTCGCGGATAGTTCCTACACGTGGCCTTTGGCAGAGTTTTGCGACGACGGTACGCTGGACGGCCAAACTGA | 2          |
| Hap_4     | ...............GTGGGGA........ | 3          |
| Hap_5     | ...............GTGGGGA........ | 3          |
| Hap_6     | ...............GTGGGGA........ | 3          |
| Hap_7     | ...............GTGGGGA........ | 3          |
| Hap_8     | ...............GTGGGGA........ | 3          |

Table 2. Haplotype of \textit{waxy} gene in studied 167 waxy maize and 14 flint maize.
formed a branch, and these maize accessions carried the \(wx-D10\) allele in \(waxy\) gene. In another branch, eight \(waxy\) maize with the \(wx-hAT\) allele formed a subgroup, and then clustered with one \(waxy\) maize harboring the \(wx-124\) mutant allele. The other two branches contained two and four \(waxy\) maize inbred lines, respectively. No insertion or deletion mutations were detected in the amplification region of \(waxy\) gene in these maize lines, suggesting that they had other mutation alleles in other regions of the \(waxy\) gene. Furthermore, the remaining 147 \(waxy\) inbred lines carrying the \(wx-D7\) mutant allele clustered together and formed an independent branch, which was significantly separated from \(waxy\) maize carrying \(wx-D10\), \(wx-124\) and \(wx-hAT\) mutant alleles (Fig. 2).

**Identification of accessions with novel \(waxy\) genotype.** After sequencing, four types of \(waxy\) mutation were identified among the \(waxy\) maize accessions: 147 accessions carried the \(wx-D7\) allele with a 30-bp deletion in the seventh exon–intrin region; five maize accessions carried the \(wx-D10\) allele with a 15-bp deletion in the 10th exon region; one maize accession carried the \(wx-D124\) allele with a 125-bp insertion in the seventh exon; and eight maize accessions carried an allele with a new insertion mutation. This mutation was identified as a 2.2 kb insertion in the middle of the third exon of \(waxy\) gene, which would lead to abnormal gene coding (Fig. 3). The conservative domain of \(waxy\) gene was analyzed in Pfam database (https://pfam.xfam.org/search/sequence), and found that the 2.2 kb insertion mutation was located on the conserved starch synthase catalytic domain of \(waxy\) gene. The amyllopectin content analysis showed that all eight of these accessions had a high amyllopectin content (> 94.5%) in grain starch, like that in \(wx-D7\) mutant maize lines (Supplementary Table S3). Consistently, the GBSS activity in seeds of these accessions (19.9–28.1 nmol/min/g) was remarkably lower than that in wild type flint maize (66.7–79.3 nmol/min/g) (Supplementary Table S4).

As shown in Fig. 1b, \(waxy\) maize in the same group had different origins and carried different \(waxy\) mutation alleles based on cluster analysis using SSR data. \(Waxy\) maize from the same origin possessed different \(waxy\) mutation alleles. Therefore, based on cluster analysis, there was no correlation among the groups, origins and \(waxy\) mutation alleles in \(waxy\) maize in this study.

**Figure 2.** Phylogenetic analysis of maize accessions based on \(waxy\) gene. The neighbor-joining phyllogenetic tree based on the Kimura 2-parameter model was constructed with MEGAX64 software using \(waxy\) gene sequence data with 1000 bootstrap replicates to assess tree reliability. Different colors of branch lines represent different groups. Different colors of taxon names represent \(waxy\) maize inbred lines carrying different \(waxy\) gene mutation alleles, and flint maize, wild relatives of maize, as well as landraces from Southwestern China. Subtree markers pointed out landraces from Southwestern China and \(waxy\) maize inbred lines from Shanxi province of China. \(wx-hAT\): \(waxy\) maize with \(wx-hAT\) mutant allele; \(wx-D7\): \(waxy\) maize with \(wx-D7\) mutant allele; \(wx-D10\): \(waxy\) maize with \(wx-D10\) mutant allele; \(wx-124\): \(waxy\) maize with \(wx-124\) mutant allele; Other mutation: \(waxy\) maize had other mutation in \(waxy\) gene, which was different from \(wx-hAT\), \(wx-D7\), \(wx-D10\) and \(wx-124\); Shanxi, China: \(waxy\) maize inbred lines originated from Shanxi province of China; Southwestern China: landraces from Southwestern China.
Molecular characteristics of \textit{wx-hAT}. The newly identified insertion mutation was 2286 bp in length and was inserted after the 43rd nucleotide in the third exon of the \textit{waxy} gene (Fig. 3a). Transposon analysis using the CENSOR program\(^{21}\) indicated that the 2.2-kb insertion contained part of the \textit{hAT} transposon with sequences ranging in length from 813 to 1558 bp (Fig. 3c). We named this \textit{waxy} allele \textit{wx-hAT}. Further analyses revealed that \textit{wx-hAT} had a 5'-CAG-3' terminal inverted repeat (TIR) and a 5'-GGC GGA TCT-3' near-terminal inverted repeat, generated a 5'-AGT ACA AG-3' target site repeat (TSD). The 3-bp TIR was flanked by the 8-bp TSD and the 3-bp and the 9-bp TIRs were separated by one nucleotide (Fig. 3c). BLAST searches in the MaizeGDB database revealed that the sequences from 9 to 1559 bp and from 1561 to 2286 bp of \textit{wx-hAT} showed 99.03% (\textit{e}-value = 0) and 99.59% (\textit{e}-value = 0) identities to the regions of 27,081,986–27,083,536 bp and 27,084,569–27,085,294 bp on chromosome 4 (Supplementary Fig. S2), respectively. This result suggests that this transposon might have jumped from chromosome 4 to the \textit{waxy} gene on chromosome 9. Those regions of chromosome 4 were not annotated.

**Intragenic selection marker for recessive \textit{waxy} gene.** As shown in Fig. 3b, using the primers of E1-4F/E1-4R two amplicons 3.1 kb and 884 bp in length were amplified from \textit{waxy} lines and flint lines, respectively. We anticipated that the mutation of the \textit{waxy} gene in these \textit{waxy} lines was caused by the insertion of the 2.2 kb fragment in its allelic dominant gene. However, the heterozygous type and the wild type of the \textit{wx-hAT} allele could not be differentiated by analyzing fragments amplified using this set of PCR primers, because only the 884 bp fragment amplified from the wild-type \textit{waxy} gene was amplified in the heterozygous \textit{wx-hAT} allele, which may due to the effect of long fragment amplification being affected by the priority amplification of short fragment. A forward primer \textit{waxyF2} was designed on the basis of the 2.2 kb insertion sequence (Supplementary Table S1). Using the three primers E1-4F, E1-4R and \textit{waxyF2}, a 608-bp and an 884-bp fragment were amplified for the homozgyous mutant \textit{wx-hAT} gene and the wild type \textit{wx-hAT} gene, respectively, while both the 608-bp and the 884-bp fragments were amplified for the heterozygous-type \textit{wx-hAT} gene (Fig. 4). Therefore, this set of three primers functioned as a molecular marker for selection of the 2.2-kb insertion mutant of the \textit{waxy} locus. This molecular marker was able to distinguish among maize lines with wild-type, homozgyous mutant-type, and heterozygous-type \textit{waxy} genes.
wx-D7 allele formed a separate branch with one flint maize (Jing2416) (Fig. 2). Ing 147 waxy maize carrying the (SXBN1, SXBN3, BN9 and 6013) and two flint maize (Ye478 and Zheng58) formed another branch. The remain-

parviglumis wild relatives (M106) and P1566891) to form an independent branch. Four waxy maize (SXBN1, SXBN3, BN9 and 6013) and two flint maize (Ye478 and Zheng58) formed another branch. The remain-
ing 147 waxy maize carrying the wx-D7 allele formed a separate branch with one flint maize (Jing2416) (Fig. 2). Waxy maize and flint maize were intermixed in each branch, which indicated that some waxy maize might be domesticated from flint maize. Moreover, waxy maize carrying the wx-D7 mutation was most abundant in the

Discussion

Because of the popularity of glutinous maize in China, waxy maize lines are frequently selected by maize breeders. An abundant of waxy maize germplasm has been obtained through decades of waxy maize breeding. The aim of our study was to reveal the genetic diversity and relationships among modern waxy maize inbred lines, as well as to identify new mutant alleles of the waxy gene in maize.

In previous studies, SSR markers have been widely used for genetic analyses of date palm22, barley23, angelia gigas24 and maize14,25. In this study, 40 core SSR markers with proven performance for maize genotype identification26 were used to genetically analyze 200 waxy maize accessions. The high average PIC value (0.7), high average number of alleles (11.45) and low average genetic similarity coefficient (0.31) are indicative of a high degree of genetic diversity among the waxy maize germplasm.

Compared with rice, maize shows a much higher level of sequence variation at the waxy locus8. To explore the genetic differentiation of the waxy gene in maize, the waxy gene sequences (3523 bp) from 167 waxy maize and 14 flint maize were compared and analyzed. The results show that the sequence diversity at the waxy locus in waxy maize is only 11.8% of that in flint maize (Table 1). This result is consistent with the findings of previous studies, i.e., that the DNA polymorphism of the waxy gene is much reduced in waxy maize10,13,14. Consistent the lower level of genetic diversity at the waxy locus in waxy maize than in non-glutinous maize, the neutral test revealed negative selection for the waxy gene in waxy maize, but not in non-glutinous maize (Table 1). Similar domestication selection in the waxy genomic region has also been detected in rice17–20, suggesting that the waxy gene in waxy crops has experienced a genetic bottleneck and strong artificial selection has acted on this locus during the improvement of waxy crops.

We conducted a phylogenetic analysis using the nucleotide sequences of the waxy gene in 167 waxy maize accessions, including five waxy maize accessions (Jing2, HN17, 1029, 80,453, 80,452) with the wx-D10 allele, one waxy maize accession with the wx-124 allele (SXBN4), eight waxy maize with the wx-hAT allele (SKN6, 80,482, SKN5, JN1, JN2, 16–585, YN1M and HN2), six waxy maize accessions with other waxy allele (SXBN1, SXBN3, BN9, 6013, 6003 and Zhongbang3M), and 147 waxy maize accessions with the wx-D7 allele. A previous study showed that the waxy maize with the wx-D10 genotype originated from the Yunnan-Guangxi region while that with the wx-D7 genotype originated from the Yangzi River region8. Consistent with these findings, waxy maize harboring wx-D10 formed a branch while waxy maize with the wx-D7 mutation was on a distinct independent branch in the tree (Fig. 2). Some waxy maize accessions had wx-124 and wx-hAT mutations and were located on a separate branch, suggesting that there may be additional origins of waxy maize.

The phylogenetic tree showed a clear genetic relationship among waxy maize, flint maize, waxy maize landraces and their wild relatives. Five wild maize relatives (parviglumis P1331783, P11331786 and P1384061, and mexicana P1566682 and P156685) formed a branch that was basal to all flint maize and waxy maize accessions, indicating that wild maize relatives (Mexicana and parviglumis) might be the ancestor of maize. Seven waxy maize landraces (CWM057, CWM056, CWM069, CWM052, CWM050) and five inbred lines (Jing2, HN17, 1029, 80,453, 80,452) clustered together first and then with one wild relative of maize (parviglumis M106) and two flint maize accessions (B73 and D9H) to form an independent branch. All waxy maize in this branch carried the wx-D10 allele. Next to this branch, eight maize inbred lines (SKN6, 80,482, SKN5, JN1, JN2, 16–585, YN1M and HN2) carrying the wx-124 allele clustered together first and then with one waxy maize inbred line (SXBN4) harboring the wx-124 allele, as well as six flint maize (Jing464, MC01, Jing724, HZS, C92 and Chang7–2) to form a separate branch. At the same time, far away from these two branches, two waxy maize (Zhonghang3M and 6003) were intermixed with three flint maize (Dan340, P178 and Q1319), and then clustered with two maize wild relatives (parviglumis P1331783 and mexicana P1566691) to form an independent branch. Four waxy maize (SXBN1, SXBN3, BN9 and 6013) and two flint maize (Ye478 and Zheng58) formed another branch. The remaining 147 waxy maize carrying the wx-D7 allele formed a separate branch with one flint maize (Jing2416) (Fig. 2). Waxy maize and flint maize were intermixed in each branch, which indicated that some waxy maize might be domesticated from flint maize. Moreover, waxy maize carrying the wx-D7 mutation was most abundant in the

Figure 4. Molecular marker detection of wx-hAT mutation by 1.5% agarose electrophoresis of PCR products. M, marker; lane 1–6, SKN5, JN1, HN2, YN1M, 16–585, 80,482, and these waxy maize lines carried the homozygous-type wx-hAT; lane 7–12, SKN5 × B73, JN1 × B73, HN2 × B73, YN1M × B73, 16–585 × B73, 80,482 × B73, and these lines carried the heterozygous-type waxy gene; lane 13–15, Jing2416, MC01, Jing 464, and these lines carried the wild-type waxy gene.
collected waxy maize inbred lines, suggesting that wx-D7 might be the main mutation type used in modern waxy maize breeding.

More than 50 waxy maize mutations had been described previously. Previous studies identified two deletion mutations including wx-D7 and wx-D10[6,11], together with three insertion mutations including wx-Cin4[11], wx-124[11] and wx-125[11] in Chinese waxy maize accessions. The wx-D7, wx-D10, and wx-124 mutations were also identified in our study, and wx-D7 was the main waxy maize allele type among the accessions we studied. Furthermore, we found a new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation wx-hAT. The hAT element is 3182-bp DNA transposon containing a 14-bp TIR flanked by an 8-bp TSD.

wx-hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions. A truncated hAT element was found in the insertion sequence, so we named this new insertion mutation allele, wx-hAT, in eight of 200 waxy maize accessions.
Phylogenetic tree reconstruction. A neighbor-joining phylogenetic tree based on the Kimura 2-parameter model was constructed with MEGAX64 software using waxy gene sequence data with 1000 bootstrap replications to assess tree reliability.

PCR marker development. The sequence of a newly identified insertion mutant of the waxy gene was used to develop a PCR molecular marker. The molecular marker was designed with Primer3 software. The primers were 1-4F (5′-AGAAGTTGACTGCTCGTCGCC-3′), 1-4R (5′-AGAATCGCTACGGTCTTGATAC-3′), and WaxyF2 (5′-AGTATGCCTCTACGGTCTGGCA-3′) (Supplementary Table S1). The PCR reaction conditions were as follows: 5 min at 94 °C, 34 cycles of 60 s at 94 °C, 60 s at 60 °C, 2 min at 72 °C, and final extension for 10 min at 72 °C. The PCR products were detected by 1.5% agarose gel electrophoresis.

Sequence analysis. Transposon prediction of the inserted sequence was performed using CENSOR (https://www.girinst.org/censor/index.php). The similarity between the inserted sequence and the B73 reference sequence was compared by conducting BLAST searches against the MaizeGDB database (https://www.maizegdb.org). Gene structure was drawn by Gene Structure Display Server (https://gsds.cbi.pku.edu.cn)

Received: 17 December 2019; Accepted: 7 September 2020
Published online: 28 September 2020

References
1. Singh, N., Inouchi, N. & Nishinari, K. Structural, thermal and viscoelastic characteristics of starches separated from normal, sugary and waxy maize. Food Hydrocoll. 20, 923–935 (2006).
2. Dong, L. et al. Supersweet and waxy: meeting the diverse demands for specialty maize by genome editing. Plant Biotechnol. J. 17, 1853–1855 (2019).
3. Wu, X. Y. et al. New waxy allele w-r Reina found in Chinese waxy maize. Genet. Resour. Crop Evol. 66, 885–895 (2019).
4. Liu, H. M. et al. Characterization of genome-wide variation in Four-row Waxy, a waxy maize landrace with a reduced kernel row phenotype. Front. Plant Sci. 7, 667. https://doi.org/10.3389/fpls.2016.00667 (2016).
5. Macdonald, F. D. & Preiss, J. Partial purification and characterization of granule-bound starch synthases from normal and waxy maize. J. Plant Physiol. 78, 849–852 (1985).
6. Shure, M., Wessler, S. & Fedoroff, N. Molecular identification and isolation of the waxy locus in maize. Cell 35, 225–233 (1983).
7. Hossain, F. et al. Molecular analysis of mutant granule-bound starch synthase-I (waxy I) gene in diverse waxy maize inbreds. 3 Biotech 9, 3. https://doi.org/10.1007/s13205-018-1530-6 (2019).
8. Huang, B. Q., Tian, M. L., Zhang, J. J. & Huang, Y. B. waxy locus and its mutant types in maize Zea mays L. J. Integr. Agr. 9, 1–10 (2010).
9. Bao, J. D. et al. Identification of gluminous maize landraces and inbred lines with altered transpiration of waxy gene. Mol. Breed. 30, 1707–1714 (2012).
10. Fan, L. J. et al. Post-domestication selection in the maize starch pathway. PLoS ONE 4, e7612. https://doi.org/10.1371/journal.pone.0007612 (2009).
11. Wu, X. Y. et al. Molecular characteristics of two new waxy mutations in China waxy maize. Mol. Breeding 37, 27. https://doi.org/10.1007/s11032-016-0612-6 (2017).
12. Hu, B. J. & Zhou, M. B. Active miniature inverted-repeat transposable elements transplants in plants: a review. Chin. J. Biotechnol. 34, 204–215 (2018).
13. Fan, L. J. et al. Molecular evidence for post-domestication selection in the Waxy gene of Chinese waxy maize. Mol. Breeding. 22, 329–338 (2008).
14. Zheng, H. J. et al. Genetic diversity and molecular evolution of Chinese waxy maize germplasm. PLoS ONE 8, e66066. https://doi.org/10.1371/journal.pone.0066066 (2013).
15. Teng, B., Zhang, Y., Wu, D. L. & Cong, X. H. Association between allelic variation at the Waxy locus and starch physicochemical properties using single-segment substitution lines in rice (Oryza sativa L.). Starch-Starke 65, 1069–1077 (2013).
16. Chen, M. H., Fjellstrom, R. G., Christensen, E. F. & Bergman, C. J. Development of three allele-specific codominant rice Waxy gene PCR markers suitable for marker-assisted selection of amylose content and paste viscosity. Mol. Breeding. 26, 513–523 (2010).
17. Liu, J., Rong, T. Z. & Li, W. C. Mutation loci and intragenic selection marker of the granule-bound starch synthase gene in waxy maize. Mol. Breed. 20, 93–102 (2007).
18. Kamalar, L. C. J., Dassanayaka, P. N., Perera, K. L. N. & Perera, S. A. C. N. SSR markers reveal the population structure of Sri Lankan yellow dwarf coconuts (Cocos nucifera L.). Euphytica 213, 36. https://doi.org/10.1007/s10681-016-1830-z (2017).
19. Luo, X. M., Wight, C. P., Zhou, Y. H. & Tinker, N. A. Characterization of chromosome-specific genomic DNA from hexaploid oat. Genome 53, 265–268 (2012).
20. Al Asafa, M. A., Duiwiyri, M. A., Qasem, J. R. & Al Abdallat, A. M. Characterization and assessment of genetic diversity of wild Date Palm (Phoenix dactylifera L.) in Jordan. Jordan J. Agr. Sci. 11, 75–94 (2015).
21. Wang, G. R. et al. Genetic diversity in hullless barley by retrotransposons-based markers and simple sequence repeat (SSR) markers. Acta Agr. Zhejiangensis 30, 357–365 (2018) (in Chinese).
22. Gill, J. S. et al. Development of genome-wide SSR markers from Angelica gigas Nakai using next generation sequencing. Genes 8, 238. https://doi.org/10.3390/genes8100238 (2017).
23. Watanabe, T. et al. Development of an SSR-based standard fingerprint database for corn variety authorized in China. Sci. Agr. Sinica 50, 1–14 (2017) (in Chinese).
24. Wang, F. G. et al. Construction of an SSR-based standard fingerprint database for corn variety authorized in China. Sci. Agr. Sinica 50, 1–14 (2017) (in Chinese).
27. Olsen, K. M. & Michael, D. P. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* **162**, 941–950 (2002).
28. Yamanaka, S., Nakamura, I., Watanabe, K. N. & Sato, Y. Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. *Theor. Appl. Genet.* **108**, 1200–1204 (2004).
29. Wei, X. *et al.* Domestication and geographic origin of *Oryza sativa* in China: insights from multilocus analysis of nucleotide variation of *O. sativa* and *O. rufipogon*. *Mol. Ecol.* **21**, 5073–5087 (2012).
30. Du, C. G., Hoffman, A., He, L. M., Caronna, J. & Dooner, H. K. The complete Ac/Ds transposon family of maize. *BMC Genom.* **12**, 588. https://doi.org/10.1186/1471-2164-12-588 (2011).
31. Luo, M. J. *et al.* Mapping of quantitative trait loci for seedling salt tolerance in maize. *Mol. Breed.* **39**, 64. https://doi.org/10.1007/s11032-019-0974-7 (2019).
32. Hara, T., Sato, K. & Ohya, M. MTRAP: pairwise sequence alignment algorithm by a new measure based on transition probability between two consecutive pairs of residues. *BMC Bioinform.* **11**, 235. https://doi.org/10.1186/1471-2105-11-235 (2010).
33. Hu, B. *et al.* GSDS 20: an upgraded gene feature visualization server. *Bioinformatics* **31**, 1296–1297 (2015).

**Acknowledgements**

This work was supported by the Beijing Municipal Natural Science Foundation (6204041), the Beijing Scholars Program (BSP041), the Science and Technology Planning Project of Beijing (D161100005716002), the Innovative Team Construction Project of BAAFS (JNKYT201603).

**Author contributions**

M. L., Y. Z., Y. Y. and J. Z. designed the experiment. B. L., Y. S., Y. Z., Y. S., M. K., C. L., Z. F., Y. F., L. X., and S. X. participated in data collection. M. L. and Y. Y. analyzed the data. M. L. wrote the manuscript. All authors revised and approved the final manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

*Supplementary information* is available for this paper at https://doi.org/10.1038/s41598-020-72764-3.

**Correspondence**

and requests for materials should be addressed to M.L., B.L. or J.Z.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020