Published online 11 January 2023

Nucleic Acids Research, 2023, Vol. 51, No. 2 ***595–609***
https://doi.org/10.1093/nar/gkac1195

A novel active transposon creates allelic variation through altered translation rate to influence protein abundance

Guo Chen[1,2,†], Ruilin Wang[1,†], Yizhe Jiang[1,†], Xiaoxiao Dong[1], Jing Xu[1], Qiang Xu[1], Qiuxin Kan[1], Zhixiang Luo[1], Nathan M. Springer[3] and Qing Li[1,4,*]

[1]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China, [2]Institute of Nuclear and Biological Technology, Xinjiang Academy of Agricultural Sciences/Xinjiang Key Laboratory of Crop Biotechnology/Improvement and Germplasm Innovation of Crop Resistance in Arid Desert Regions (Preparation), Urumqi 830091, China, [3]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108, USA and [4]Hubei Hongshan Laboratory, Wuhan 430070, China

Received July 13, 2022; Revised November 29, 2022; Editorial Decision November 29, 2022; Accepted December 03, 2022

ABSTRACT

Protein translation is tightly and precisely controlled by multiple mechanisms including upstream open reading frames (uORFs), but the origins of uORFs and their role in maize are largely unexplored. In this study, an active transposition event was identified during the propagation of maize inbred line B73. The transposon, which was named BTA for 'B73 active transposable element hAT', creates a novel dosage-dependent hypomorphic allele of the hexose transporter gene *ZmSWEET4c* through insertion within the coding sequence in the first exon, and results in reduced kernel size. The BTA insertion does not affect transcript abundance but reduces protein abundance of *ZmSWEET4c*, probably through the introduction of a uORF. Furthermore, the introduction of BTA sequence in the exon of other genes can regulate translation efficiency without affecting their mRNA levels. A transposon capture assay revealed 79 novel insertions for BTA and BTA-like elements. These insertion sites have typical euchromatin features, including low levels of DNA methylation and high levels of H3K27ac. A putative autonomous element that mobilizes BTA and BTA-like elements was identified. Together, our results suggest a transposon-based origin of uORFs and document a new role for transposable elements to influence protein abundance and phenotypic diversity by affecting the translation rate.

INTRODUCTION

Allelic variation is a major driver of heritable phenotypic variation. There are many types of allelic variation including, but not limited to, coding variation, regulatory variation and copy number variation. These types of variation contribute differently to phenotypic diversity. For example, a compilation of the functional allelic variations that result in crop quantitative trait loci suggested that regulatory variation and coding variation are the two major types of variants (1). Though many functional variants have been identified, there are relatively few cases of allelic variation that affect phenotypes through protein translation.

Protein translation can be tightly controlled by many mechanisms including microRNA, circular RNA (2), epigenetic modifications on mRNAs such as m^6A (3), mRNA secondary structure (4) and upstream open reading frames (uORFs). uORFs occur in 5′-untranslated regions (UTRs) of ∼40% of human mRNA and can affect the translation of the downstream main ORF (mORF) (5). Generally, uORFs regulate gene expression by reducing the translation of the mORF (6,7). Some studies have suggested that loss of uORF function is associated with human disease (8,9). Studies in plants have found that the uORFs are associated with disease resistance, nutrient absorption and other traits (7,10,11). Due to the importance and widespread occurrence of uORFs, many studies have investigated the utilization of uORFs to manipulate translation and phenotypes (12–14).

A recent study in maize suggested that allelic variation in uORFs among inbred genotypes can lead to different rates of protein translation and thus variable protein abundance, hinting at an important role for uORFs in creating maize phenotypic diversity (15). However, the mechanisms that create uORFs are not well characterized. A study in

*To whom correspondence should be addressed. Email: qingli@mail.hzau.edu.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

humans suggested that ∼10% of the ATGs of uORFs have a possible transposon origin (16). Transposons comprise ∼85% of the maize genome (17). There are many examples in maize of transposon insertions that regulate the expression of nearby genes to contribute to phenotypic variations in a range of traits, such as flowering time (18,19), tillering (20,21), disease resistance (22,23) and drought tolerance (24,25). Transposons can generate variability at all levels of gene expression. Transposon insertions within genes can disrupt the ORFs and result in altered expression levels, which were possibly caused by DNA methylation on transposable elements (TEs) or other mechanisms (26). Transposon insertions within promoters or 5′-UTRs of genes can provide novel regulatory sequences such as enhancers or promoters, which have been widely reported in human, flies and plants (26–28). In addition, transposons can repress translation through the production of small interfering RNA (siRNA) (29). Nevertheless, it is unclear whether transposons can affect the phenotype through the creation of uORFs to influence protein abundance.

The Ac/Ds transposon was first identified in maize by McClintock (30), and was a founding member of the hAT transposon family. The Ac element is an autonomous transposon that encodes a transposase containing four domains: the BED domain, DNA-binding domain, catalytic domain and an additional domain inserted into the catalytic domain (31). The Ds element, representing an internal deletion of Ac, is a non-autonomous transposon that requires the transposase encoded by Ac for transposition (32). Several active Ac-like elements have been identified in multiple species (31,33).

Previous research has shown that *ZmSWEET4c*, which encodes a hexose transporter, plays a key role in seed filling (34). This gene has been subject to convergent selection in both maize and rice (35). Loss of function of this gene leads to arrested kernel development in both maize and rice. In this study, we identified a novel hypomorphic allele (named '*smk*') of this gene with reduced protein translation efficiency and protein levels caused by a transposon (BTA) insertion. This allele has a dosage-dependent effect on kernel size. Our results provide evidence that transposon insertions within an exon can create uORFs, adding a new layer of regulation by transposons as 'controlling elements'. In addition, we found that BTA and its closely related transposons (BTA-like) are active and preferentially insert into regions with euchromatin features, and identified a novel Ac-like autonomous element. Our results identified a novel active transposon in maize and provided new insights into the potential regulatory role of transposons in protein translation.

MATERIALS AND METHODS

Plant materials

The *smk* allele was identified in B73, which is the inbred line used to generate the initial maize reference genome (17). Our B73 stock was originally obtained from Dr Peter Balint-Kurti (North Carolina State University) and had been self-pollinated for 3–4 generations prior to the identification of the *smk* mutant. It is known that identically named inbred lines might contain large amounts of uncharacterized genetic variations (36). To assess how this B73 stock is different from the B73 stock (Iowa State University) that was used for making the reference genome, we called single nucleotide polymorphism (SNPs) using RNA-seq data that were generated on kernels at 14 and 20 days after pollination (DAP) of the *smk* plant (see below). The HaplotypeCaller of the Genome Analysis Toolkit was used to call SNPs (37), and VariantFiltration was used to filter out low-quality SNPs which were defined as 'QualByDepth <2, FisherStrand >60.0, RMSMappingQuality <40.0, MappingQualityRankSumTest < −12.5 or ReadPosRankSumTest < −8.0'. We further require coverage of at least 10 reads for each SNP. Homozygous SNPs were defined as those for which >99% of the reads were from the non-reference allele. Heterozygous SNPs were defined as those for which >20% of the reads were from the minor allele. In total, 1060 SNPs were identified, comprising 276 heterozygous SNPs, 684 homozygous SNPs and 100 unassigned SNPs. About 48.7% (516/1060) of the SNPs were located within two small blocks on chromosome 1: 295 314 030–297 603 421 and chromosome 5: 212 831 212–217 338 586. This suggests that our B73 is indeed a B73 stock but with non-B73 haplotypes in some regions, which probably represent introgression of non-B73 haplotypes and residual heterozygosity.

Fine-mapping and gene cloning

For bulked segregant RNA sequencing (BSR-seq) to map the regions underlying *smk*, the homozygous *smk* mutant was crossed with Mo17 to generate an F_1 hybrid, which was then self-pollinated to produce a segregating F_2 population. Two pools of samples were collected from this F_2 population: one pool has kernels with mutant phenotype and the other pool has kernels with wild-type (WT) phenotype. Each pool contains at least 30 kernels with pericarp removed. Samples were frozen immediately using liquid nitrogen and were stored at –80°C until used. To fine-map the gene for *smk*, nine molecular markers were developed within the primary mapping region to survey a population of 971 mutant kernels for recombinants (Supplementary Table S1). These markers were designed by comparing the genomic sequence of B73 and Mo17 from the MaizeGDB (www.maizegdb.org).

Phenotypic characterization

For measuring kernel size and weight, kernels from the middle of at least 10 ears were collected for the homozygous *smk* mutant and the WT, respectively. Kernel size, including length, width and thickness, was measured using a Vernier caliper. Kernel length was defined as the distance from the tip of the kernel to the top. Kernel width and thickness are measured on the widest part of the kernel (38). For measuring agronomic traits, including plant height and ear height, at least 30 plants of the homozygous mutant and WT were used. At least 10 ears were used to measure ear-related traits, including ear diameter, ear length and cob diameter using a Vernier caliper. Ear row number was counted based on the kernel rows in the middle part of the ear.

To characterize kernel size across developmental stages, whole kernels, embryo and endosperm of the homozygous mutant and the WT were harvested every 2 days either from 1 to 24 DAP (for whole kernel) or from 12 to 24 DAP (for embryo and endosperm). The samples were fixed in 4% paraformaldehyde (Sigma, USA) overnight, and were photographed through a Leica MZFLIII microscope.

Cytological sections

For analysis of cytological sections, kernels at 10, 14 and 18 DAP were sampled for paraffin section preparation. The sections were fixed in 4% paraformaldehyde (Sigma, USA) overnight and were dehydrated in a graded ethanol series (30%, 50%, 70% ethanol). The sections were then embedded in Paraplast Plus, and 12 μm slices were cut using a Leica RM2265 microtome, followed by de-paraffinization as described previously (39). The sections were stained with 1% toluidine blue O.

Plasmid construction and plant transformation

To produce CRISPR/Cas9 [clustered regularly interspaced palindromic repeats (CRISPR)/CRISPR-associated protein 9) knockout lines, two single guide RNAs (sgRNAs) located at the third exon of *ZmSWEET4c* were designed using CRISPR-P (http://crispr.hzau.edu.cn/CRISPR2/). The fragment containing promoter-driven sgRNA (maizeU6.1-sgRNA1 and maizeU6.2-sgRNA2, Supplementary Table S1) was generated by polymerase chain reaction (PCR) and was inserted into the CRISPR/Cas9 vector. The construct was transformed into the maize inbred line B104 through *Agrobacterium*-mediated maize transformation at BoMeiXingAo (Beijing, China).

RNA isolation and qRT-PCR

Quantitative real-time PCR (qRT-PCR) was performed for the following purposes: (i) comparing the mRNA levels between the *smk* mutant and WT in 14 DAP kernels; (ii) Ribo-qRCR to compare mRNA levels bounded by different numbers of ribosomes; and (iii) determination of mRNA levels of genes that are transiently expressed in the protoplast for luciferase (LUC)-based experiments. For all these experiments, RNAs were isolated using TRIzol reagent (Invitrogen), and reverse transcribed with a cDNA Synthesis Mix (TransGen Biotech). qRT-PCR assays were performed with Hieff qPCR SYBR Green Master Mix (YEASEN) on a Bio-Rad CFX96 Touch Real-time PCR detection system. The maize actin gene was used as the internal control. Primers are listed in Supplementary Table S1. Relative expression was presented with the $2^{-\Delta\Delta Ct}$ method. For each qRT-PCR sample, three biological replicates with three technical replicates for each biological replicate were analyzed.

For Ribo-qPCR analysis, ribosome complexes were extracted from ∼500 mg of powder of kernels using Polysome Extraction Buffer (20 mM HEPES-KOH, pH 7.5, 100 mM KCl, 10 mM magnesium acetate, pH 7.0, 15 mM β-mercaptoethanol, 100 μg of cycloheximide/ml). The samples were loaded onto a sucrose density gradient (10–50%) and subjected to ultracentrifugation at 38 000 rpm (Beckman, SW40 rotor) for 3 h at 4°C. A gradient profiler (BioComp, http://www.biocompinstruments.com/index.php) with an EM-1 UV data acquisition system (Bio-Rad, http://www.bio-rad.com/) was employed for component analysis according to the manufacturer's instructions. RNA was extracted from each fraction embedded with either monosomes or polysomes using TRIzol reagent (Invitrogen). Reverse transcription and qRT-PCR assays were performed as described above.

Subcellular localization

The full-length coding sequences (CDSs) of the WT and *smk* allele were cloned into the RL2-GFP (green fluorescent protein) vector. Recombinant plasmid was extracted, and a total of ∼10 μg of plasmid was introduced into the maize protoplasts. The transfected protoplasts were cultured at room temperature for 16–18 h in the dark and the GFP signals were detected by an Olympus FluoView™ FV1200 confocal microscope (Olympus, Tokyo, Japan). Marker Red was used as reference maker for the plasma membrane.

RNA-seq and Ribo-seq

For RNA-seq and Ribo-seq, kernels at 14 and 20 DAP were collected from both the *smk* mutant and WT plants. Total RNA was isolated using TRIzol reagent (Invitrogen) and RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). RNA integrity was assessed using the RNA Nano 6000 Assay Kit on the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). A total of 1 μg of RNA per sample was used for RNA-seq library preparation using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations. Indexes were added to differentiate samples. Raw data of fastq format were first subject to quality control using FastQC (v.0.11.9). Reads with high quality were then mapped to the maize genome (B73 RefGen_v4, AGPv4) using HISAT2 (v.2.1.0) (40) with default parameters. Meanwhile, clean data of WT and *smk* were mapped to a database containing the cDNA sequences of the original WT allele and the *smk* allele using HISAT2 (v.2.1.0). Fragments per kilobase per million mapped reads (FPKM) were calculated using (number of fragments mapped to the WT or *smk*) × 10^9/(number of fragments mapped to the B73 genome × length of cDNA).

For Ribo-seq analysis, ribosome complexes were extracted using extraction buffer (50 mM Tris–HCl, pH 7.5, 200 mM KCl, 15 mM $MgCl_2$, 50 mg/ml cycloheximide, 15 mM β-mercaptoethanol, 1% Triton X-100, 10 U/ml DNase I). The Ribo-seq library preparation and sequencing was performed at Novogene (Beijing). The raw data were analyzed using the RiboToolkit (41) to check sequencing quality. The majority of the reads have a length between 25 and 29 nt, and the expected 3 nt periodicity for Ribo-seq data is good. Next, reads were mapped to both the B73 reference genome (version 4) and the respective cDNA sequences of either the WT or the *smk* allele. Reads per kilobase of transcript per million reads mapped (RPKM) was

calculated as (number of reads mapped to the WT or *smk*) × 10^9/(number of reads mapped to the B73 genome × length of cDNA).

To identify the putative ORF of each read, all the reads were first mapped to the CDS of the WT and the *smk* allele. Reads with a length of 25–32 nt were used for translation start codon analysis. The A-site of each aligned read was assigned based on the length of each read. The offset from the 5′ end of the alignment was 12 bp for reads with 25 nt, 13 bp for reads with 26 nt, 14 bp for reads with 27 nt, 15 bp for reads with 28 or 29 nt and 16 bp for reads with 30–32 nt. If the position of the A base in ATG1 of the WT allele is regarded as 1, the position of the A base in ATG2 would be 144 in the *smk* allele. Therefore, if the re-calibrated position of the reads is 1 + 3*n*, the reads were assigned to initiate translation from ATG1; if the re-calibrated position of the mapped reads is 144 + 3*n*, the reads were assigned to initiate translation from ATG2.

Luciferase assay

To produce the constructs used in the dual-luciferase assay, the various fragments, as described in Figures 4 and 5, were each cloned upstream of the LUC driven by a minimal sequence from the 35S promoter. The same vector also contains a strong 35S promoter driven Renilla LUC (RLUC) as an internal control. A total of 5 μg of recombinant plasmid was transformed into maize protoplasts using polyethylene glycol-mediated transformation. Measurement of LUC/RLUC activity was performed as described previously (42).

Western blot

A peptide containing 15 amino acids (NTKQIMEARRRKADQ) was synthesized to immunize white rabbits to produce antibody (QiWeiYiCheng Tech Co. Ltd, Beijing). In the western blot assay, the antibody was diluted using a ratio of 1:500. Total proteins were prepared either from 14 DAP kernels of the WT and *smk* mutant, from the leaves of B73 or from a transient assay in maize protoplast that express a fused protein between Flag and the WT or *smk* allele. In all cases, total proteins were extracted with extraction buffer [50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 1% sodium deoxycholate, 0.1% SDS, 1 mM EDTA, 2 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF), 10% glycerol and 1× protease inhibitor cocktail]. The extracted protein samples were mixed with 1× sodium dodecylsulfate (SDS) loading buffer and were separated by SDS–polyacrylamide gel elecrophoresis (PAGE). Separated protein samples were then transferred to a polyvinyldifluoridene (PVDF) membrane (0.2 μm; Bio-Rad). The membrane with the protein sample attached was incubated with primary and secondary antibodies. The signal was visualized using the clarity western ECL substrate kit (Bio-Rad) on GENEGNOME according to the manufacturer's instructions. The target protein was prepared using either the gene-specific antibody or anti-Flag (ABclonal, Cat. No. AE005, 1:3000). Actin used as a control was detected with anti-Actin antibody (ABclonal, Cat. No. AE009, 1:3000).

Identification and characterization of BTA-like elements

Sequences similar to BTA were identified using the Blast tool in MaizeGDB. The 156 bp BTA sequences containing both terminal inverted repeats (TIRs) were used as input to search the reference genome B73. The sequences that are similar to BTA at an E-value of 10^{-7} (bit score >60) were identified. We also required the target sequences to cover at least 90% of the query sequence. The aligned sequences along with flanking sequences on both sides were downloaded to identify putative target site duplications (TSDs) and TIRs. A similar method was used to annotate BTA-like transposons in Mo17 (43) and W22 genomes (44). To establish the orthologous relationship of the BTA-like elements in the three genomes (B73, Mo17 and W22), the 1 kb flanking sequences on each side were obtained for all the elements in each genome and were blasted against each other. The best target match that is on the same chromosome as the query and that is within 10 Mb of query was considered as the orthologous BTA-like element.

Transposon capture assay

The capture probes were designed at both ends of BTA-like elements. The length of the probes is 100 bp. Genomic DNA was extracted with CTAB (cetyltrimethylammonium bromide) and was used to prepare next-generation sequencing libraries using the Enzyme Plus Library Prep Kit. The libraries were captured using the designed probes with the TargetSeq One® Kit (Ada-block, for BGI) and were sequenced on the Illumina platform (Illumina, San Diego, CA, USA) with 150 bp paired-end reads mode. The probe synthesis, library preparation and sequencing were performed by iGenetech (Beijing, China).

To identify novel insertion sites of BTA-like transposons, raw reads were subject to quality control using FastQC (v.0.11.9). High-quality data were obtained after removing low-quality bases (Q <20) and adapters using TrimGalore/0.6.6 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The remaining reads were then used for identification of new insertion sites using a previous method with slight modification (45). Briefly, the reads were first aligned to the BTA-like transposons using Bowtie2/2.4.1 (parameters: -local -very sensitive -K 4) allowing at most four positions (46). The sequences mapped to BTA-like elements were then clipped and the remaining sequences were mapped to the B73 reference genome using Bowtie2/2.4.1 (parameters: -local -very sensitive). New insertion sites with at least three reads were identified and an 8 bp TSD was identified manually in IGV_2.8.13.

To verify the new insertion sites, primers were designed on the flanking sequence of insertion sites. Alternatively, one primer was designed on the transposon, and the other was designed on the flanking sequence of insertion sites (Supplementary Table S1). The PCR template was either B73 or the samples used for sequence capture.

To analyze which transposon was inserted in the novel insertion sites, reads mapped to the same position were extracted and were assembled using soapdenovo2 (version 2-r241) (47). For the insertion sites which soapdenovo2 could not assemble successfully, MEGA11 was used to align

the reads together with a reference with 300 bp upstream and downstream sequences of the insertion site (48). Next, the flanking sequences were deleted and the remaining sequences were regarded as transposon sequences, which were then aligned to the BTA-like transposons (detected in the B73 reference genome) using BLAST+ 2.9.0 (BLASTN -outfmt 6 -evaluate 1e-5). The transposon with the highest score is identified as the transposon on the insertion site.

For TSD bias analysis, the numbers of A, T, C and G at each position of the TSD were counted using run.sh (github.com/biozhp/motifStack_input) to derive a count matrix, which was then used to draw the figures using the motifstack package in R.

For epigenetic feature analysis, the chromatin immunoprecipitation (ChIP)-seq data from a previous study (49) were downloaded from the NCBI. The high-quality data were obtained after removing low-quality bases and adapters using TrimGalore/0.6.6 with default parameters. Next, they were mapped to the B73_v4 genome using Bowtie2/2.4.1 (parameters: -no mixed, -no discordant) (46). Only reads that are uniquely mapped were retained. Duplicate reads were removed using Picard-tools 1.119. The metaplots of reads around the insertion sites were made using Deeptools/3.5.0 (50).

Identification of an autonomous element for BTA transposition

To identify the possible autonomous element, the B73 genome was *de novo* annotated for transposons using the EDTA package (51), and the TIRs of the EDTA-annotated DTA (hAT) transposons were retrieved and compared with the TIR of BTA (which is 15 bp in length). A criterion of ≤3 bp mismatches was applied to find transposons that have TIRs highly similar to that of BTA. Next, we filtered out the DTAs that have a length of >3000 bp and have >10 RNA-seq reads in the RNA-seq libraries that are prepared from 14 DAP and 20 DAP WT and *smk* kernels. This leads to the identification of eight DTA elements. The entire sequences of these eight DTA elements were then compared with the entire sequences of BTA to identify putative autonomous elements that have similar internal sequences to BTA. To confirm that the putative autonomous element can be expressed, we fused the CDS in-frame with the gene encoding GFP, which was then transiently expressed in maize protoplasts. We also tagged the CDS with hemagglutinin (HA) and expressed it in *Escherichia coli*. Total protein was extracted and the target protein was detected using western blot.

RESULTS AND DISCUSSION

smk is a novel allele of *ZmSWEET4c*

A small kernel mutant (*smk*) was identified during the propagation of the inbred line B73 (Figure 1A). Compared with WT kernels, the mutant kernels show reduced kernel size and hundred kernel weight (Figure 1B; Supplementary Figure S1A), but can germinate normally (Supplementary Figure S1B). A significant difference was also observed for ear-related phenotypes (Supplementary Figure S1C). The plant architecture is largely similar between the WT and mutant except that the mutant plant is slightly taller than the WT (Supplementary Figure S1D, E). Paraffin section analysis revealed that the kernels of the mutant, including embryo and endosperm were normal but smaller than those of the WT across all developmental stages (Supplementary Figure S2).

The self-pollinated ear of the heterozygous plant showed a segregation of WT and mutant kernels at a ratio of 3:1 (Figure 1A, 308:81, $\chi^2 = 3.13$, $P > 0.05$), suggesting that the small kernel phenotype was caused by a single recessive mutation. To identify the gene underlying the *smk* mutant, the homozygous mutant was crossed with Mo17 to develop a large F_2 population. BSR-seq of WT and mutant kernels from a single ear was performed to map the underlying gene (52). This analysis mapped the mutation to a single chromosomal region on chromosome 5: base pairs 51 590 060–168 226 276. Using 971 mutant F_2 kernels, the gene was fine-mapped to a 14 Mb region between 118 Mb and 132 Mb (Figure 1C). Based on the B73 reference genome (version 4), there are 102 genes in this region. Most of these genes are either not expressed or constitutively expressed across different tissues, and three genes (*Zm00001d015860*, *Zm00001d015868* and *Zm00001d015912*) are specifically expressed in kernels where the most obvious phenotypic change was observed (Supplementary Figure S3A). Sequence analysis suggested no difference in *Zm00001d015860* and *Zm00001d015868*, but a 156 bp insertion was found in the first exon of *Zm00001d015912* in the *smk* mutant (Figure 1C). The 156 bp insertion is highly similar to a hAT transposon based on comparisons with transposons in Repbase (53). It has a perfect 8 bp TSD and an imperfect 15 bp TIR (Figure 1D). The TIR sequence does not share similarity with any reported transposons, and therefore we designated this novel transposon as BTA for 'B73 active TE hAT'. BTA insertion results in a premature stop codon of the *Zm00001d015912* gene (Figure 1D). *Zm00001d015912* encodes ZmSWEET4c, a glucose and fructose transporter that was previously shown to affect kernel development in maize and rice (34). Therefore, *Zm00001d015912* was considered as the candidate gene for *smk*.

One interesting observation is that, while the reference mutant allele (*zmsweet4c-umu1*) identified in a previous study showed a dramatic loss of endosperm and non-viable kernels (34), the *smk* allele identified in this study produces smaller but functional kernels. One possible explanation for this phenotypic discrepancy is the presence of genetic modifiers in the background since these alleles were generated in different backgrounds, with *smk* and *zmsweet4c-umu1* in B73 and W22 genetic backgrounds, respectively. To test for such a possibility, the homozygous *smk* mutants were crossed with the inbred line W22. This would allow us to test the effect of the *smk* allele in a W22-like genetic background. The resulting F_1 kernels are indistinguishable from WT W22 (Supplementary Figure S3B), suggesting that the W22 allele is dominant to the *smk* allele. These kernels were then planted and the plants were self-pollinated to generate F_2 kernels. The F_2 segregates large and small kernels at the expected 3:1 ratio (663:205, $\chi^2 = 0.88$, $P > 0.05$), fitting a single gene model, and no empty pericarp kernels were identified (Supplementary Figure S3B). Therefore, the phe-

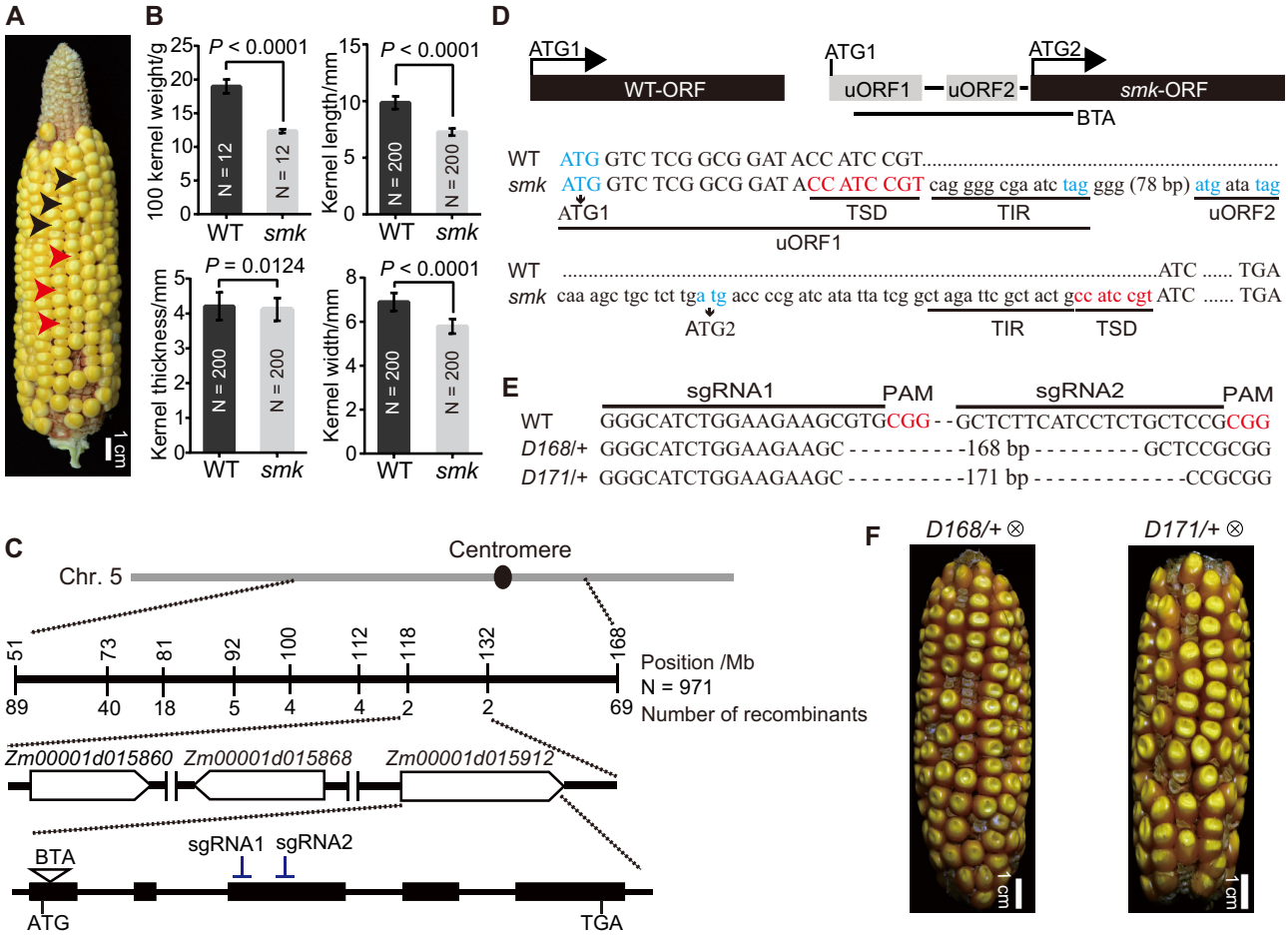

Figure 1. Phenotypic characterization of *smk* and identification of the underlying gene. (**A**) A representative ear from the self-pollination of a heterozygous plant. Black and red rows indicate WT and mutant kernels, respectively. (**B**) Comparison of hundred kernel weight and kernel size between the WT and mutant. The WT and *smk* kernels are from WT plants and homozygous *smk* plants, respectively. The *P*-value was from two-tailed Student's *t*-test. (**C**) Map-based cloning of the gene underlying *smk*. A hAT insertion (BTA) was identified in the first exon of the *Zm00001d015912* gene. (**D**) ORF maps of the WT and the *smk* alleles. The DNA sequences of BTA (lower case) and part of the flanking regions (upper case) are shown. The TSDs and TIRs of this insertion are underlined. ATG1 is the original start codon of the gene. ATG2 indicates a putative ATG that is provided by BTA. Also shown are two predicted uORFs, uORF1 and uORF2. (**E**) Genotypes of two CRISPR-based knockout lines. –: deletion. (**F**) Self-pollinated ears of heterozygous plants with the mutant allele corresponding to that in (E).

notypic differences among different alleles are not likely to be related to differences in genetic background.

The *smk* allele is hypomorphic and dosage dependent

To test the possibility that the *smk* allele is hypomorphic, two approaches were taken. In the first approach, we created loss-of-function alleles of *ZmSWEET4c* using CRISPR/Cas9 in the B104 background and introgressed the *smk* allele into this background. The CRISPR/Cas9 mutants were generated using two guide RNAs targeting the third exon of *ZmSWEET4c* (Figure 1C). A total of six independent events with deletions or insertions around the target sites were obtained, including two heterozygous lines (Figure 1E, F) and four homozygous lines (Supplementary Figure S3C, D). The self-pollinated ear of the two heterozygous plants was segregating for WT and empty pericarp kernels at a ratio of 3:1 (Figure 1E, F, 184:70, $\chi^2 = 0.89$, $P > 0.05$; 141:57, $\chi^2 = 1.52$, $P > 0.05$). The four homozygous mutant plants produced ears with empty pericarp kernels (Supplementary Figure S3C, D).

Homozygous *smk* (*smk*/*smk*) plants were crossed with a heterozygous CRISPR/Cas9 mutant that contained a 168 bp deletion (*D168*/+; Figures 1E, F and 2A). Two types of kernels (large and small) corresponding to *smk*/+ and *smk*/*D168* were observed in the resulting F_1 segregating at a ratio of 1:1 (85:88, $\chi^2 = 0.82$, $P > 0.05$, Figure 2B). Furthermore, both types of kernels were planted and the plants were self-pollinated (Figure 2C, D). The self-pollinated ears of the *smk*/+ plants were segregating for large and small kernels at a ratio fitting 3:1 (523:158, $\chi^2 = 1.18$, $P > 0.05$). Genotyping reveals that large kernels are of genotypes of *smk*/+ or +/+, and small kernels are of genotypes of *smk*/*smk* (Supplementary Figure S4A). In contrast, three types of kernels were segregating at a ratio of 1:2:1 in the self-pollinated ear of the *smk*/*D168* plant (58:100:59, $\chi^2 = 1.34$, $P > 0.05$, Figure 2D, E). Genotyping reveals that the small and empty kernels were of geno-

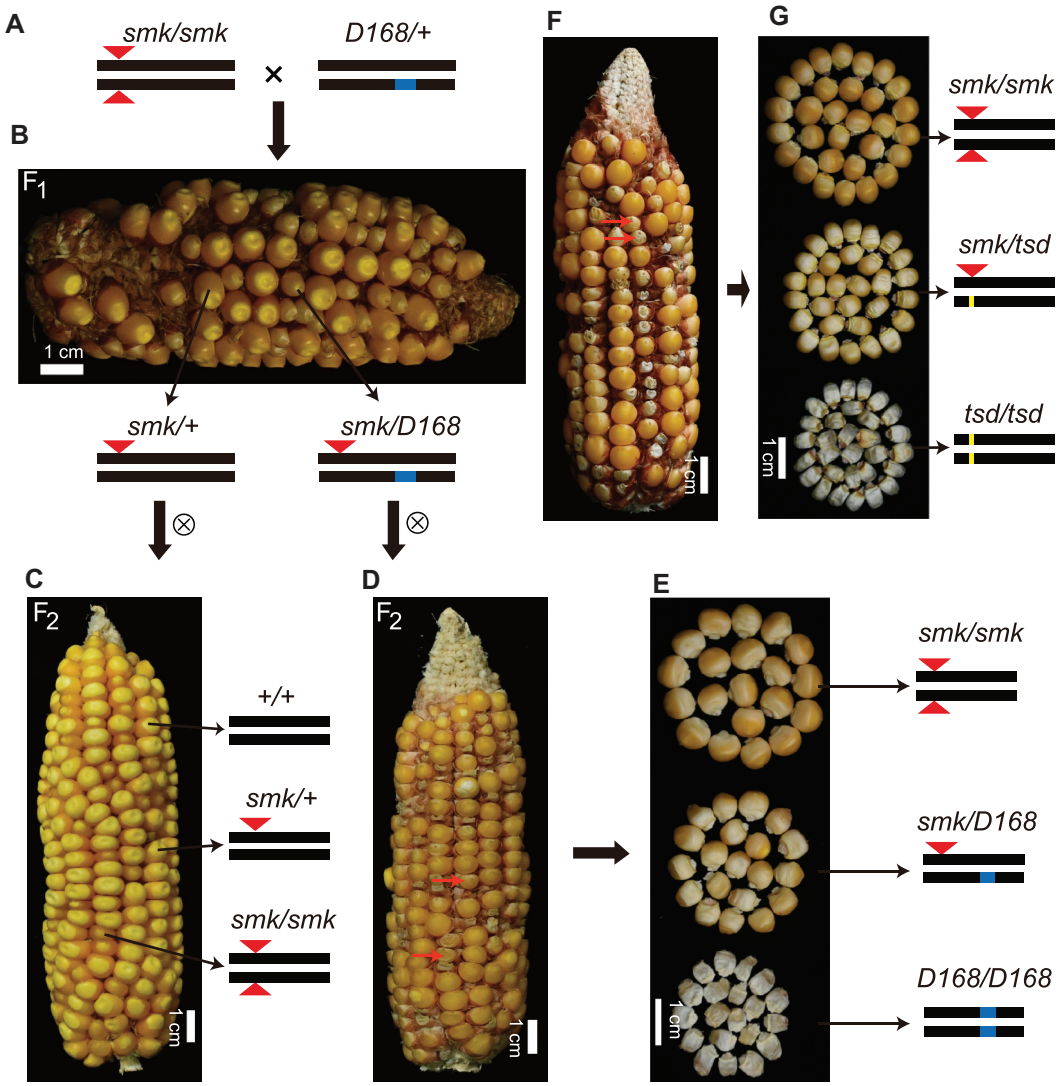

Figure 2. *smk* is an allele with partial function. (**A**) A schematic to show the crossing scheme between the *smk*/*smk* mutant and the heterozygous *D168*/+ plants. Red triangles indicate the BTA insertion; blue segments indicate the 168 bp deletion. (**B**) The F_1 ear from the cross in (A). (**C** and **D**) The self-pollinated ears of the F_1 plants and the genotypes of the kernels with different phenotypes. The red arrows in (D) indicate two heterozygous kernels (*smk*/*D168*). (**E**) The kernels from *smk*/*D168* selfing and their corresponding genotypes. (**F**) The ear from *smk*/*tsd* selfing. The red arrows indicate two heterozygous kernels (*smk*/*tsd*). (**G**) The kernels from *smk*/*tsd* selfing and their corresponding genotypes. +, WT allele; *smk*, allele with BTA insertion; *D168*, allele with 168 bp deletion; *tsd*, allele with BTA jumping away from the *smk* allele.

types *smk*/*smk* and *D168*/*D168*, respectively, while the kernels showing intermediate phenotypes were genotyped as *smk*/*D168*. This suggests that the *smk* allele is allelic to the *D168* mutation and is a hypomorphic allele rather than a null allele. Besides, it suggests that two copies of *smk* alleles can produce larger kernels than one copy, and a possible dosage effect of the *smk* allele. A similar analysis was performed between *smk* and another CRISPR/Cas9 allele (*D171*, 171 bp deletion) and the results are similar (Supplementary Figure S4B–D).

In the second approach, we screened for BTA excision from the *smk* allele. A segregating ear with small and empty kernels was identified during the propagation of the *smk* plants (Figure 2F). Genotyping of the empty kernels suggests that BTA is excised from the *smk* allele, leaving a 7 bp footprint that would lead to a frameshift of the original protein (Supplementary Figure S4E, named as the '*tsd*' allele for 'target site duplication'). This further supports that the *smk* allele is functional. Interestingly, an intermediate type of kernels was also observed, similar to *smk*/*D168* selfing (Figure 2E, G). The small, intermediate and empty kernels were segregating at a ratio of 1:2:1 (76:131:63, $\chi^2 = 1.49$, $P > 0.05$). Genotyping reveals that their genotypes were *smk*/*smk*, *smk*/*tsd* and *tsd*/*tsd*, respectively (Figure 2G). This suggests a dosage effect of the *smk* allele. Moreover, two types of kernels were observed for kernels with either the *smk*/*tsd* or *smk*/*D168* genotype (Figure 2D, F), reflecting a possible difference in the *smk* dosage of the triploid endosperm (two *smk* alleles versus one) and further supporting a dosage effect of this allele.

The *smk* allele can be translated into a protein similar to the WT allele

To provide insight into why the *smk* allele is hypomorphic, we analyzed the BTA insertion allele in more detail. While the protein originating at the original ATG of the gene (ATG1) terminates within BTA, a new ATG (ATG2) was found within BTA (Figure 1D). The putative protein starting from ATG2 is different from the WT protein only in the first 13 amino acids (Supplementary Figure S5A). To test whether such a putative protein exists, Ribo-seq which can detect translating mRNAs, was performed in kernels at 14 and 20 DAP (Supplementary Figure S5B). Reads covering the *smk* allele, including BTA itself, were detected at both 14 and 20 DAP, suggesting that *smk* can be translated.

Since the *smk* allele has two ATGs (ATG1 and ATG2 in Figure 1D) that are located within two different reading frames, we determined which ATG the Ribo-seq reads were derived from. We assign each read to the A-site (see the Materials and Methods; 54) which would be 3*n* (*n* is any integer) distance away from the start codon ATG. This analysis suggested that the Ribo-seq reads in *smk* were mainly from ATG2 (within BTA) translation both at 14 DAP (152/209) and at 20 DAP (163/202). In contrast, in the WT sample, the majority of reads were derived from ATG1 translation (1482/1816 reads at 14 DAP, and 246/299 reads at 20 DAP, Supplementary Figure S5C). These results provide further evidence that the *smk* allele can be translated. Moreover, subcellular localization analysis suggested that the protein encoded by *smk* is located in the plasma membrane which is similar to the protein encoded by the WT allele (Supplementary Figure S5D) and is consistent with the prediction that the products of *smk* and WT alleles have similar transmembrane protein structures (Supplementary Figure S6A). Together, these results support that the *smk* allele is translated and functional.

The *smk* allele has reduced translation efficiency

To provide insight into why the *smk* plants have kernels smaller than the WT plants, we investigated the mRNA expression level and splicing. Both RNA-seq and qRT-PCR suggest that the *smk* plant has RNA expression levels that are not significantly different from the WT allele (Supplementary Figure S6B, C). Moreover, no splicing differences in RNA were observed between the WT and *smk* (Supplementary Figure S6D). This suggests that the small kernel phenotype is not caused by reduced mRNA levels or aberrant mRNA splicing. Another possibility is that protein levels were reduced in the *smk* plant. To investigate this possibility, two vectors with the UBI promoter driving the expression of a fusion protein of either the WT or the *smk* allele (including the 5′-UTR and full CDS but not including the TGA stop codon) and the 3×Flag tag were generated and transiently expressed in maize protoplasts (Figure 3A). As expected, the *smk* allele produces less protein than the WT allele (the protein mol. wt is 31.64 kDa and 32.41 kDa for the WT-3×Flag and SMK-3×Flag respectively; Figure 3A). To investigate whether the *smk* allele has reduced protein abundance compared with the WT *in vivo*, an antibody that can detect both the WT and SMK protein was generated (Supplementary Figures S5A and S6E). This antibody can detect the endogenous WT and SMK protein (the protein mol. wt is 28.28 kDa and 29.05 kDa, respectively) in 14 DAP maize kernels, while no protein was detected in the leaves where the mRNA is not expressed. Compared with the WT, the SMK protein level is significantly reduced (Figure 3B). Taken together, these results indicated that the *smk* allele produces less protein compared with the WT allele.

To study the reason for reduced protein in *smk* plants, Ribo-seq and polysome profiling coupled with qRT-PCR were used to investigate translation efficiency. Ribo-seq suggested that there were fewer Ribo-seq reads in *smk* compared with the WT (Supplementary Figure S5B), raising the possibility of reduced translation. Therefore, translation efficiency was quantified by dividing the number of Ribo-seq reads by the number of RNA-seq reads. Indeed, translation efficiency was reduced in *smk*, especially at 14 DAP (Figure 3C). To further support this possibility, polysome profiling qRT-PCR was performed. Ribosomes were separated in sucrose gradients by continuous measurement of RNA absorbance ($A_{260\,nm}$) and were isolated at different elution times (fractions 4–11 in Figure 3D). The mRNAs from each fraction were isolated to perform qRT-PCR with a primer that can amplify both the WT and *smk* allele. There is significantly more mRNA that is bound by multiple ribosomes in the WT than in *smk* (fractions 10–11 in Figure 3E). Together, the Ribo-seq and polysome profiling qRT-PCR results suggested that the translational efficiency of the *smk* allele was reduced.

The reduced translation efficiency of the *smk* allele was caused by BTA insertion

To explore why translation efficiency of the *smk* allele was reduced, a dual-luciferase reporter system was exploited. Different alleles of the target gene were fused in-frame to the luciferase gene and cloned downstream of the 35S minipromoter (Figure 4A). The Renilla protein (RLUC) that was on the same plasmid with the LUC protein was used as an internal control for transformation efficiency, and the plasmid was transformed into maize protoplasts. In an initial experiment, the WT and the *smk* alleles were cloned into this vector, and were used to confirm the different translation efficiency of these two alleles. Compared with the WT allele, the LUC/RLUC level of the *smk* allele decreases ∼24.5-fold (Figure 4B). This decrease is not due to differences in sequence length between the two alleles or differences in mRNA levels (Supplementary Figure S7). Next, various mutations and deletions were generated on the BTA insertions of the *smk* allele to assess how these mutations affect translation efficiency. As shown in Figure 1D, the two start codons ATG1 and ATG2 were in different reading frames, and ATG1 would initiate a uORF (named uORF1). Careful examination of the BTA insertion found another small uORF with only three codons (uORF2) before ATG2 (Figure 1D). To test whether these two uORFs affect translation, the start codon of each uORF was mutated to TTG. While mutation of uORF1 leads to a slight increase in translation efficiency, no change in translation efficiency was found for the mutation of uORF2 (Figure 4C). Simultaneous mutation of both start codons had a similar effect to that of uORF1 (Figure 4C). These findings suggest that

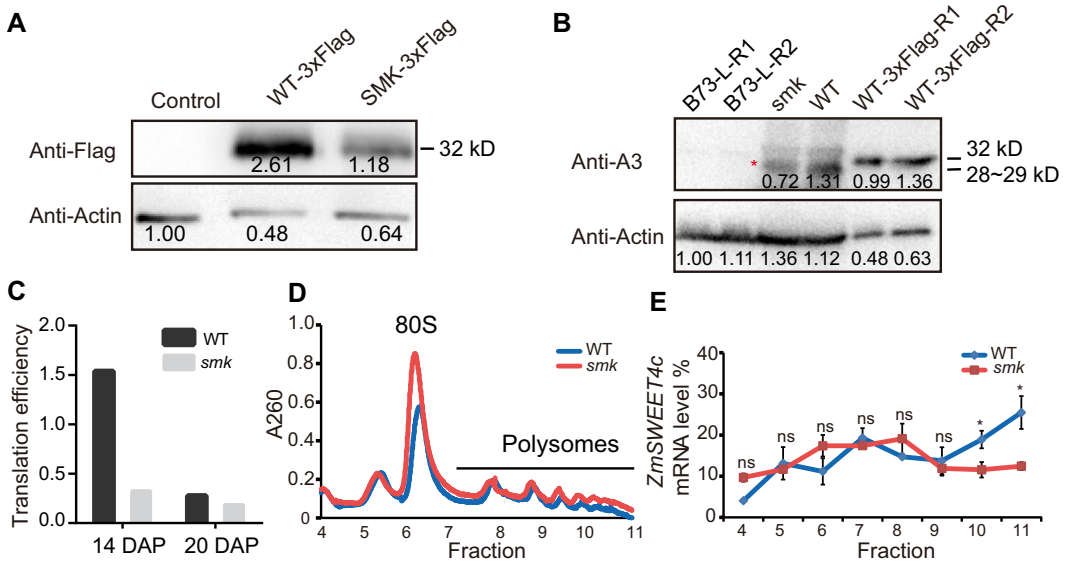

Figure 3. The *smk* allele shows reduced translation efficiency. (**A**) The *smk* allele show reduced protein levels based on transient expression in maize protoplasts. Control: empty vector. (**B**) Western blot showing that the *smk* mutant has reduced protein levels compared with the WT in 14 DAP kernels. The red asterisk indicates the position of the target protein. B73-L, the leaf of the B73 inbred line, was used as a negative control. The WT-3×Flag, which is WT protein fused with a 3×Flag and expressed in maize protoplasts, was used as a positive control. Note that the WT-3×Flag protein was expected to be larger than the endogenous protein because of the presence of the 3×Flag tag. R1 and R2 represent two replicates. (**C**) Comparison of translation efficiency between the WT and *smk* based on Ribo-seq. (**D**) Ribosomes isolated from the kernels of the WT and *smk* by sucrose density gradient. A260: the absorbance at 260 nm. (**E**) The relative mRNA levels of *ZmSWEET4c* by polysome profiling qRT-PCR in WT and *smk* kernels. The error bar represents the standard deviation of three replicates. $*P < 0.05$; ns, not significant. The *P*-value was based on two-tailed Student's *t*-test.

uORF1 can affect translation of the downstream main reading frame.

Though the translation efficiency was increased by mutating uORF1, it is still significantly lower than that of the WT allele, suggesting that there are other factors affecting translation efficiency. To further explore the mechanism of BTA inhibition on translation efficiency, we analyzed the secondary structure of BTA and its flanking sequences (Figure 4D); a complex hairpin–loop structure was predicted. Therefore, we tested whether deletion of partial sequences would affect translation (Figure 4E). Three constructs with varying sizes of the 5′ end sequences removed were generated. JD1-39 has a deletion of the first 39 bp starting from ATG1, JD1-120 has a deletion of the first 120 bp and JD1-129 has a deletion of the first 129 bp. Compared with the *smk* allele, all three newly created alleles had higher translation efficiency, with the alleles with longer sequences deleted having higher translation efficiency. In addition, protein levels were also significantly reduced when BTA was inserted before the annotated ATG of the original *ZmSWEET4c* gene, or when BTA was replaced with an unrelated sequence (Supplementary Figure S7). These results suggest that secondary structure introduced by BTA may also influence the translation efficiency. Furthermore, we mutated ATG2 on the basis of JD1-39 and JD1-120 to test whether ATG2 is the start codon that is employed for translation (Figure 4F). Compared with the corresponding allele with no mutation, the mutated allele showed reduced translation efficiency, providing further evidence that ATG2 could initiate the translation of a protein. These results suggested that the precise insertion of BTA within the *ZmSWEET4c* gene not only results in a similar protein to the WT allele, but also creates a uORF and possible secondary structures that can regulate the protein translation rate without affecting mRNA levels to affect phenotypic diversity (Supplementary Figures S7 and S8). In addition, it is interesting to note that the different mutations/deletions could lead to a series of (from 0.1 to 0.7) translation efficiencies relative to the original WT allele. This provides the possibility to change the translation efficiency by manipulating BTA, and also provides an alternative route to create quantitative variations, except the widespread use of promoter editing (55).

BTA can modulate translation efficiency of other genes

To explore whether BTA can affect translation efficiency of other genes, four maize genes were chosen because of their well-known function, namely *tb1*, *fea2*, *ZmCCT10* and *ZmDREB1D*. The *tb1* gene is a domestication gene regulating tiller branching (21). The *fea2* gene encodes a protein whose activity is negatively correlated with maize kernel row number (56). The *ZmCCT10* gene regulates flowering time and resistance to Gibberella stalk rot (23,57). The *ZmDREB1D* (*Zm00001d002618*) gene encodes a protein homologous to CBF3 in *Arabidopsis* that is involved in chilling tolerance (58). The coding region of these four genes without BTA (–BTA) as well as that with BTA insertion (+BTA) was cloned into the same dual-reporter vector as that in Figure 4 (Figure 5A). The two constructs for each gene were transformed into maize mesophyll protoplasts. Interestingly, all four genes with a BTA insertion showed reduced LUC levels compared with the WT allele (Figure 5B). Quantification of the mRNA levels suggests that the reduced LUC levels were due to reduced translation but not mRNA levels

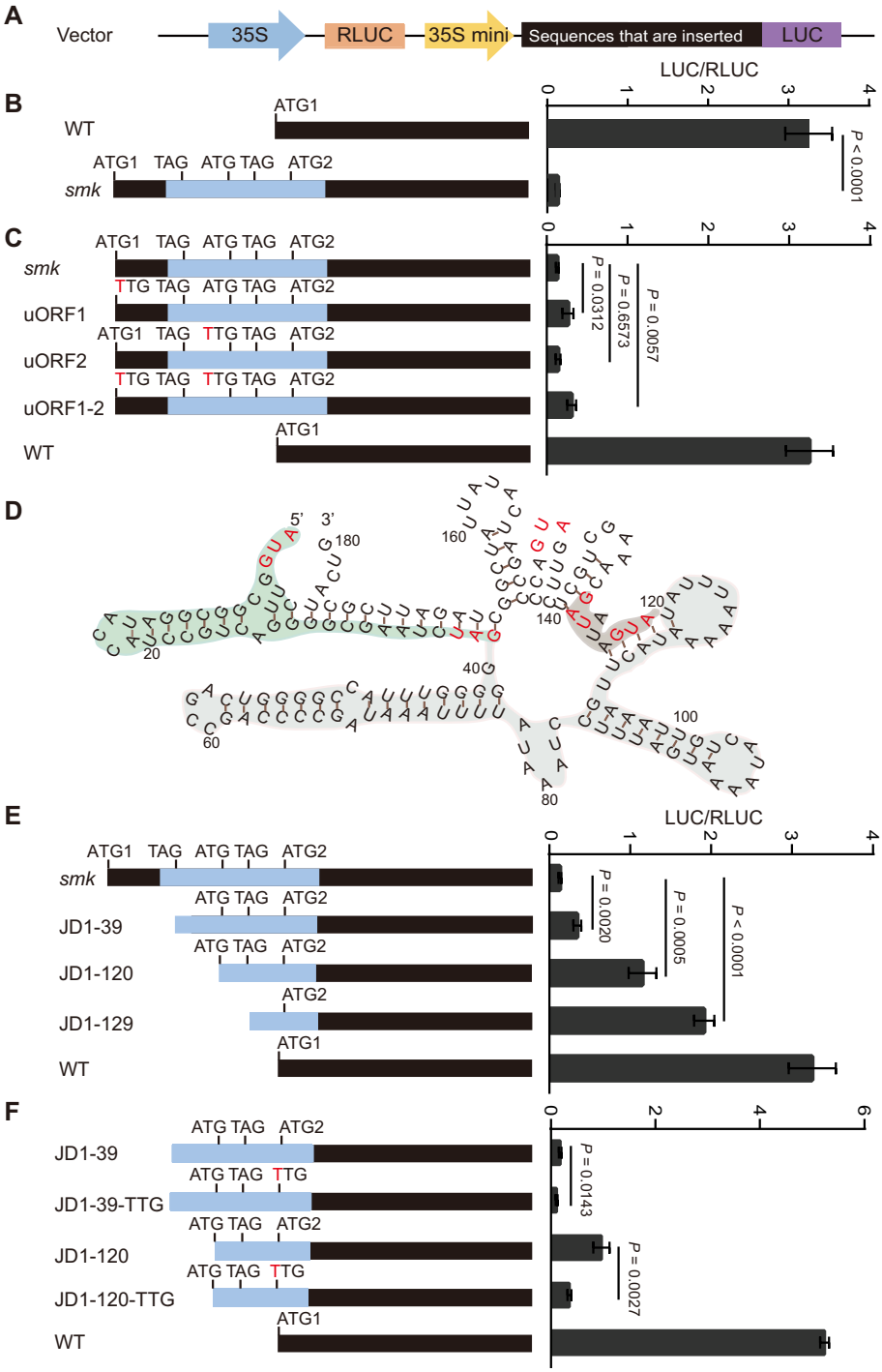

Figure 4. BTA regulates translation of *smk*. (**A**) A schematic plot to show the vector made for the LUC experiment. (**B**) Comparison of LUC activity between the WT and *smk* allele. The blue segment indicates BTA. Note that the *smk* allele shows reduced translation due to BTA insertion. (**C**) Effect of the two uORFs (Figure 1D) on the translation of the *smk* allele. (**D**) The predicted RNA secondary structure of BTA and its flanking sequences. The structure was predicted using RNAfold (rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) with default parameters. (**E**) The effect of various deletions on the translation of the *smk* allele. (**F**) Mutation of ATG2 (Figure 1D) leads to reduced translation. In (B), (C), (E) and (F), values and bars represent the mean and standard deviation of three biological replicates, respectively. Note: the WT samples in (B), (C) and (E) are same, which is different from that in (F). The *P*-value was determined by two-tailed Student's *t*-test.

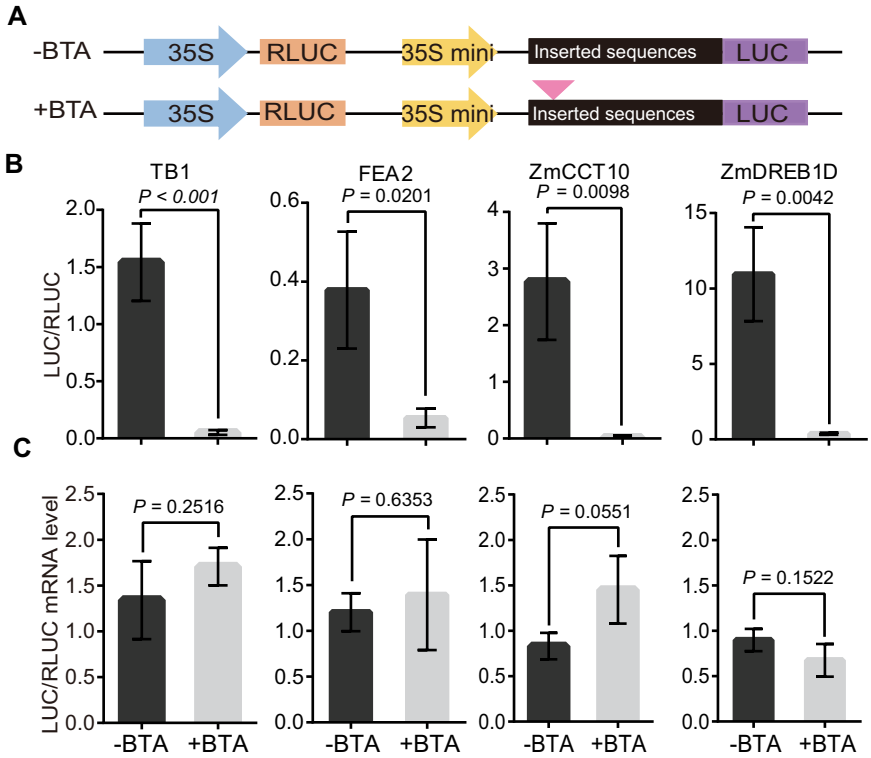

Figure 5. BTA can regulate translation efficiency of other genes. (**A**) A schematic to show the construction of vectors. (B and C) Comparison of translation efficiency (**B**) and mRNA levels (**C**) of the allele with or without BTA insertion. The *P*-value was determined by two-tailed Student's *t*-test.

(Figure 5C). These data indicate that BTA can be used to manipulate translation of other genes. With the development of CRISPR/Cas9 technology, it is possible to create *in vivo* precise insertion of BTA within genes to control their translation rate to produce the desired phenotypes.

Genomic and epigenomic features associated with BTA and BTA-like transposition

To further characterize the BTA transposon family, we first determined the copy number of BTA in the B73 reference genome. A total of 123 BTA-like elements were identified using either B73 version4 or version5 references, which utilized different B73 individuals for sequencing (Supplementary Table S2). A similar search in the Mo17 and W22 genome identifies 112 and 130 BTA-like elements, respectively (Supplementary Table S2). These include 74 elements that can be identified in all three genomes (Supplementary Table S2).

We hypothesized that BTA has been active in our B73 stocks. Therefore, a transposon capture assay was developed to identify novel insertions using seven B73 samples that are related to the *smk* mutant (Supplementary Figure S9). To maximize the identification of novel insertion sites, DNA from multiple individuals for each sample were mixed. This method is quite sensitive as 121 of the 123 BTA-like elements in their original positions can be identified in at least one of the seven samples. Moreover, the detection sensitivity is at least 10% as the BTA insertion in the *smk* allele can be identified in a sample which has 10% of individuals that contain the *smk* insertion (Supplementary Table S3), suggesting that the capture assay is highly efficient.

Using the transposon capture assay, a total of 79 non-redundant novel insertion sites distributed over the 10 maize chromosomes were identified from the seven samples, with each sample having 1–21 novel insertions (Figure 6A; Supplementary Table S3). A total of 49 insertion sites were randomly chosen for PCR-based validation, and the validation rate is ∼69% (34/49) (Supplementary Figure S9D). Interestingly, these novel insertion sites were due to transposition of six distinct BTA-like elements, including BTA itself which contributes 36 insertions (TE38 in Figure 6B).

The genomic distribution pattern of the 79 novel insertion sites is quite similar to the full set of 123 previously existing insertions. Interestingly, there is a high enrichment within and around gene bodies for novel insertions (Figure 6C). It is likely that these types of insertions may be reduced in the previously existing insertion due to possible deleterious effects on gene expression. Examination of the TSDs suggests a sequence bias at the eighth position for C and G, which is similar to that of the historical transposition events (Figure 6D). Interestingly, the insertion sites display unique epigenomic features (Figure 6E). Specifically, the levels of H3K27ac are higher at the insertion sites than at the flanking regions. In contrast, DNA methylation levels in all three contexts (CG/CHG/CHH) are lower at the insertion sites (Figure 6E). These results suggest that the BTA-like transposons are preferentially inserted into euchromatic regions with specific chromatin features.

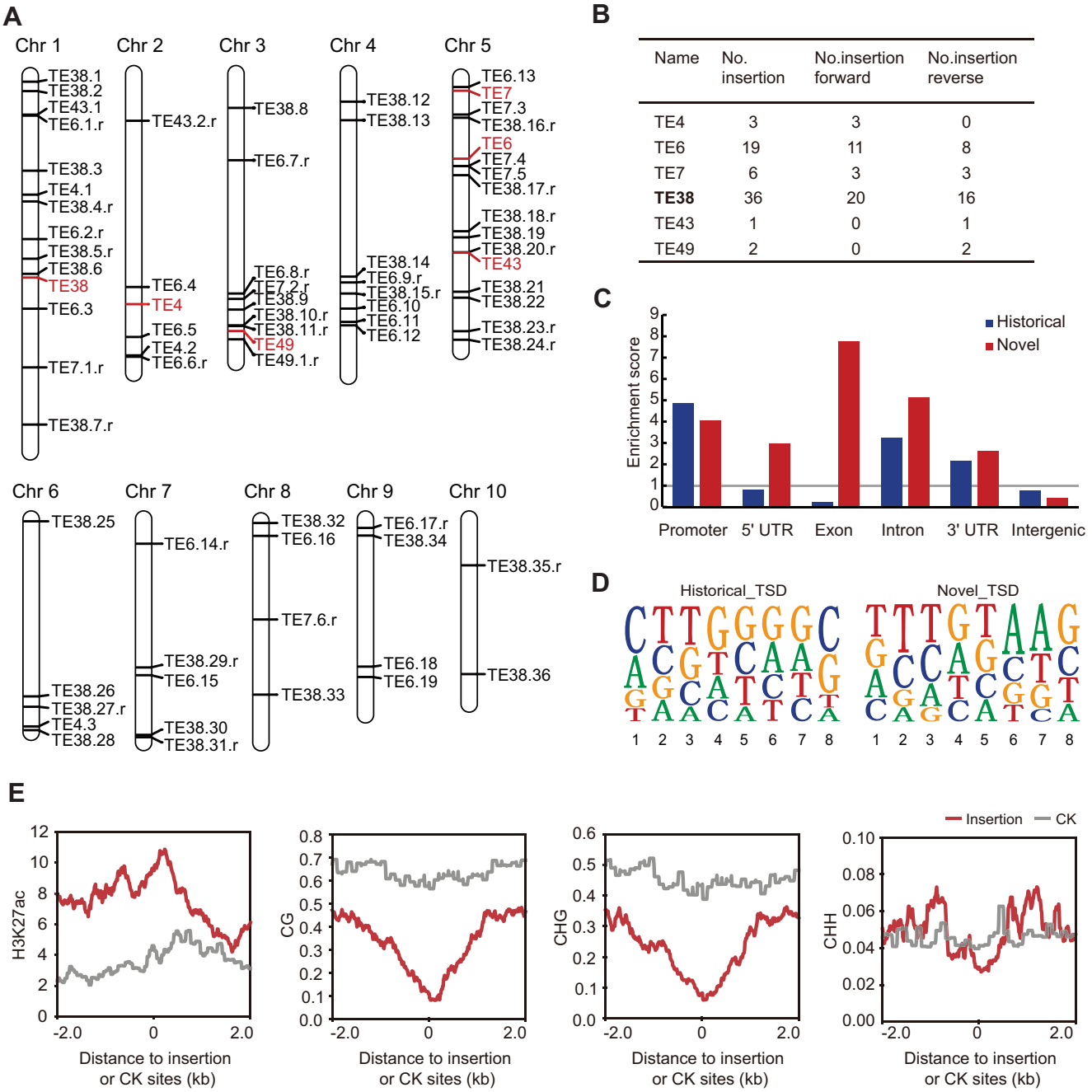

Figure 6. The features associated with BTA and BTA-like transpositions. (**A**) Distribution of the novel insertions as well as their parental origins (highlighted in red) in the B73 genome. Novel transposition of the same transposon was named using sequential numbers after the name of the original transposon. For example, TE38.15.r represents the 15th insertion for TE38 and this insertion is reverse complementary to TE38. Note that the insertions that are not verified by PCR were not included. (**B**) Summary of novel insertions detected for BTA (TE38) and five other BTA-like elements. (**C**) Enrichment of novel insertions as well as historical insertions over different features. (**D**) Sequence features of TSD for novel insertions and historical insertions. (**E**) Epigenetic features of the novel insertion sites in the B73 genome. Gray lines represent control genomic regions that are randomly selected.

Identification of a putative autonomous element for BTA

One possible autonomous element for BTA transposition was identified in the B73 genome (see the Materials and Methods). This element is located on chromosome 6: 106 095 770–106 098 809 (version4, named hAT2908, 3040 bp), and causes an 8 bp duplication at the target site (5′-AATAGGAT-3′). hAT2908 ends in 15 bp imperfect TIRs which show 86.67% and 73.33% similarity to those of BTA for the 5′ and 3′ TIRs, respectively (Figure 7A). The internal sequences of hAT2908, especially at the 5′ end, are also highly similar to that of BTA, suggesting that BTA is a possible deletion derivative of hAT2908 (Figure 7A). Besides TSD and TIRs, hAT2908 contains two types of repetitive sequences (5′-GACCCC-3′ and 5′-TATTT-3′) or the reverse complement within the subterminal 300 bp at either end

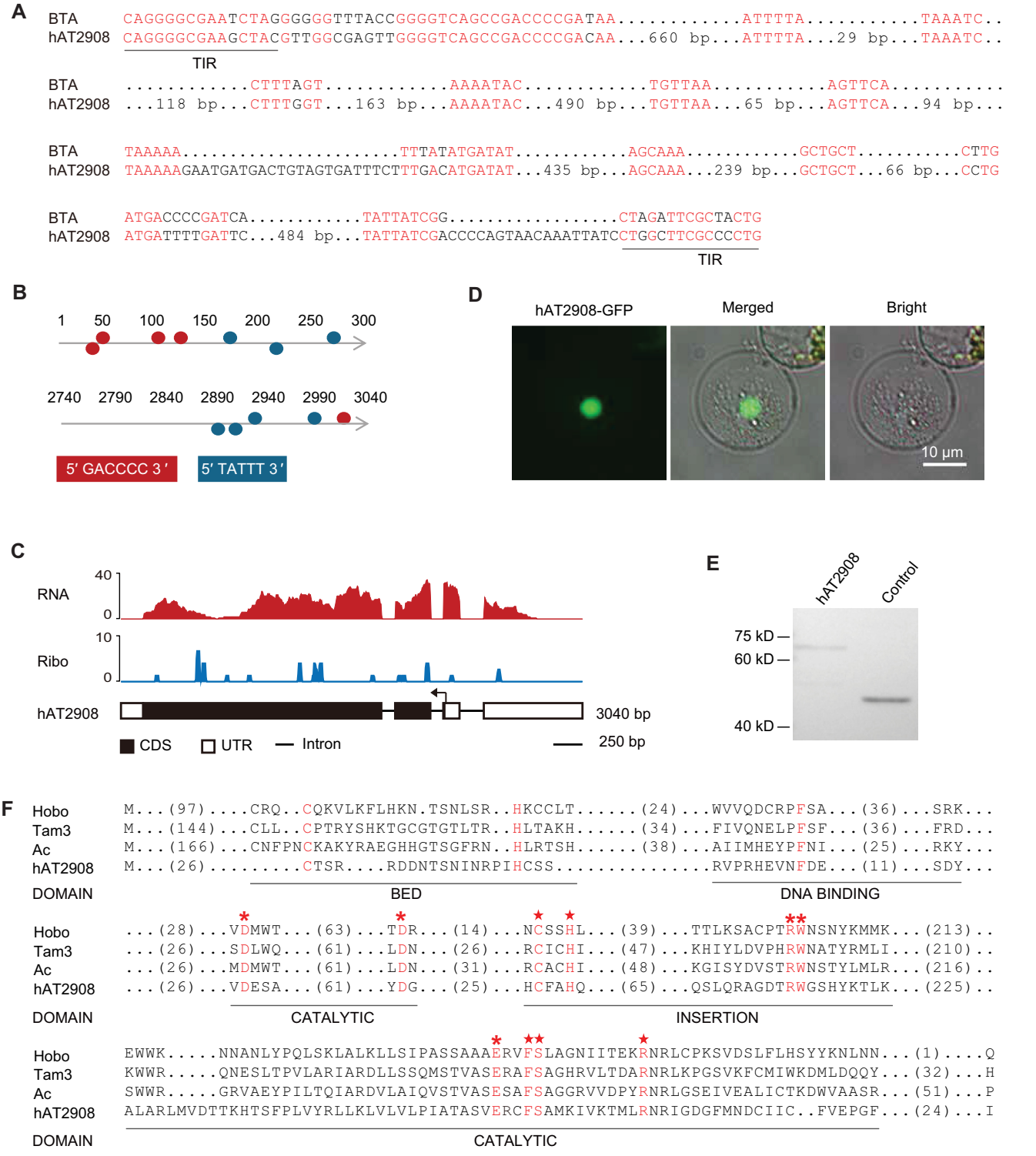

Figure 7. Characterization of the putative autonomous element. (**A**) Sequence comparison between BTA and hAT2908. The TIRs are labeled. The numbers represent the number of nucleotides that cannot be aligned. (**B**) A diagram to show subterminal direct repeats in hAT2908. The two types of repeat are shown by two different colors. The repeats on the top or bottom strand of hAT2908 are depicted by circles above or below the transposon, respectively. (**C**) An IGV view to show that hAT2908 can be transcribed and translated. (**D**) Transient expression of the hAT2908 CDS fused in-frame with GFP in maize protoplast. (**E**) Western blot of hAT2908 that was tagged with HA. Control is an unrelated protein tagged with HA. (**F**) Comparison of the putative domains of hAT2908 with three known autonomous hAT transposases. The amino acids that are conserved based on previous studies are labeled in red and/or marked by asterisks and stars.

(Figure 7B). hAT2908 can be transcribed with clear exon–intron structure in our RNA-seq data, and has Ribo-seq reads on the exons (Figure 7C), suggesting that the RNAs can be translated. Furthermore, the CDS of hAT2908 can be expressed in maize protoplast (Figure 7D). The expected protein can also be detected using western blot when expressed in *E. coli* (Figure 7E). The hAT2908 transcript has four exons and encodes a putative protein with 602 amino acids. Sequence alignment between the protein encoded by hAT2908 and the three founding autonomous members of the hAT superfamily (Hobo from *Drosophila*, Ac from maize and Tam3 from snapdragon) showed that the known domains of the hAT transposase are highly conserved (Figure 7F). This includes a BED domain, a DNA-binding domain, a catalytic domain and an additional domain inserted into the catalytic domain (31). Notably, the five key amino acids essential for the activity of hAT transposases (59; marked by asterisks in Figure 7F) were all found unchanged in hAT2908 protein. In addition, another five highly conserved amino residues based on alignment of 44 putative 'active' transposases in plants (60; marked by stars in Figure 7F) were also unchanged. Together, these results suggest that hAT2908 is a potential autonomous element that can mobilize BTA and BTA-like elements.

DATA AVAILABILITY

Sequence data from this article can be found in the GenBank data libraries under project number PRJNA821340.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Professor Zhibing Lai for providing the B73 stock used in this study, Professor Zhipeng Zhou for helpful discussions, and Professor Ning Yang for providing transposon annotation of the B73 genome. Data analyses were performed on the high-performance computing platform at the National Key Laboratory of Crop Genetic Improvement in Huazhong Agricultural University.

Author contributions: Q.L. designed this study. G.C., R.W., Y.J., X.D., Q.K. and Z.L. conducted the experiments. R.W., Y.J., J.X. and Q.X. analyzed the data. Q.L., G.C. and N.M.S. drafted and revised the manuscript.

FUNDING

This work was supported by National Natural Science Foundation of China [32072076]; the Fundamental Research Funds for the Central Universities [2662022SKPY002]; and the 111 Project Crop Genomics and Molecular Breeding [B20051]. N.M.S. is supported by a grant from the National Science Foundation [IOS-1934384].

Conflict of interest statement. None declared.

REFERENCES

1. Liang,Y., Liu,H.J., Yan,J. and Tian,F. (2021) Natural variation in crops, realized understanding, continuing promise. *Annu. Rev. Plant Biol.*, **72**, 357–385.
2. Liu,C.X. and Chen,L.L. (2022) Circular RNAs, characterization, cellular roles, and applications. *Cell*, **185**, 2016–2034.
3. Zhou,L., Gao,G., Tang,R., Wang,W., Wang,Y., Tian,S. and Qin,G. (2022) m6A-mediated regulation of crop development and stress responses. *Plant Biotechnol. J.*, **20**, 1447–1455.
4. Zhang,Y. and Gross,C.A. (2021) Cold shock response in bacteria. *Annu. Rev. Genet.*, **55**, 377–400.
5. Jayaram,D.R., Frost,S., Argov,C., Liju,V.B., Anto,N.P., Muraleedharan,A., Ben-Ari,A., Sinay,R., Smoly,I., Novoplansky,O. *et al.* (2021) Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of pkcs. *Proc. Natl Acad. Sci. USA*, **118**, e2018899118.
6. Wu,H.W., Fajiculay,E., Wu,J.F., Yan,C.S., Hsu,C.P. and Wu,S.H. (2022) Noise reduction by upstream open reading frames. *Nat. Plants*, **8**, 474–480.
7. Zhang,T., Wu,A., Yue,Y. and Zhao,Y. (2020) uORFs: important cis-regulatory elements in plants. *Int. J. Mol. Sci.*, **21**, 6238.
8. Lee,D.S.M., Park,J., Kromer,A., Baras,A., Rader,D.J., Ritchie,M.D., Ghanem,L.R. and Barash,Y. (2021) Disrupting upstream translation in mRNAs is associated with human disease. *Nat. Commun.*, **12**, 1515.
9. Chen,H.H. and Tarn,W.Y. (2019) uORF-mediated translational control: recently elucidated mechanisms and implications in cancer. *RNA Biol.*, **16**, 1327–1338.
10. Halterman,D.A. and Wise,R.P. (2006) Upstream open reading frames of the barley Mla13 powdery mildew resistance gene function co-operatively to down-regulate translation. *Mol. Plant Pathol.*, **7**, 167–176.
11. Guo,Z., Cao,H., Zhao,J., Bai,S., Peng,W., Li,J., Sun,L., Chen,L., Lin,Z., Shi,C. *et al.* (2022) A natural uORF variant confers phosphorus acquisition diversity in soybean. *Nat. Commun.*, **13**, 3796.
12. Xu,G., Yuan,M., Ai,C., Liu,L., Zhuang,E., Karapetyan,S., Wang,S. and Dong,X. (2017) uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature*, **545**, 491–494.
13. Zhang,H., Si,X., Ji,X., Fan,R., Liu,J., Chen,K., Wang,D. and Gao,C. (2018) Genome editing of upstream open reading frames enables translational control in plants. *Nat. Biotechnol.*, **36**, 894–898.
14. Xing,S., Chen,K., Zhu,H., Zhang,R., Zhang,H., Li,B. and Gao,C. (2020) Fine-tuning sugar content in strawberry. *Genome Biol.*, **21**, 230.
15. Gage,J.L., Mali,S., McLoughlin,F., Khaipho-Burch,M., Monier,B., Bailey-Serres,J., Vierstra,R.D. and Buckler,E.S. (2022) Variation in upstream open reading frames contributes to allelic diversity in maize protein abundance. *Proc. Natl Acad. Sci. USA*, **119**, e2112516119.
16. Kitano,S., Kurasawa,H. and Aizawa,Y. (2018) Transposable elements shape the human proteome landscape via formation of cis-acting upstream open reading frames. *Genes Cells*, **23**, 274–284.
17. Schnable,P.S., Ware,D., Fulton,R.S., Stein,J.C., Wei,F., Pasternak,S., Liang,C., Zhang,J., Fulton,L., Graves,T.A. *et al.* (2009) The B73 maize genome, complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.
18. Salvi,S., Sponza,G., Morgante,M., Tomes,D., Niu,X., Fengler,K.A., Meeley,R., Ananiev,E.V., Svitashev,S., Bruggemann,E. *et al.* (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl Acad. Sci. USA*, **104**, 11376–11381.
19. Huang,C., Sun,H., Xu,D., Chen,Q., Liang,Y., Wang,X., Xu,G., Tian,J., Wang,C., Li,D. *et al.* (2018) ZmCCT9 enhances maize adaptation to higher latitudes. *Proc. Natl Acad. Sci. USA*, **115**, E334–E341.
20. Doebley,J., Stec,A. and Hubbard,L. (1997) The evolution of apical dominance in maize. *Nature*, **386**, 485–488.
21. Studer,A., Zhao,Q., Ross-Ibarra,J. and Doebley,J. (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.*, **43**, 1160–1163.
22. Liu,Q., Deng,S., Liu,B., Tao,Y., Ai,H., Liu,J., Zhang,Y., Zhao,Y. and Xu,M. (2020) A helitron-induced RabGDIα variant causes quantitative recessive resistance to maize rough dwarf disease. *Nat. Commun.*, **11**, 495.
23. Wang,C., Yang,Q., Wang,W., Li,Y., Guo,Y., Zhang,D., Ma,X., Song,W., Zhao,J. and Xu,M. (2017) A transposon-directed epigenetic

change in ZmCCT underlies quantitative resistance to gibberella stalk rot in maize. *New Phytol.*, **215**, 1503–1515.
24. Mao,H., Wang,H., Liu,S., Li,Z., Yang,X., Yan,J., Li,J., Tran,L.S. and Qin,F. (2015) A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. *Nat. Commun.*, **6**, 8326.
25. Wang,X., Wang,H., Liu,S., Ferjani,A., Li,J., Yan,J., Yang,X. and Qin,F. (2016) Genetic variation in ZmVPP1 contributes to drought tolerance in maize seedlings. *Nat. Genet.*, **48**, 1233–1241.
26. Hirsch,C.D. and Springer,N.M. (2017) Transposable element influences on gene expression in plants. *Biochim. Biophys. Acta*, **1860**, 157–165.
27. Sundaram,V. and Wysocka,J. (2020) Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philos. Trans. R. Soc. B: Biol. Sci.*, **375**, 20190347.
28. Moschetti,R., Palazzo,A., Lorusso,P., Viggiano,L. and Marsano,R.M. (2020) 'What you need, baby, I got it': transposable elements as suppliers of cis-operating sequences in *Drosophila*. *Biology*, **9**, 25.
29. Shen,J., Liu,J., Xie,K., Xing,F., Xiong,F., Xiao,J., Li,X. and Xiong,L. (2017) Translational repression by a miniature inverted-repeat transposable element in the 3′ untranslated region. *Nat. Commun.*, **8**, 14651.
30. McClintock,B. (1950) The origin and behavior of mutable loci in maize. *Proc. Natl Acad. Sci. USA*, **36**, 344–355.
31. Atkinson,P.W. (2015) hAT transposable elements. *Microbiol. Spectr.*, **3**, 10.
32. Lazarow,K., Doll,M.L. and Kunze,R. (2013) Molecular biology of maize Ac/Ds elements, an overview. *Methods Mol. Biol.*, **1057**, 59–82.
33. Xu,Z. and Dooner,H.K. (2005) Mx-rMx, a family of interacting transposons in the growing hAT superfamily of maize. *Plant Cell*, **17**, 375–388.
34. Sosso,D., Luo,D., Li,Q.B., Sasse,J., Yang,J., Gendrot,G., Suzuki,M., Koch,K.E., McCarty,D.R., Chourey,P.S. *et al.* (2015) Seed filling in domesticated maize and rice depends on SWEET-mediated hexose transport. *Nat. Genet.*, **47**, 1489–1493.
35. Chen,W., Chen,L., Zhang,X., Yang,N., Guo,J., Wang,M., Ji,S., Zhao,X., Yin,P., Cai,L. *et al.* (2022) Convergent selection of a WD40 protein that enhances grain yield in maize and rice. *Science*, **375**, eabg7985.
36. Liang,Z. and Schnable,J.C. (2016) RNA-seq based analysis of population structure within the maize inbred B73. *PLoS One*, **11**, e0157942.
37. McKenna,A., Hanna,M., Banks,E., Sivachenko,A., Cibulskis,K., Kernytsky,A., Garimella,K., Altshuler,D., Gabriel,S., Daly,M. *et al.* (2010) The genome analysis toolkit, a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.
38. Liu,J., Huang,J., Guo,H., Lan,L., Wang,H., Xu,Y., Yang,X., Li,W., Tong,H., Xiao,Y. *et al.* (2017) The conserved and unique genetic architecture of kernel size and weight in maize and rice. *Plant Physiol.*, **175**, 774–785.
39. Liu,Y.J., Xiu,Z.H., Meeley,R. and Tan,B.C. (2013) Empty pericarp5 encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. *Plant Cell*, **25**, 868–883.
40. Kim,D., Paggi,J.M., Park,C., Bennett,C. and Salzberg,S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*, **37**, 907–915.
41. Liu,Q., Shvarts,T., Sliz,P. and Gregory,R.I. (2020) RiboToolkit: an integrated platform for analysis and annotation of ribosome profiling data to decode mRNA translation at codon resolution. *Nucleic Acids Res.*, **48**, W218–W229.
42. Li,M., Noshay,J.M., Dong,X., Springer,N.M. and Li,Q. (2021) A capture-based assay for detection and characterization of transposon polymorphisms in maize. *G3 (Bethesda)*, **11**, jkab138.
43. Sun,S., Zhou,Y., Chen,J., Shi,J., Zhao,H., Zhao,H., Song,W., Zhang,M., Cui,Y., Dong,X. *et al.* (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.*, **50**, 1289–1295.
44. Springer,N.M., Anderson,S.N., Andorf,C.M., Ahern,K.R., Bai,F., Barad,O., Barbazuk,W.B., Bass,H.W., Baruch,K., Ben-Zvi,G. *et al.* (2018) The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.*, **50**, 1282–1288.
45. Quadrana,L., Silveira,A.B., Caillieux,E. and Colot,V. (2021) Detection of transposable element insertions in Arabidopsis using sequence capture. *Methods Mol. Biol.*, **2250**, 141–155.
46. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
47. Luo,R., Liu,B., Xie,Y., Li,Z., Huang,W., Yuan,J., He,G., Chen,Y., Pan,Q., Liu,Y. *et al.* (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, **1**, 18.
48. Tamura,K., Stecher,G. and Kumar,S. (2021) MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.*, **38**, 3022–3027.
49. Ricci,W.A., Lu,Z., Ji,L., Marand,A.P., Ethridge,C.L., Murphy,N.G., Noshay,J.M., Galli,M., Mejía-Guerra,M.K., Colomé-Tatché,M. *et al.* (2019) Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants*, **5**, 1237–1249.
50. Ramírez,F., Ryan,D.P., Grüning,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2, a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.
51. Ou,S., Su,W., Liao,Y., Chougule,K., Agda,J., Hellinga,A.J., Lugo,C., Elliott,T.A., Ware,D., Peterson,T. *et al.* (2019) Benchmarking transposable element annotation methods for creation of a streamlined,comprehensive pipeline. *Genome Biol.*, **20**, 275.
52. Liu,S., Yeh,C.T., Tang,H.M., Nettleton,D. and Schnable,P.S. (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS One*, **7**, e36406.
53. Kohany,O., Gentles,A.J., Hankus,L. and Jurka,J. (2006) Annotation, submission and screening of repetitive elements in Repbase: repbaseSubmitter and Censor. *BMC Bioinformatics*, **7**, 474.
54. Lei,L., Shi,J., Chen,J., Zhang,M., Sun,S., Xie,S., Li,X., Zeng,B., Peng,L., Hauck,A. *et al.* (2015) Ribosome profiling reveals dynamic translational landscape in maize seedlings under drought stress. *Plant J.*, **84**, 1206–1218.
55. Rodríguez-Leal,D., Lemmon,Z.H., Man,J., Bartlett,M.E. and Lippman,Z.B. (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell*, **171**, 470–480.
56. Bommert,P., Nagasawa,N.S. and Jackson,D. (2013) Quantitative variation in maize kernel row number is controlled by the FASCIATED EAR2 locus. *Nat. Genet.*, **45**, 334–337.
57. Yang,Q., Li,Z., Li,W., Ku,L., Wang,C., Ye,J., Li,K., Yang,N., Li,Y., Zhong,T. *et al.* (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc. Natl Acad. Sci. USA*, **110**, 16969–16974.
58. Zeng,R., Li,Z., Shi,Y., Fu,D., Yin,P., Cheng,J., Jiang,C. and Yang,S. (2021) Natural variation in a type-A response regulator confers maize chilling tolerance. *Nat. Commun.*, **12**, 4713.
59. Hickman,A.B., Perez,Z.N., Zhou,L., Musingarimi,P., Ghirlando,R., Hinshaw,J.E., Craig,N.L. and Dyda,F. (2005) Molecular architecture of a eukaryotic DNA transposase. *Nat. Struct. Mol. Biol.*, **12**, 715–721.
60. Karakülah,G. and Pavlopoulou,A. (2018) In silico phylogenetic analysis of hAT transposable elements in plants. *Genes*, **9**, 284.