The soybean (*Glycine max* L.) cytokinin oxidase/dehydrogenase multigene family; identification of natural variations for altered cytokinin content and seed yield

Nguyen Ngoc Hai[1] ORCID 0000-0001-5155-4212, Shrikaar Kambhampati[1,2*] ORCID 0000-0003-1582-0733, Anna Kisiala[1] ORCID 0000-0002-4918-7836, Mark Seegobin[1], RJ Neil Emery[1] ORCID 0000-0003-1304-6430

[1] Department of Biology, Trent University, 1600 West Bank Drive, Peterborough, ON K9J 7B8, Canada

[2] Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO 63132, USA

[*] Corresponding author: SKambhampati@danforthcenter.org

Declarations

Funding: This work was supported by Sevita International and a grant from the Natural Sciences and Engineering Research Council of Canada (NSERC Collaborative Research & Development Grant No: CRDPJ-387250) to RJNE.

Conflicts of interest/Competing interests: The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval: Not applicable

Consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability: Not applicable

Author's contributions: HNN: Data curation, Writing - Reviewing and Editing; SK: Investigation, Data curation, Visualization, Writing - Original draft preparation, Writing - Reviewing and Editing; AK: Data curation, Writing - Reviewing and Editing; MS: Data curation, Writing - Reviewing and Editing; RJNE: Conceptualization, Supervision, Funding acquisition, Writing - Reviewing and Editing. HNN and SK contributed equally.

Acknowledgements: Authors thank Ms. Amy Galer for analyzing soybean phytohormone data by HPLC-MS/MS.

Abstract

Cytokinins (CKs) play a fundamental role in regulating dynamics of organ source/sink relationships during plant development, including flowering and seed formation stages. As a result, CKs are key drivers of seed yield. The cytokinin dehydrogenase/oxidase (CKX) gene family is of critical importance for regulating plant CK levels by causing their irreversible degradation. Variation of *CKX* gene activity is significantly correlated with seed yield in many crop species while in soybean (*Glycine max* L.), the possible associations between *CKX* gene family members (GFMs) and yield parameters have not yet been assessed. In this study, seventeen *GmCKX* GFMs were identified, and natural variations among *GmCKX* genes were probed among soybean cultivars with varying yield characteristics. The key *CKX* genes responsible for regulating CK content during seed filling stages of reproductive development were highlighted using comparative phylogenetics, gene expression analysis and CK metabolite profiling. Five of the seventeen identified *GmCKX* GFMs, showed natural variations in the form of single nucleotide polymorphisms (SNPs). Three non-synonymous variations, T69A, H105Q and A159G, in the gene *GmCKX14* with high expression during critical seed filling stages, were predicted to affect protein stability, resulting in a higher CK accumulation in soybean lines with desired yield characteristics. The potential for marker-assisted selection based on these natural variations within *GmCKX14*, is discussed in the context of hormonal control that can result in higher soybean yield.

Keywords: *CKX*, CK, natural genetic variation, SNP, phylogenetic analysis, protein modeling

Key Message

Natural variations in soybean cytokinin oxidase/dehydrogenase gene, *GmCKX14*, with high expression during seed development, were linked to increased sink strength via altered cytokinin profiles in high yielding cultivars.

Introduction

Soybean (*Glycine max* L.) is a legume species that is cultivated on a global scale, and is important to food and biomaterials industries (Liu et al. 2020). The high oil and protein content of mature seeds make it a valuable crop. However, soybean cultivation performs well below its full potential due to high rates of flower and pod abortion during reproductive development (Nico et al. 2016; He et al. 2019). Advanced genomics techniques including gene mapping, genome wide association studies, marker development and molecular breeding can be used to enhance the efficiency of classic selection and accelerate soybean yield improvement (Hua et al. 2019; Afzal et al. 2020). Thus, employing genomic studies for the identification and characterization of relevant genes and the development of molecular markers responsible for yield (e.g. higher flowering and pod setting rates, increased number of seed and 100 seed weight) is important for achieving a Green Revolution in soybean (Liu et al. 2020). Notably, previous research has reported a significant role for certain signaling molecules, the cytokinins (CKs), in regulating flower abortion rates and for increasing the number of pods being set in several crop species (Carlson et al. 1987; Atkins and Pigeaire 1993; Nonokawa et al. 2012; Emery et al. 2000; Emery and Atkins 2006). This includes soybean, as enhanced CK content was reported in reproductive tissues during pod set and seed filling phases (Jameson and Song 2016; Liu et al. 2018; Zuñiga-Mayo et al. 2018).

Cytokinins are adenine derivatives which are differentiated based on side chain substitutions at the N^6 position of the purine ring. Four primary isoprenoid CKs occur in nature: isopentenyladenine (iP), *trans*-zeatin (*trans*Z), *cis*-zeatin (*cis*Z) and dihydrozeatin (DZ) (Sakakibara 2006). The occurrence and physiological activity of CKs can differ not only among plant species, but also among different plant organs (Gajdosová et al. 2011). CK homeostasis is balanced through the rate of their biosynthesis, interconversion, inactivation and degradation (Sakakibara 2006). The rate limiting step in CK biosynthesis is regulated by an enzyme isopentenyl transferase (IPT), while their breakdown is mainly controlled by another enzyme, cytokinin oxidase/dehydrogenase (CKX) (Hirose et al. 2008). CKX catalyzes CK degradation by cleavage of the N^6-isoprenyl sidechain (Kieber and Schaller 2014).

The evolutionary development of the CK-catabolizing genes and their distinct individual properties indicate there are important roles for CKX enzymes in CK regulation (Wang et al. 2020). In plant species, CKX enzymes are produced by multi-gene families. Recently, progress has been made in identifying the mechanism of *CKX* gene family members (GFMs) to improve seed grain yield in several important crops, such as rice (*Oryza sativa* L.) (Gao et al. 2014; Gao et al. 2019; Mao et al. 2020), barley (*Hordeum vulgare* L.) (Zalewski et al. 2014; Holubová et al. 2018; Gasparis et al. 2019), and wheat (*Triticum aestivum* L.) (Zhang et al. 2012; Lu et al. 2015; Chang et al. 2015; Chen et al. 2020), making *CKX* an emerging genetic target for crop yield improvement (Jameson and Song 2016; Chen et al. 2020). Considering the potential

of *CKX* genes in biotechnological strategies for agronomy, Le et al. (2012) identified and characterized expression of *CKX*s in soybean seedlings. However, no functional studies and no genetic markers for yield in relation to CKs have been reported in soybean to date.

An enhanced level of endogenous CKs in plants is often a result of loss of function or knock out of *CKX* genes, that can lead to increased seed numbers and seed weights (Chen et al. 2020). In Arabidopsis (*Arabidopsis thaliana* L.), *AtCKX3/AtCKX5* double mutants formed larger inflorescences, with increased flowers and seed numbers (Bartrina et al. 2011). In cotton (*Gossypium* spp.), moderate suppression of *CKX* genes led to higher yields of seed and fiber and this was brought about by several CK controlled traits including: delayed leaf senescence, increased photosynthetic activity, enhanced formation of fruiting branches and bolls, and enlarged seed size (Zhao et al. 2015). A similar effect has been documented in barley, for which silencing of *HvCKX1* and *HvCKX9* stimulated seed sink strength, and effectively increased grain weight (Zalewski et al. 2010, 2014). Subsequently, silencing of *HvCKX1* via RNAi led to enhanced plant productivity (Holubová et al. 2018).

Landmark reports in CK research using rice and wheat reported that natural genetic variation within *CKX* genes can significantly impact yield traits. One of the first reports on *CKX*-associated yield increases was a natural variant of rice *OsCKX2* gene which had a premature stop codon (Ashikari et al. 2005). Association of *OsCKX2* activity with grain number was further confirmed by knocking out its expression which resulted in increased tiller number and improved yield, as well as reduced yield loss under salinity stress conditions (Joshi et al. 2018). Similarly, Yeh et al. (2015) reported that reduced expression of *OsCKX2* resulted in higher CK accumulation in reproductive organs, thereby enhancing grain yield. Furthermore, the natural variation in *TaCKX6-D1* – a wheat ortholog of rice *OsCKX2*, was significantly associated with grain weight (Zhang et al. 2012). Novel allelic variation within another wheat *CKX* gene - *TaCKX6a02*, was also correlated with grain size, weight, and seed filling rate (Lu et al. 2015).

These functional studies suggest that, though plant *CKX* genes belong to multigene families, loss of function of a single gene that is expressed at critical stages of seed development can result in a dramatic yield change. Therefore, manipulation of CK metabolism throughout modifying the *CKX* activity can be effectively used to enhance yield in important crop species. The cultivation of genetically modified crops has become a contentious topic because of a range of issues related to food security, environmental risks and human health (e.g. gene dispersal, loss of biodiversity, the emergence of superweeds and superpests, raising antibiotic resistance, food allergies and other side effects; Gassmann et al. 2011; Ahmad and Mukhtar 2017). As these matters will be difficult to resolve in the foreseeable future, the identification and exploitation of natural variation resources becomes an increasingly valued approach for breeding programs

with the purpose of improving crop yields, especially for cases when GMO technology is either not available or meets with consumer or industry adoptive reluctance. The analysis of the single nucleotide polymorphisms (SNPs) is one of the effective techniques of exploring natural genetic variation in plants. Each SNP represents a difference in a single DNA base which can effectively change the phenotype of the modified alleles if they occur at functionally important locations (Mammadov et al. 2012; Su et al. 2019).

This work reports on a multi-pronged study of *CKX* gene family members in soybean including phylogenetic analysis, screening for natural variations, transcript expression and CK abundances among cultivars. A total of sixteen soybean cultivars with varying yield parameters were probed for natural variations in *GmCKX* genes. A number of single nucleotide polymorphisms (SNPs) were identified and significant evidence for association between the SNPs in *GmCKX14* and soybean yield parameters is presented. The outcome is a natural variant marker of CK metabolism for immediate use in plant breeding programs or, for further genetic engineering to develop new, high-yielding soybean cultivars.

Materials and Methods

Plant materials and growth conditions

Sixteen soybean cultivars - 9 large size beans and 7 natto beans (Sevita International, Canada) with varying seed sizes and yield levels (**Table S1**) were used in this study. Soybean plants were grown in the Trent University Aurora research greenhouse (Conviron, Canada; photoperiod 16/8h, 27/19 ^{o}C, 60-80% relative humidity (RH)). Seeds were planted in Sunshine Mix #1, professional growing mixture (Sun Gro Horticulture, Canada). Plants were watered daily and fertilized with Jack's 15-16-17 fertilizer (JR Peters, USA) twice per week until the initiation of reproductive development. Leaves, pods and/or seeds were collected at developmental stages specific to the experiments. Stages of seed development were determined by visual appearance following descriptions previously indicated [https://extension.umn.edu/growing-soybean/soybean-growth-stages] including: R4 (full pod) at 20-30 days after flowering, R5 (beginning seed) at 30-45 days after flowering, R5.5 (intermediate stage between R5 and R6), and R6 (full seed) at 45-65 days after flowering.

Identification of *CKX* genes in soybean and phylogenetic analysis

In order to identify the soybean *CKX* GFMs, full length sequences of the previously identified genes of *Arabidopsis thaliana CKXs* and *Oryza sativa CKXs* (Gu et al. 2010) were used as a query to search the soybean genome database in phytozome v9.1 (http://phytozome.net/soybean.php) (Schmutz et al. 2010). The 17 putative soybean *CKX* genes were confirmed based on the presence of domain signatures

(Schmülling et al. 2003), necessary for enzyme function, e.g. FAD binding domain (PF01565) and CK binding domain (PF09265) against the Protein Family (PFAM) database (Finn et al. 2014). Gene nomenclature was based on the chromosome on which they occurred and, in case of tandem duplications, the order of their location. The identified *GmCKX* GFMs along with their respective gene and protein characteristics are presented in **Table 1**.

For phylogenetic analysis, the soybean CKX protein sequences along with previously reported protein sequences of the CKX proteins from the model plant species Arabidopsis, and two crops, rice and corn (Gu et al. 2010) were first aligned using the ClustalW interface in MEGA software (Penn State University, USA) (gap open penalty of 10 and gap extension penalty of 0.2; Rédei 2008; Hall 2013). The neighbourhood-joining method adopted from Gu et al. (2010) was used to construct a phylogenetic tree. Pair-wise deletion was applied to circumvent the gaps and Poisson correction distance was used to estimate the evolutionary distance. A bootstrap test with 1000 bootstrap replicates was conducted to check the reliability of the clusters.

DNA isolation and amplification of *GmCKX* genes

Young leaves were collected from each soybean cultivar at the third trifoliate stage (V3). Approximately, 100-150 ng/µL of DNA was extracted from 100 mg of ground leaf tissue using DNeasy Plant Mini Kit (Qiagen Inc., Canada) following the manufacturer's protocol. The levels of DNA in the extracted samples were quantified using NanoDrop 8000 (Fisher Scientific, Canada) spectrophotometer. All extracted DNA samples were diluted to a standard concentration of 30 ng/µL prior to further analysis.

Degenerate primers specific to each gene were designed using Primer3plus software [https://primer3plus.com/] for all 17 identified *CKX* genes following the guidelines for primer design provided by the PCR amplification kit manufacturer (Promega Corporation, USA; **Table S2**). PCR amplification of the identified *CKX* GFMs was performed using the reaction kits selected based on the gene size. Genes smaller than 4kb were amplified using the GoTaq® Flexi DNA polymerase kit (Promega Corporation, USA). Cycling conditions were: 95 ^{o}C for 1 min, 35 cycles (95 ^{o}C for 10 s, 50-59 ^{o}C for 5 s and 72 ^{o}C for 2-4 min), 72 ^{o}C for 5 min, followed by a hold at 4 ^{o}C. Genes bigger than 4kb were amplified using Expand Long Range dNTPack (Roche, Switzerland). Cycling conditions were: 92 ^{o}C for 2 min, 10 cycles (92 ^{o}C for 10 s, 50-57 ^{o}C for 5 s and 68 ^{o}C for 2-7 min), 25 cycles (92 ^{o}C for 10 s, 50-55 ^{o}C for 5 s and 68 ^{o}C for 4.5-7 min), 68 ^{o}C for 7 min, followed by a hold at 8 ^{o}C. Detailed information on the annealing temperature and extension time for each of the 17 genes is included in **Table S2**. PCR products were purified using the QIAquick PCR purification kit (Qiagen Inc., Canada), quantified using a NanoDrop™ 8000 (Fisher Scientific, Canada) spectrophotometer, and diluted to a standard concentration of 20ng/µL.

Sequencing analysis of *GmCKX* genes and identification of single nucleotide polymorphisms (SNP)

Sequencing primers were designed for every 400 bp, excluding larger intron regions, with Primer3plus software [https://primer3plus.com/] (**Table S3**). To amplify the 400bp regions, sequencing reaction mixes were prepared using the Big Dye Terminator V3.1 Cycle Sequencing Kit (Fisher Scientific, Canada). Cycling conditions were: 96 ^{o}C for 1 min followed by 35 cycles of 96 ^{o}C for 10 s, 50 ^{o}C for 5 s and 60 ^{o}C for 4 min. To determine if the PCR products were derived from a single locus or from two or more homoeologous regions, the resultant products were separated using a 3730 DNA Analyzer (Fisher Scientific, Canada) and analysed using Sequence Analysis software V5.4 (Fisher Scientific, Canada).

The obtained sequences from 16 soybean cultivars for each of the 17 *CKX* genes were compared to the reference genome from the phytozome v9.1 database using the MEGA software (Penn State University, USA) for any differences. The electropherograms of the identified differences were manually verified to eliminate ambiguity in nucleotide base calls. The regions of single nucleotide polymorphism (SNP) from the cultivars that showed differences from the reference gene sequence were re-sequenced along with 2 control cultivars, using fresh plant material and newly extracted DNA, in both forward and reverse direction with new primers (**Table S4**) to confirm the presence of SNPs.

Predicted functional effects of the identified SNPs

To examine the potential protein structure of SNPs, swiss-models were constructed using Expert Protein Analysis System (ExPASy; Swiss Institute of Bioinformatics, Switzerland; Biasini et al. 2014). A PDB file of the constructed protein model was analysed using PyMol v2.0 software (The PyMOL Molecular Graphics System 2.0, Schrodinger, LLC., https://pymol.org/2/). The SNP location was identified in the primary protein structure and any effects were predicted based on the properties of the substituted amino acid (Betts and Russell 2007). Protein stability was predicted using the Site Directed Mutator software (SDM; UK; Topham et al. 1997; Pandurangan et al. 2017).

Expression analysis of *GmCKX* genes by real-time qPCR

Total RNA was extracted from soybean seeds (OAC Wallace) at 3 developmental stages: R5, R5.5, and R6, using RNeasy Lipid Tissue Mini Kit (Qiagen Inc., Canada) and treated with DNase I (Qiagen Inc., Canada) before reverse transcription and cDNA synthesis to remove any genomic DNA contamination (Song et al. 2012). RNA samples were quantified using Nanodrop 8000 (Fisher Scientific, Canada) before reverse transcription. The first strand cDNA was synthesized using Omniscript Reverse Transcriptase Kit (Qiagen Inc., Canada) and Oligo-dT primers (Integrated DNA Technologies, Inc., Canada). The first strand cDNA and the gene-specific primers of 11 *GmCKX* genes were used in qRT-PCR reactions (**Table 5**).

Primers used for qRT-PCR were designed with Eurofins' Primer Design Tools (Eurofins, North York, ON, Canada) using following parameters: primer sequence length: 20-22 nt, GC content: 50% and 55%, melting temperature (Tm): 55 – 60 ^{o}C, and the size of the amplicon: 90-100 bp. The GoTaq qPCR Master Mix Kit (Promega Corporation, USA) was used for the gene expression profiling. The housekeeping gene *GmGAPDH* was used as an endogenous control for normalization of expression analysis. The qRT-PCR was performed using a StepOne Plus Real-Time PCR Thermo-cycler (Fisher Scientific, Canada), with an initial GoTaq® Hot Start Polymerase activation step at 95 ^{o}C for 2 minutes, followed by 40 cycles of denaturation at 95 ^{o}C for 15 seconds, then annealing and extension at 60 ^{o}C for 1 minute under the Standard Cycling Conditions.

The expression levels of the *GmCKX* genes were tested in 3 biological replicates and each reaction was analysed in 2 technical replicates. Relative gene expression levels and fold changes were determined by comparison with the most highly expressed *GmCKX* gene for each corresponding seed developmental stage (Le et al. 2012). The differences at each growth stage and between the stages were compared statistically for significance (ANOVA and Duncan's multiple range test; $p<0.05$).

Cytokinin extraction and quantification using High Performance Liquid Chromatography – Electrospray Ionization - Tandem Mass Spectrometry (HPLC-(ESI+)-MS/MS)

Cytokinin profiles were analysed from pods (collected at R4 and R5 stage) and seeds (collected at R5.5 and R6 stages) of 4 soybean cultivars: OAC 06-34, DH420, OAC Wallace and DH3604. Each sample type was analysed in 3 biological replicates. The CK extraction procedure was performed as previously described (Farrow and Emery 2012). Profiles of 25 CKs were scanned for and analyzed, if present, including: four CK free bases (FB), their corresponding riboside (RB), nucleotide precursors (NT), nine glucoside (GLUC) and four 2-methylthiol conjugates (MET).

Purified CKs were identified and quantified by HPLC-(ESI+)-MS/MS (Agilent 1100 series HPLC (Agilent, USA) connected to a QTrap 5500 mass spectrometer (Sciex, USA)), according to the conditions outlined in Farrow and Emery (2012). An aliquot was injected on a Luna C18 reversed-phase column (3 µm, 150 × 2.0 mm; Phenomenex, USA), and the CK were eluted with an increasing gradient of 0.08% CH_3COOH in CH_3CN (A) mixed with 0.08% CH_3COOH in ddH_2O (B), at a flow rate of 0.2 mL/min. The initial conditions were 5% A and 95% B, changing linearly in 17 min to 95% A and 5% B. Conditions remained constant for 5 min, and then, immediately returned to initial conditions for 18 min. The effluent was introduced into the electrospray source using conditions specific for each CK, and quantification was obtained by multiple reaction monitoring of the protonated intact CK molecule $[M+H]^+$ and the specific product ion.

The differences in the levels of the select CK groups (FB, RB, NT) and in CK types (tZ, cZ, DZ and iP-type) belonging to these three groups were compared statistically for significance (ANOVA and Duncan's multiple range test; $p<0.05$) among the 4 cultivars at each growth stage.

Results

Identification and genetic analysis of the *CKX* GFMs in soybean

The BLAST search report using Arabidopsis and rice *CKX* genes produced 17 putative soybean *CKX* genes (*GmCKX*s). Sequence analysis using the Protein Family PFAM database confirmed the presence of domain signatures required for enzyme function (i.e. FAD binding and CK binding domains) in all 17 soybean *CKX* homologs.

The identified soybean *CKX* genes were predicted to be located on 10 different chromosomes and disrupted by introns. All *GmCKX* genes displayed a minimum of 4 exons, with a few having 6 and 7 exons, revealing a high complexity of the sequences (**Table 1**). The predicted *GmCKX* genes ranged from 2596 to 7205 bp, with amino acids ranging from 442 to 552 aa per protein sequence, molecular weight ranging from 51.0 to 62.3 KDa, and theoretical isoelectric point ranging from 4.95 to 9.22. The predicted molecular weights and isoelectric points of their proteins are similar to those of other plant species exhibiting similar protein functions. Soybean *CKX* GFMs were predicted to have 1 to 6 glycosylation sites (**Table 1**) suggesting post translational modification and a broad range of functions in plants.

Phylogenetic analysis of *GmCKX* GFMs

To dissect the phylogenetic relationships of the *GmCKX* GFMs, a multiple alignment analysis was performed using the full amino acid sequences of the *CKX* GFMs of soybean, Arabidopsis, rice and corn. Based on the alignment analysis, *GmCKX* genes were distributed into 4 out of 6 groups of the generated phylogenetic tree – I, III, IV and V (**Fig. 1**, **Fig. S1)** with well supported bootstrap values. Group I consisted only of dicot genes, groups II and VI consisted of only monocot genes, while groups III, IV and V contained both dicot and monocot genes. This is congruent with a similar set of 6 groups that was generated in a phylogenetic study on corn *CKX* GFMs (Gu et al. 2010)

*Gm*CKX gene sequencing and identification of single nucleotide polymorphism (SNP)

To identify potential variations in *GmCKX* GFMs, completed sequencing of 17 *GmCKX* genes was performed from 16 soybean cultivars. Among the natural variation detected in the analysed genes, total of 9 non-synonymous SNPs were found in 5 *GmCKX* genes (*GmCKX04-1*, *GmCKX04-2*, *GmCKX06*,

GmCKX09-1, GmCKX14). Those SNPs are listed in **Table 2** which identifies their potential destabilizing effect and the cultivars in which each SNP is present.

The *GmCKX14* sequence showed 8 nucleotide variations within the first exon, present in 4 regular bean type cultivars: OAC Wallace, IA1010LF, and DH420 - characterized by high seed weight (>20 g per 100 seeds), and in high yielding DH748 (over 4 t/ha; **Table S1**). Out of the 8 SNPs identified in *GmCKX14*, 3 were non-synonymous and resulted in amino acid changes: T59A, H105Q and A159G (**Table 2**). A comparison of the homology model proteins of the two haplotypes created by ExPASy and a prediction of the effect of variation on the protein using SDM revealed that the SNP presence affects the stability of the enzyme produced by the modified *GmCKX14* gene (ΔΔG of -0.75). The predicted destabilization was strengthened by the fact that one of the *GmCKX14* natural polymorphisms, where alanine is substituted for glycine in 159 position (A159G), occurred close to the active site of the protein (**Fig. 2**). This substitution could affect the activity of the CKX enzyme by reducing its affinity to the CK substrate and leading to higher CK accumulation and hence, improved yield parameters.

Based on the presence of the destabilizing SNPs, 2 cultivars, OAC Wallace and DH420 were selected for further analysis of CK metabolites and compared to 2 other cultivars that did not have the *GmCKX14* SNP. In addition, one cultivar, OAC Wallace was used for the analysis of the selected *GmCKX* GFMs.

Expression patterns of Gm*CKX* GFMs during soybean reproductive development

In the developing soybean seeds of OAC Wallace, *CKX* GFMs, 6 of the eleven analysed *GmCKX* genes had relatively high expression levels at R5, R5.5, and R6 growth stages: *GmCKX14, GmCKX03, GmCKX19, GmCKX04-1, GmCKX04-2,* and *GmCKX15* (**Fig. 3**). *GmCKX14* had, by far, the highest expression level in all three seed stages compared to the other *GmCKX*s (ANOVA, Duncan's multiple range test, $p<0.05$). Additionally, the transcripts of *GmCKX14* and *GmCKX19* showed clear, increasing trends along the seed development progression from R5 to R6 stage (ANOVA, Duncan's multiple range test, $p<0.05$).

Cytokinin quantification by HPLC-MS/MS during soybean reproductive development

The levels of 25 forms of CKs were measured during soybean seed development using HPLC-MS/MS (**Fig. 4**, **Table S6a-b**). Notably, distinct patterns in the CK form and levels were recorded across the 4 stages of pods (R4, R5) and seeds (R5.5, R6) among the 2 high yielding cultivars, OAC Wallace and DH420 (both of which have the *GmCKX14* SNPs), and the other 2 analysed cultivars (OAC 06-14 and DH3604 – which do not have the *GmCKX14* SNPs). The observed trends are consistent with the reduced *GmCKX14* activity. For example, DH420 had the highest level of CK precursors, NTs, in pods at R4 and R5 stages (ANOVA,

Duncan's multiple range test, $p<0.05$). In OAC Wallace, at the later R6 stage, the FB CK levels were the highest among the tested soybeans (ANOVA, Duncan's multiple range test, $p<0.05$). This strongly aligns with the results of gene expression analysis for *GmCKX14* in R6 seeds of this cultivar, since *GmCKX14* at R6 was the most highly expressed gene of any timepoint. Regarding CK-type (cZ, tZ, DZ and iP-type) distribution, the highest yielding cultivar, OAC Wallace, revealed the highest accumulation of the most active, tZ-type CKs in seeds at both R5.5 and R6 stages (ANOVA, Duncan's multiple range test, $p<0.05$). Additionally, at R6 stage, cZ levels in OAC Wallace exceeded those in the remaining 3 cultivars (ANOVA, Duncan's multiple range test, $p<0.05$). Cytokinin-type profiles were different for the second *GmCKX14* SNP cultivar, DH420, but there were still remarkable patterns which may explain its high yield. Specifically, the levels of cZ and iP derivatives were the most abundant at R4 compared to the other 3 soybeans.

Discussion

Cytokinins (CKs) play essential roles in numerous plant physiology processes. Among their multiple functions, CKs are a key drivers of seed yield (reviewed by Jameson and Song 2016; Chen et al. 2020). In soybean, CKs are: involved in preventing flower abortion and in pod setting (Nonokawa et al. 2012); associated with seed development and fatty acid biosynthesis (Nguyen et al. 2016); and ultimately, responsible for seed yield (Kambhampati et al. 2017). Studies on genetic analysis of CKs aid greatly in the understanding of fundamentals of plant physiology and they help build genetic resources for high yielding crop breeding programs. Although CK metabolic genes have been extensively studied in several plant species (Chen et al. 2020; Wang et al. 2020a, b), only a few examples have investigated the function of CK metabolic genes in soybean yield formation (Le et al. 2012). Thus, the genetic characterization and identification of CK metabolic genes, along with the selection of unique natural variants for high-yielding traits should be a high priority for the progress in soybean breeding. To date, the present study of *GmCKX* genes is the most focused example of the drive towards the identification of CK-based markers for modern soybean breeding programs.

This work focuses on the soybean CKX GFMs because it is the best-known way to effectively regulate endogenous CKs in reproductive tissues, and thus improve yield. While IPT is the limiting step for CK biosynthesis, it is more complicated to control, and no variants exist in any crops that are linked to high yield. By contrast CKX can degrade CKs in a spatially controlled manner and examples of both natural variants (e.g. Ashikari et al. 2005) and manipulation through genetic engineering (e.g. Zalewski et al. 2014) for enhanced crop yields are convincing. Our work employed a multi-pronged approach to study CK

metabolism and its mechanistic association with soybean yield parameters. This included a *CKX* protein sequence based phylogenetic analysis, SNP identification in *CKX* GFMs among different yielding cultivars, bioinformatic predictions of SNP destabilization potentials, and *CKX* gene expression analyses as well as comprehensive CK metabolite profiling during pod and seed filling.

In this study, 17 *CKX* GFMs in soybean were identified (**Fig. S1**). Previously, the presence of paralogous *CKX* genes were reported from multiple species including rice (11 *CKX* GFMs; Mameaux et al. 2012), foxtail millet (11 *CKX* GFMs; Wang et al. 2014), corn (13 *CKX* GFMs; Zalabák et al. 2014), potato (5 *CKX* GFMs; Suttle et al. 2014), wild strawberry (8 *CKX* GFMs; Jiang et al. 2016), apple (12 *CKX* GFMs; Tan et al. 2018) and wheat (11 *CKX* GFMs; Chen et al. 2020), all of which were much lower numbers than those identified here for soybean. A higher number of *CKX* metabolic genes in the soybean genome compared to that of other species is likely a consequence of genome evolution and duplication, which produces homologous genes, increasing the total gene number (Kaltenegger et al. 2018; Wang et al. 2020a).

A multiple alignment analysis of the *CKX* GFMs of rice, corn, Arabidopsis and soybean revealed 6 distinctive clusters that grouped the CK degradation genes (**Fig. 1**). Group I consisted of only dicot genes, and groups II and VI of only monocot genes, suggesting that a relative genetic divergence may have occurred in *CKX* genes during evolution and speciation of the two plant groups (Tan et al. 2018). Groups III, IV and V contained both dicot and monocot genes, indicating that the expansion of these genes occurred before the monocot/dicot divergence (Chaw et al. 2004; Hertweck et al. 2015).

A phylogenetic tree further revealed that most of the soybean *CKX* genes have counterparts in Arabidopsis. For example, 4 of the *GmCKX* GFMs: *GmCKX09-2*, *GmCKX13-1*, *GmCKX15*, *GmCKX17-1* and Arabidopsis *AtCKX3*; *GmCKX09-1*, *GmCKX11*, *GmCKX12* and *AtCKX6*; *GmCKX03*, *GmCKX19* and *AtCKX1*; *GmCKX04-2*, *GmCKX14*, *GmCKX17-3* and *AtCKX7*; *GmCKX04-1*, *GmCKX06* and *AtCKX5* belong to the same clades (**Fig. 1**), suggesting soybean and Arabidopsis *CKX* genes descended from the same ancestor. The fact that not every soybean *CKX* gene has a counterpart in Arabidopsis may reflect that these two species have undergone differentiated expansions.

The genes assigned to group V (*GmCKX04-2*, *GmCKX14*, *GmCKX17-3 and ZmCKX10*, *AtCKX7*, *OsCKX11*) were previously found to have essential roles in plant growth and development. A cytosolic corn enzyme, *ZmCKX10*, preferentially degrades cisZeatin (cZ) and has an abundant transcript in embryos, leaves and pedicels (Šmehilová et al. 2009). In Arabidopsis, *AtCKX7* displays high affinity towards cZ (Gajdošová et al. 2011) while the gain-of-function mutant of *AtCKX7* exhibited an increased CKX enzymatic activity, leading to CK deficiency and altered vascular development (Köllmer et al. 2014). *OsCKX11* coordinates source and sink relationship, thus regulating leaf senescence and grain number in

rice (Zhang et al. 2020). In soybean, *GmCKX04-2* and *GmCKX14* were highly expressed in the leaves, roots and root hairs at reproductive stage under normal and drought conditions (Le et al. 2012). These reports suggest the *CKX* genes of phylogeny group V have critical effects on vegetative and embryonic growth, somatic embryogenesis, seed development, lateral root formation, petal cell identity, floral meristem identity and transition to flowering. Therefore, the polymorphisms among the soybean *CKX* genes would be most meaningful if found among the genes of group V (*GmCKX04-2*, *GmCKX14*, *GmCKX17-3*) because of their reproductive tissue specific expression, which would likely most impact soybean yield and seed weight.

Single nucleotide polymorphism (SNP) makers have attracted interest in molecular breeding because of their genome-wide availability and amenability for high throughput platforms (Su et al. 2019). Several works by industrial sector representatives such as Monsanto (Eathington et al. 2007; Rosso et al. 2011), Syngenta (Ribaut and Ragot 2007), Dow AgroSciences (Ren et al. 2008; Mammadov et al. 2012) or Hi-bred (Zheng et al. 2008) proved that development of SNP markers is of a great of interest to plant breeding industry. The use of SNP markers has been recognized as convenient method due to the common presence of SNPs in plant genotypes, their stability caused by the relatively low mutation rates, flexibility, as well as method speed, high-throughput genotyping options, and cost effectiveness (Rasheed et al. 2017; Su et al. 2019). Previously reported SNP markers in soybean are potentially of practical use in breeding (Rosso et al. 2011; Song et al. 2013; Lee et al. 2015). For example, mutations in soybean D-myo-inositol 3-phosphate synthase 1 gene (*MIPS1*) lead to the increase in inorganic phosphorous and sucrose levels, and reduced phytate and raffinosaccharide accumulation in seeds, improving the nutritional value of the food products (Maupin et al. 2011).

To identify any potential SNPs in soybean *CKX* GFMs, sequencing of 17 *GmCKX* genes was performed from 16 soybean cultivars. A total of 9 non-synonymous SNPs were found among 5 *GmCKX* genes: *GmCKX04-1*, *GmCKX04-2*, *GmCKX06*, *GmCKX09-1* and *GmCKX14* (**Table 2**). We had hypothesized that the presence of SNPs in *GmCKX14* sequences could be critical for determining reproductive organ sink strength because *GmCKX14* is normally active during seed development stages (Le et al. 2012). It was predicted that that disruptive SNPs would be present in high seed weight (100 seed weight > 20g) soybean cultivars, like OAC Wallace, DH420 and IA1010LF. Homology modelling supports the hypothesis that reducing the GmCKX14 enzymatic activity can lead to the increased accumulation of CKs, resulting in the increased sink strength of the developing seeds. Phylogeny analysis suggested that *GmCKX14* might be involved in degradation of cZ-type CKs, similarly to its *AtCKX7* ortholog (Köllmer et al. 2014; Schäfer et al. 2015). The potential substrate specificity of the *GmCKX14* gene towards cZ and any lower activity caused by the observed SNP substitutions (**Table 2**) could lead to the higher cZ-type CK levels and hence,

higher seed weight (Kudo et al. 2012; Powell et al. 2013; Hluska et al. 2016; Kambhampati et al. 2017). Therefore, a possible explanation for the higher seed weight observed in OAC Wallace, DH420 and IA1010LF cultivars, could be the potentially lower activity of the *CKX* enzyme caused by the three substitutions in *GmCKX14* sequence: T69A, H105Q and A159G, which, according to our models, affect protein stability. In particular, one of these substitutions, A159G, occurred close to the protein active site presumably altering its substrate binding affinity (**Fig. 2**). Previous reports of CK metabolite profiles further support this hypothesis, since they showed a positive correlation between soybean seed weight and levels of cZ (Kambhampati et al. 2017).

The manipulation of *CKX* GFMs can alter endogenous active CK levels, and substantially modify plant physiology and development (Jameson and Song 2016; Chen et al. 2020). Genetically modified plants, with CK profiles altered via targeting *CKX* GFMs, show enhanced tolerance to environmental stresses and improved yield, as it has been observed in: barley (Zalewski et al. 2014; Holubová et al. 2018), rice (Ashikari et al. 2005; Yeh et al. 2015; Li et al. 2016; Joshi et al. 2018) and wheat (Zhang et al. 2012; Li et al. 2018; Wang et al. 2018). In the present study, one of the cultivars, OAC Wallace, was selected for gene expression analysis since it was high yielding and it contained the SNPs in *GmCKX14* sequence. Accordingly, transcript analysis showed that that patterns of *GmCKX*s expression in developing soybean seeds differed among three reproductive growth stages (R5, R5.5, and R6; **Fig. 3**). *GmCKX14*, was exceptional since it had the highest transcript levels in each of the seed growth stages. Therefore, it can be considered as a high priority gene candidate for the further studies of the functional characteristics (CRISPR/Cas9 or overexpression) and as a target in modern breeding programs, particularly in regards to high yield and abiotic stress tolerance (Pospíšilová et al. 2016; Holubová et al. 2018; Gasparis et al. 2019). Our results align with another study that showed that *GmCKX14* had high expression in flowers, during early pod development and seed formation (Le et al. 2012). Under normal conditions, high expression of *CKX* genes reduces quantities of the active forms of endogenous CKs, and their transport (RBs) and conjugate (GLUCs) forms (Avalbaev et al. 2012). This results in the reduced organ/tissue sink strength, limited photosynthetic activity, altered distribution of carbohydrates and glycolytic enzymes and, ultimately, reduced seed growth (Werner et al. 2008; Prerostova et al. 2018). Therefore, in our work, the SNPs identified in the 4 high-yield soybean cultivars can conceivably diminish the activity of *GmCKX14*, and thereby enhance CK accumulation during the reproductive stages. This, would thus delay senescence, enhance photosynthesis and primordia formation, which, in turn, can increase the spike number, flower number and size, pod number, seed number and seed weight (Jameson and Song 2016; He et al. 2018; Panda et al. 2018). Taken together, our data support the hypothesis that the substitution in *GmCKX14* is closely linked to seed yield in soybean. These findings highlight that the natural variations observed in *GmCKX14* could be important markers for developing high-yielding cultivars in soybean breeding

programs. Further characterizations are ongoing, and new SNP markers will be developed for selection of high yielding soybean materials.

CKs metabolites are strongly associated with the enhancement of soybean reproductive development, since HPLC-MS/MS analyses have clearly established their distinct correspondence with cell division, cell expansion, and dehydration during embryo development, seed protein and lipid accumulation, and in the DNA endoreduplication (Nguyen et al. 2016; Kambhampati et al. 2017). Therefore, the present study surveyed the complete set of CKs during three reproductive stages that critically define pod and seed development and yield formation in soybean. This was done for two high yield cultivars (OAC Wallace and DH420) which contained the *GmCKX14* SNPs and two lower-yielding cultivars (OAC 06-14 and DH3604) that did not (**Fig. 4**).

Distinct CK metabolite profile attributes were observed for the two high yield cultivars. Notably, in the cultivar DH420 (*GmCKX14* SNPs identified), the inactive CK precursors, NTs, were a dominant form in R4 and R5 stages. Furthermore, the peak in cZ and iP-type CKs observed in DH420 during early pod development suggest that potential activity of the altered *GmCKX14* is important for the control of plant development and grain yield, as was the case for a previously reported mechanism of *HvCKX1* action which modified barley productivity (Holubová et al. 2018). In another high yield cultivar, OAC Wallace, (also with the *GmCKX14* SNPs identified), the active FB forms significantly peaked at the R6 stage. Two CK groups (NTs and RBs), that are transportable and are precursors to the most active FB forms were observed to preferentially accumulate in developing pods, thereby representing sources for the higher FB content in seed (Witte and Herde 2020). The FB peak was previously reported in soybean seeds at R6, at the onset of mitotic cell division events (Lin et al. 2017). The activity of FBs strongly contributes to the sink strength and accumulation of assimilates in soybean seed development (Kambhampati et al. 2017). As predicted by the protein model, the *GmCKX14* SNPs can bring about an increased accumulation of the active CK types (tZ, cZ) during R5.5 and R6 stages in OAC Wallace (**Fig. 4**). According to previous reports, high levels of these CKs accumulate in different plant organs at the reproductive stages (Le et al. 2012) and correlate with grain yield in soybean (Kambhampati et al. 2017). This indicates potential effect of *GmCKX* SNPs on CK metabolism and increased soybean productivity, and reinforces the idea that the SNPs identified in *GmCKX14* can serve as a new, highly effective marker in soybean breeding programs.

Our comprehensive CK metabolite analysis is one of the rare examples in crops, where detailed endogenous hormone profiling has been shown to be functionally important in relation to the natural variation of *CKX* genes. However, not all the data point clearly toward high yield traits. Some of the discrepancies found in this study among the metabolic, gene expression data and bioinformatics prediction of *GmCKX14* function can be at least partially explained by the redundancy effects of 17 *CKX* GFMs in soybean (Liu et al. 2008;

Lakhssassi et al. 2017). There is also the challenge of profiling metabolites that are derived from whole tissues for which resolution cannot be discerned at sub-cellular levels. It is now known that CKs are precisely sensed at multiple sites at the plasma-membrane and endoplasmic reticulum (ER) (Kubiasová et al. 2020) and that CKX isoforms are likewise specifically located, most frequently on the ER (Niemann et al. 2018). Thus, different CKX proteins would control different cytokinin pools depending on their subcellular localizations, which our methods cannot resolve.

Finding and exploiting genomic DNA variation is of utmost importance for crop genetics and breeding. In this research, soybean *CKX* gene activity has been demonstrated to be significantly associated with yield components. Gene sequencing in multiple soybean cultivars and bioinformatics predictions revealed that SNPs present among the *GmCKX* GFMs potentially affect CKX enzyme functioning. The observed *GmCKX* gene expression patterns and CK metabolite profiles supported the suggestion that the reduced *CKX* activity and increased CK accumulation, during soybean reproductive development enhanced yield parameters in SNP *GmCKX14* cultivars. The revealed natural genetic polymorphisms can be used as molecular indicators for identifying desirable soybean lines in larger genetic populations through marker assisted selection, which can save costs and time associated with breeding projects. Discovering useful genes, improving agricultural traits hidden in the plant genome, and applying these findings to crop breeding can pave the way for development of new, non-GMO soybean cultivars with higher yield and better abiotic stress resistance.

References

Afzal M, Alghamdi SS, Migdadi HH et al (2020) Legume genomics and transcriptomics: From classic breeding to modern technologies. Saudi J Biol Sci 27(1):543-555. https://doi.org/10.1016/j.sjbs.2019.11.018

Atkins CA, Pigeaire A (1993) Application of cytokinins to flowers to increase pod set in *Lupinus angustifolius* L. Aust J Agric Res 44(8):1799-1819. https://doi.org/10.1071/AR9931799

Ahmad N, Mukhtar Z (2017) Genetic manipulations in crops: challenges and opportunities. Genomics 109(5-6):494-505. https://doi.org/10.1016/j.ygeno.2017.07.007

Ashikari M, Sakakibara H, Lin S et al (2005) Plant science: Cytokinin oxidase regulates rice grain production. Science 309(5735):741-745. https://doi.org/10.1126/science.1113373

Avalbaev AM, Somov KA, Yuldashev RA, Shakirova FM (2012) Cytokinin oxidase is key enzyme of cytokinin degradation. Biochemistry (Mosc) 77(12):1354-1361.

https://doi.org/10.1134/S0006297912120024

Bartrina I, Otto E, Strnad M et al (2011) Cytokinin regulates the activity of reproductive meristems, flower organ size, ovule formation, and thus seed yield in *Arabidopsis thaliana*. Plant Cell 23(1):69-80. https://doi.org/10.1105/tpc.110.079079

Betts MJ, Russell RB (2003) Amino acid properties and consequences of substitutions. Bioinformatics for Geneticists 317:289. https://doi.org/10.1002/0470867302.ch14

Biasini M, Bienert S, Waterhouse A et al (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42(W1):W252-W258. https://doi.org/10.1093/nar/gku340

Carlson DR, Dyer DJ, Cotterman CD, Durley RC (1987) The physiological basis for cytokinin induced increases in pod set in IX93-100 soybeans. Plant Physiol 84(2):233-239. https://doi.org/10.1104/pp.84.2.233

Chang C, Lu J, Zhang HP et al (2015) Copy number variation of cytokinin oxidase gene *Tackx4* associated with grain weight and chlorophyll content of flag leaf in common wheat. PLoS One 10(12):e0145970. https://doi.org/10.1371/journal.pone.0145970

Chaw SM, Chang CC, Chen HL, Li WH (2004) Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. J Mol Evol 58(4):424-441. https://doi.org/10.1007/s00239-003-2564-9

Chen L, Zhao J, Song J, Jameson PE (2020) Cytokinin dehydrogenase: a genetic target for yield improvement in wheat. Plant Biotechnol J 18(3):614-630. https://doi.org/10.1111/pbi.13305

Eathington SR, Crosbie TM, Edwards MD et al (2007) Molecular markers in a commercial breeding program. Crop Sci 47:S154-S163. https://doi.org/10.2135/cropsci2007.04.0015IPBS

Emery RJN, Ma Q, Atkins CA (2000) The forms and sources of cytokinins in developing white lupine seeds and fruits. Plant Physiol 123(4):1593-1604. https://doi.org/10.1104/pp.123.4.1593

Farrow SC, Emery RN (2012) Concurrent profiling of indole-3-acetic acid, abscisic acid, and cytokinins and structurally related purines by high-performance-liquid-chromatography tandem electrospray mass spectrometry. Plant Methods 8(1):42-60. https://doi.org/10.1186/1746-4811-8-42

Finn RD, Bateman A, Clements J et al (2014). Pfam: the protein families database. Nucleic Acids Res 42(D1):D222-D230. https://doi.org/10.1093/nar/gkt1223

Gajdosova S, Spichal L, Kaminek M et al (2011) Distribution, biological activities, metabolism, and the conceivable function of cis-zeatin-type cytokinins in plants. J Exp Bot 62:2827–2840. https://doi.org/10.1093/jxb/erq457

Gao S, Fang J, Xu F et al (2014) CYTOKININ OXIDASE/DEHYDROGENASE4 integrates cytokinin and auxin signaling to control rice crown root formation. Plant Physiol 165(3):1035-1046. https://doi.org/10.1104/pp.114.238584

Gao S, Xiao Y, Xu F et al (2019) Cytokinin-dependent regulatory module underlies the maintenance of zinc nutrition in rice. New Phytol 224(1):202-215. https://doi.org/10.1111/nph.15962

Gasparis S, Przyborowski M, Kała M, Nadolska-Orczyk A (2019) Knockout of the *HvCKX1* or *HvCKX3* gene in barley (*Hordeum vulgare* L.) by RNA-guided Cas9 nuclease affects the regulation of cytokinin metabolism and root morphology. Cells 8:782-806. https://doi.org/10.3390/cells8080782

Gassmann AJ, Petzold-Maxwell JL, Keweshan RS, Dunbar MW (2011) Field-evolved resistance to Bt maize by Western corn rootworm. PLoS One 6(7):e22629. https://doi.org/10.1371/journal.pone.0022629

Gu R, Fu J, Guo S et al (2010) Comparative expression and phylogenetic analysis of maize cytokinin dehydrogenase/oxidase (*CKX*) gene family. J Plant Growth Regul 29(4):428-440. https://doi.org/10.1007/s00344-010-9155-y

Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol 30.5:1229-1235. https://doi.org/10.1093/molbev/mst012

He J, Jin Y, Turner NC et al (2019) Phosphorus application increases root growth, improves daily water use during the reproductive stage, and increases grain yield in soybean subjected to water shortage. Environ Exp Bot 166:103816-103827. https://doi.org/10.1016/j.envexpbot.2019.103816

He Q, Yang L, Hu W et al (2018) Overexpression of an auxin receptor *OsAFB6* significantly enhanced grain yield by increasing cytokinin and decreasing auxin concentrations in rice panicle. Sci Rep 8(1):1-11. https://doi.org/10.1038/s41598-018-32450-x

Hertweck KL, Kinney MS, Stuart SA et al (2015) Phylogenetics, divergence times and diversification from three genomic partitions in monocots. Bot J Linn Soc 178(3):375-393. https://doi.org/10.1111/boj.12260

Hirose N, Takei K, Kuroha T et al (2008) Regulation of cytokinin biosynthesis, compartmentalization and translocation. Environ Exp Bot 59:75–83. https://doi.org/10.1093/jxb/erm157

Hluska T, Dobrev PI, Tarkowská D et al (2016) Cytokinin metabolism in maize: novel evidence of cytokinin abundance, interconversions and formation of a new trans-zeatin metabolic product with a weak anticytokinin activity. Plant Sci 247:127-137. https://doi.org/10.1016/j.plantsci.2016.03.014

Holubová K, Hensel G, Vojta P et al (2018) Modification of barley plant productivity through regulation of cytokinin content by reverse-genetics approaches. Front. Plant Sci 9:1676-1694. https://doi.org/10.3389/fpls.2018.01676

Hua K, Zhang J, Botella JR et al (2019) Perspectives on the application of genome-editing technologies in crop breeding. Mol Plant 12(8):1047-1059. https://doi.org/10.1016/j.molp.2019.06.009

Kubiasová K, Montesinos JC, Šamajová O et al (2020) Cytokinin fluoroprobe reveals multiple sites of cytokinin perception at plasma membrane and endoplasmic reticulum. Nat Commun 11(1):1-11. https://doi.org/10.1038/s41467-020-17949-0

Jameson PE, Song J (2016) Cytokinin: a key driver of seed yield. J Exp Bot 67(3):593-606. https://doi.org/10.1093/jxb/erv461

Jiang Y, Mi X, Lin Y et al (2016) Evolution and expression patterns of cytokinin oxidase genes in *Fragaria vesca*. Sci Hortic 212:115–125. https://doi.org/10.1016/j.scienta.2016.09.037

Joshi R, Sahoo KK, Tripathi AK et al (2018) Knockdown of an inflorescence meristem-specific cytokinin oxidase–*OsCKX2* in rice reduces yield penalty under salinity stress condition. Plant Cell Environ 41(5):936-946. https://doi.org/10.1111/pce.12947

Kaltenegger E, Leng S, Heyl A (2018) The effects of repeated whole genome duplication events on the evolution of cytokinin signaling pathway. BMC Evol Biol 18(1):76-95 https://doi.org/10.1186/s12862-018-1153-x

Kambhampati S, Kurepin LV, Kisiala AB et al (2017) Yield associated traits correlate with cytokinin profiles in developing pods and seeds of field-grown soybean cultivars. Field Crops Res 214:175-184. https://doi.org/10.1016/j.fcr.2017.09.009

Kieber JJ, Schaller GE (2014) Cytokinins. The Arabidopsis Book/American Society of Plant Biologists. 12-47. https://doi.org/10.1199/tab.0168

Köllmer I, Novák O, Strnad M et al (2014) Overexpression of the cytosolic cytokinin oxidase/dehydrogenase (*CKX 7*) from Arabidopsis causes specific changes in root growth and xylem differentiation. Plant J 78(3):359-371. https://doi.org/10.1111/tpj.12477

Kudo T, Makita N, Kojima M et al (2012) Cytokinin activity of cis-zeatin and phenotypic alterations induced by overexpression of putative cis-zeatin-O-glucosyltransferase in rice. Plant Physiol 160(1):319-331. https://doi.org/10.1104/pp.112.196733

Lakhssassi N, Liu S, Bekal S et al (2017) Characterization of the soluble NSF attachment protein gene family identifies two members involved in additive resistance to a plant pathogen. Sci Rep 7(1):1-11. https://doi.org/10.1038/srep45226

Le DT, Nishiyama R, Watanabe Y et al (2012) Identification and expression analysis of cytokinin metabolic genes in soybean under normal and drought conditions in relation to cytokinin levels. PloS One 7(8):e42411. https://doi.org/10.1371/journal.pone.0042411

Lee YG, Jeong N, Kim JH et al (2015) Development, validation and genetic analysis of a large soybean SNP genotyping array. Plant J 81(4):625-636. https://doi.org/10.1111/tpj.12755

Li M, Li X, Zhou Z et al (2016) Reassessment of the four yield-related genes *Gn1a*, *DEP1*, *GS3*, and *IPA1* in rice using a CRISPR/Cas9 system. Front Plant Sci 7:377-390. https://doi.org/10.3389/fpls.2016.00377

Li Y, Fu X, Zhao M et al (2018) A genome-wide view of transcriptome dynamics during early spike development in bread wheat. Sci Rep 8(1):1-16. https://doi.org/10.1038/s41598-018-33718-y

Lin JY, Le BH, Chen M et al (2017) Similarity between soybean and Arabidopsis seed methylomes and loss of non-CG methylation does not affect seed development. PNAS 114(45):E9730-E9739. https://doi.org/10.1073/pnas.1716758114

Liu B, Kanazawa A, Matsumura H et al (2008) Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. Genetics 180(2):995-1007. https://doi.org/10.1534/genetics.108.092742

Liu P, Zhang C, Ma JQ et al (2018) Genome-wide identification and expression profiling of cytokinin oxidase/dehydrogenase (*CKX*) genes reveal likely roles in pod development and stress responses in oilseed rape (*Brassica napus* L.). Genes 9(3):168-184. https://doi.org/10.3390/genes9030168

Liu S, Zhang M, Feng F, Tian Z (2020) Toward a 'Green Revolution' for soybean. Mol Plant 13 (5):688-697. https://doi.org/10.1016/j.molp.2020.03.002

Lu J, Chang C, Zhang HP et al (2015) Identification of a novel allele of *TaCKX6a02* associated with grain size, filling rate and weight of common wheat. PLoS One 10(12):e0144765. https://doi.org/10.1371/journal.pone.0144765

Mameaux S, Cockram J, Thiel T et al (2012) Molecular, phylogenetic and comparative genomic analysis of the cytokinin oxidase/dehydrogenase gene family in the Poaceae. Plant Biotechnol J 10(1):67-82. https://doi.org/10.1111/j.1467-7652.2011.00645.x

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. Int J Plant Genomics 2012:1-11. https://doi.org/10.1155/2012/728398

Mao C, He J, Liu L et al (2020) *OsNAC2* integrates auxin and cytokinin pathways to modulate rice root development. Plant Biotechnol J 18(2):429-442. https://doi.org/10.1111/pbi.13209

Maupin LM, Rosso ML, Rainey KM (2011) Environmental effects on soybean with modified phosphorus and sugar composition. Crop Sci 51(2):642-650. https://doi.org/10.2135/cropsci2010.07.0396

Nguyen TQ, Kisiala A, Andreas P et al (2016) Soybean seed development: fatty acid and phytohormone metabolism and their interactions. Curr Genom 17(3):241-260. https://doi.org/10.2174/1389202917666160202220238

Niemann MC, Weber H, Hluska T et al (2018) The cytokinin oxidase/dehydrogenase *CKX1* is a membrane-bound protein requiring homooligomerization in the endoplasmic reticulum for its cellular activity. Plant Physiol 176(3):2024-2039. https://doi.org/10.1104/pp.17.00925

Nico M, Mantese AI, Miralles DJ, Kantolic AG (2016) Soybean fruit development and set at the node level under combined photoperiod and radiation conditions. J Exp Bot 67(1):365-377. https://doi.org/10.1093/jxb/erv475

Nonokawa K, Nakajima T, Nakamura T, Kokubun M (2012) Effect of synthetic cytokinin application on pod setting of individual florets within raceme in soybean. Plant Prod Sci 15(2):79-81. https://doi.org/10.1626/pps.15.79

Panda BB, Sekhar S, Dash SK et al (2018) Biochemical and molecular characterisation of exogenous cytokinin application on grain filling in rice. BMC Plant Biol 18(1):89-108. https://doi.org/10.1186/s12870-018-1279-4

Pandurangan AP, Ochoa-Montaño B, Ascher DB, Blundell TL (2017) SDM: a server for predicting effects of mutations on protein stability. Nucleic Acids Res 45(W1):W229-W235. https://doi.org/10.1093/nar/gkx439

Pospíšilová H, Jiskrova E, Vojta P et al (2016) Transgenic barley overexpressing a cytokinin dehydrogenase gene shows greater tolerance to drought stress. New Biotechnology 33(5):692-705. https://doi.org/10.1016/j.nbt.2015.12.005

Powell AF, Paleczny AR, Olechowski H, Emery RN (2013) Changes in cytokinin form and concentration in developing kernels correspond with variation in yield among field-grown barley cultivars. Plant Physiol Biochem 64:33-40. https://doi.org/10.1016/j.plaphy.2012.12.010

Prerostova S, Dobrev PI, Gaudinova A et al (2018) Cytokinins: Their impact on molecular and growth responses to drought stress and recovery in Arabidopsis. Front Plant Sci 9:655-669. https://doi.org/10.3389/fpls.2018.00655

Rasheed A, Hao Y, Xia X et al (2017) Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. Mol Plant 10(8):1047-1064. https://doi.org/10.1016/j.molp.2017.06.008

Rédei GP (2008) CLUSTAL W (improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice). In: Encyclopedia of Genetics, Genomics, Proteomics and Informatics. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6754-9_3188

Ren R, Nagel BA, Kumpatla SP et al (2018) U.S. Patent No. 10,117,411. Washington, DC: U.S. Patent and Trademark Office.

Ribaut JM, Ragot M (2007) Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. J Exp Bot 58(2):351-360. https://doi.org/10.1093/jxb/erl214

Rosso ML, Burleson SA, Maupin LM, Rainey KM (2011) Development of breeder-friendly markers for selection of *MIPS1* mutations in soybean. Mol Breed 28(1):127-132. https://doi.org/10.1007/s11032-011-9573-y

Sakakibara H (2006) Cytokinins: activity, biosynthesis, and translocation. Annu Rev Plant Biol 57:431-449. https://doi.org/10.1146/annurev.arplant.57.032905.105231

Schäfer M, Brütting C, Meza-Canales ID et al (2015) The role of cis-zeatin-type cytokinins in plant growth regulation and mediating responses to environmental interactions. J Exp Bot 66(16):4873-4884. https://doi.org/10.1093/jxb/erv214

Schmülling T, Werner T, Riefler M et al (2003) Structure and function of cytokinin oxidase/dehydrogenase genes of maize, rice, Arabidopsis and other species. J Plant Res 116(3):241-252. https://doi.org/10.1007/s10265-003-0096-4

Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. Nature 463(7278):178-183. https://doi.org/10.1038/nature08670

Šmehilová M, Galuszka P, Bilyeu KD et al (2009) Subcellular localization and biochemical comparison of cytosolic and secreted cytokinin dehydrogenase enzymes from maize. J Exp Bot 60(9):2701-2712. https://doi.org/10.1093/jxb/erp126

Song J, Jiang L, Jameson PE (2012) Co-ordinate regulation of cytokinin gene family members during flag leaf and reproductive development in wheat. BMC Plant Biol 12(1):78-95. https://doi.org/10.1186/1471-2229-12-78

Song Q, Hyten DL, Jia G et al (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PloS One 8(1):e54985. https://doi.org/10.1371/journal.pone.0054985

Su J, Zhang F, Chong X et al (2019) Genome-wide association study identifies favorable SNP alleles and candidate genes for waterlogging tolerance in chrysanthemums. Hortic Res 6(1):1-13. https://doi.org/10.1038/s41438-018-0101-7

Suttle JC, Huckle LL, Lu S, Knauber DC (2014) Potato tuber cytokinin oxidase/dehydrogenase genes: Biochemical properties, activity, and expression during tuber dormancy progression. J Plant Physiol 171:448–457. https://doi.org/10.1016/j.jplph.2013.11.007

Tan M, Li G, Qi S et al (2018) Identification and expression analysis of the *IPT* and *CKX* gene families during axillary bud outgrowth in apple (*Malus domestica* Borkh.). Gene 651:106-117. https://doi.org/10.1016/j.gene.2018.01.101

Topham CM, Srinivasan N, Blundell TL (1997) Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. Protein Engineering 10(1):7-21. https://doi.org/10.1093/protein/10.1.7

Untergasser A, Nijveen H, Rao X et al (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35(2):W71-W74. https://doi.org/10.1093/nar/gkm306

Wang W, Simmonds J, Pan Q et al (2018) Gene editing and mutagenesis reveal inter-cultivar differences and additivity in the contribution of *TaGW2* homoeologues to grain size and weight in wheat. Theor Appl Genet 131(11):2463-2475. https://doi.org/10.1007/s00122-018-3166-7

Wang X, Ding J, Lin S et al (2020a) Evolution and roles of cytokinin genes in angiosperms 2: Do ancient *CKX*s play housekeeping roles while non-ancient *CKX*s play regulatory roles? Hortic Res 7(1):1-15. https://doi.org/10.1038/s41438-020-0246-z

Wang X, Lin S, Liu D et al (2020b) Evolution and roles of cytokinin genes in angiosperms 1: Do ancient *IPT*s play housekeeping while non-ancient *IPT*s play regulatory roles? Hortic Res 7(1):1-15.

https://doi.org/10.1038/s41438-019-0211-x

Wang Y, Liu H Xin Q (2014) Genome-wide analysis and identification of cytokinin oxidase/dehydrogenase (*CKX*) gene family in foxtail millet (*Setaria italica*). Crop J 2:244–254. https://doi.org/10.1016/j.cj.2014.05.001

Werner T, Holst K, Pörs Y et al (2008) Cytokinin deficiency causes distinct changes of sink and source parameters in tobacco shoots and roots. J Exp Bot 59(10):2659-2672. https://doi.org/10.1093/jxb/ern134

Witte CP, Herde M (2020) Nucleotide metabolism in plants. Plant Physiol 182(1):63-78. https://doi.org/10.1104/pp.19.00955

Yeh SY, Chen HW, Ng CY et al (2015) Down-regulation of cytokinin oxidase 2 expression increases tiller number and improves rice yield. Rice 8(1):1-13. https://doi.org/10.1186/s12284-015-0070-5

Zalabák D, Galuszka P, Mrízová K et al (2014) Biochemical characterization of the maize cytokinin dehydrogenase family and cytokinin profiling in developing maize plantlets in relation to the expression of cytokinin dehydrogenase genes. Plant Physiol Biochem 74:283–293. https://doi.org/10.1016/j.plaphy.2013.11.020

Zalewski W, Gasparis S, Boczkowska M et al (2014) Expression patterns of *HvCKX* genes indicate their role in growth and reproductive development of barley. PLoS One 9(12):e115729. https://doi.org/10.1371/journal.pone.0115729

Zhang L, Zhao YL, Gao LF et al (2012) *TaCKX6-D1*, the ortholog of rice *OsCKX2*, is associated with grain weight in hexaploid wheat. New Phytol 195(3):574-584. https://doi.org/10.1111/j.1469-8137.2012.04194.x

Zhang W, Peng K, Cui F et al (2020) Cytokinin oxidase/dehydrogenase *OsCKX11* coordinates source and sink relationship in rice by simultaneous regulation of leaf senescence and grain number. Plant Biotechnol J 1-16 https://doi.org/10.1111/pbi.13467

Zhao J, Bai W, Zeng Q et al (2015) Moderately enhancing cytokinin level by down-regulation of *GhCKX* expression in cotton concurrently increases fiber and seed yield. Mol Breed 35(2):60-73. https://doi.org/10.1007/s11032-015-0232-6

Zheng P, Allen WB, Roesler K et al (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nature Genetics 40(3):367-372. https://doi.org/10.1038/ng.85

Zuñiga-Mayo VM, Baños-Bayardo CR, Díaz-Ramírez D et al (2018) Conserved and novel responses to cytokinin treatments during flower and fruit development in *Brassica napus* and *Arabidopsis thaliana*. Sci Rep 8(1):1-10. https://doi.org/10.1038/s41598-018-25017-3

Table 1. Characteristics of the analysed soybean *CKX* genes and their respective proteins.

Gene name	Accession no.	No. of Exons	Size of gene (bp)	Size of AA sequence	Mw (KDa)[a]	pI[a]	Glycosylation sites[b]
GmCKX03	Glyma03g28910	7	2757	545	61.1	6.71	2
GmCKX04-1	Glyma04g03130	5	6403	451	51	6.47	1
GmCKX04-2	Glyma04g05840	6	3413	494	55.1	5.22	2
GmCKX06	Glyma06g03180	5	7205	528	59.5	6	1
GmCKX09-1	Glyma09g35950	6	2998	534	60.1	6.37	3
GmCKX09-2	Glyma09g07360	6	4028	536	60.8	9.22	2
GmCKX09-3	Glyma09g07190	6	4934	533	60	6.95	6
GmCKX11	Glyma11g20860	5	2997	552	62.3	6.79	6
GmCKX12	Glyma12g01390	5	2908	442	49.2	6.46	1
GmCKX13-1	Glyma13g16430	5	4042	535	61	7.35	1
GmCKX13-2	Glyma13g16420	5	4829	524	58.8	4.95	7
GmCKX14	Glyma14g11280	4	5891	513	57.2	5.81	3
GmCKX15	Glyma15g18560	6	4071	543	61.5	8.73	2
GmCKX17-1	Glyma17g06220	5	4175	535	60.9	6.42	2
GmCKX17-2	Glyma17g06230	6	4598	528	59.5	5.41	6
GmCKX17-3	Glyma17g34330	4	7149	513	57.4	6.16	2
GmCKX19	Glyma19g31620	5	2596	544	60.7	6.12	4

[a] Molecular Weight and Isoelectric Point of the CKX proteins were predicted using the pI/Mw tool in ExPASy (http://web.expasy.org/compute_pi/)

[b] Potential glycosylation sites were predicted using NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc/)

Table 2. *GmCKX* gene family members with natural variation and the cultivars in which the variations occurred. Information associated with SNPs found in *GmCKX14* is bolded (more description in text).

Gene	Variation	Pseudo ΔΔG[a]	Cultivar
GmCKX04-1	T419P	0.60	DH3604, X5076 108B, OAC 06-14
GmCKX04-2	A66V	-1.57	DH3604, X5076-108B
GmCKX06	H16R K77E	-0.63	X5076 68B, X5076 159B
GmCKX09-1	L330F L509Q	-0.34	DH3604, OT 08 13, X5076 159B
GmCKX14	**T69A** **H105Q** **A159G**	**-0.75**	**OAC Wallace, IA1010LF, DH420, DH748**

[a] ΔΔG value as per SDM classifies a mutation as slightly destabilizing if less than -0.5; slightly stabilizing between 0.5-1.0; and destabilizing if less than -1.0

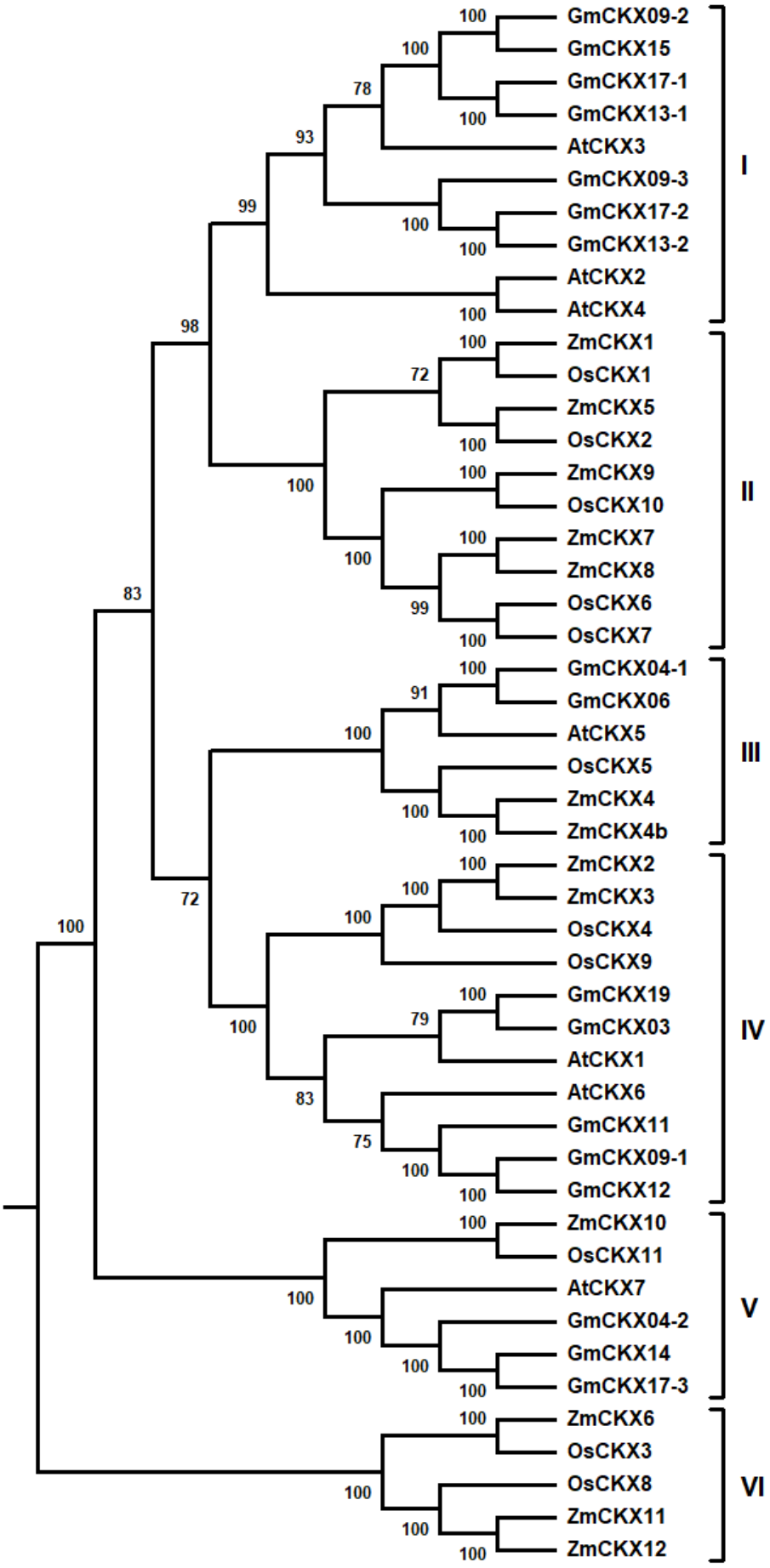
Figure Captions

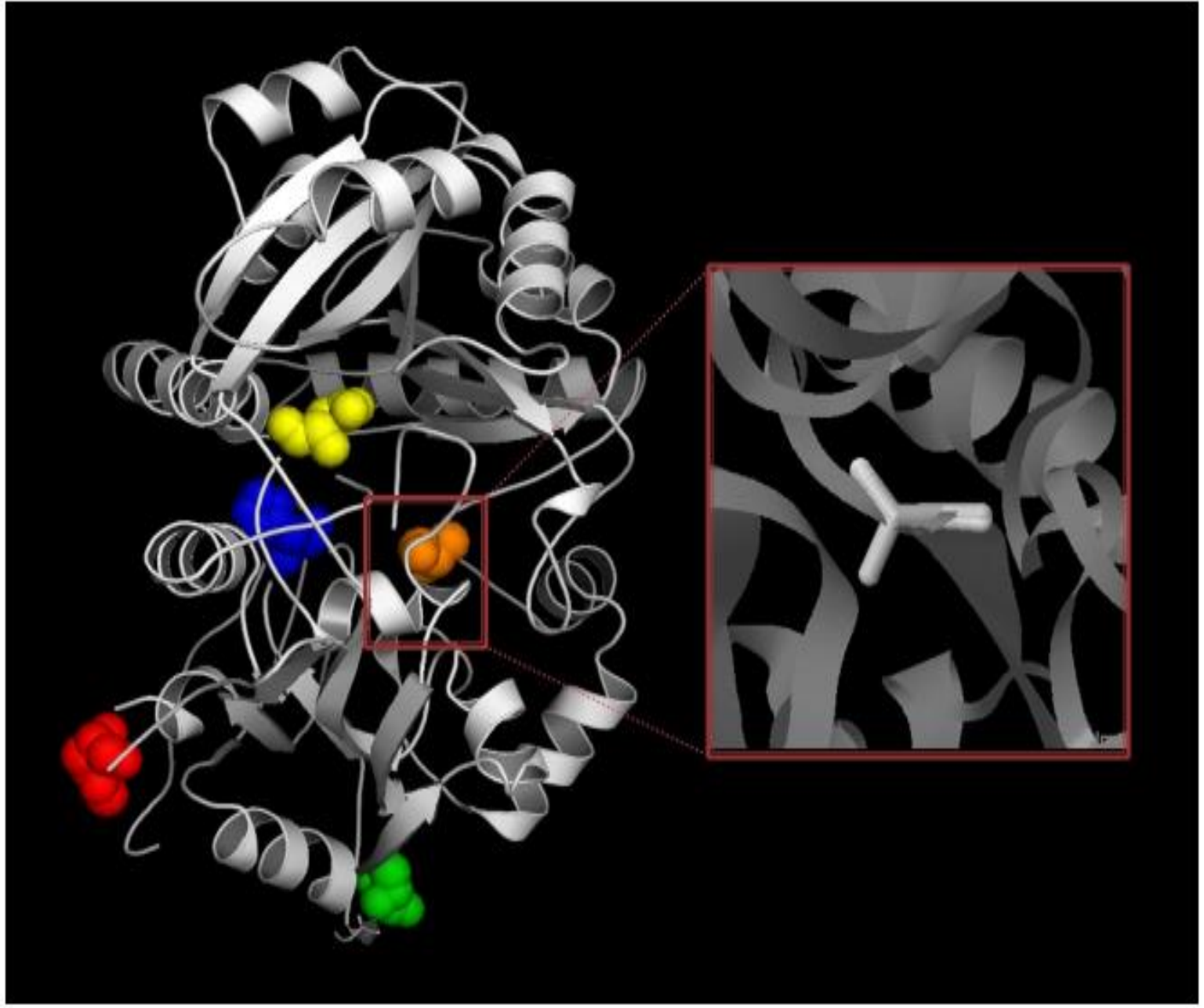
Fig. 1 Phylogenetic analysis of the CKX gene family members of two monocot species, rice (*Oryza sativa*), maize (*Zea mays*), and two dicots, Arabidopsis (*Arabidopsis thaliana*) and soybean (*Glycine max*). The tree was generated based on the protein sequences, upon alignment with ClustalW, using the neighbor-joining method. The CKX proteins were assigned into 6 groups of which four (I, III, IV and V) contained GmCKX proteins and the other two (II and VI) were exclusive to monocots

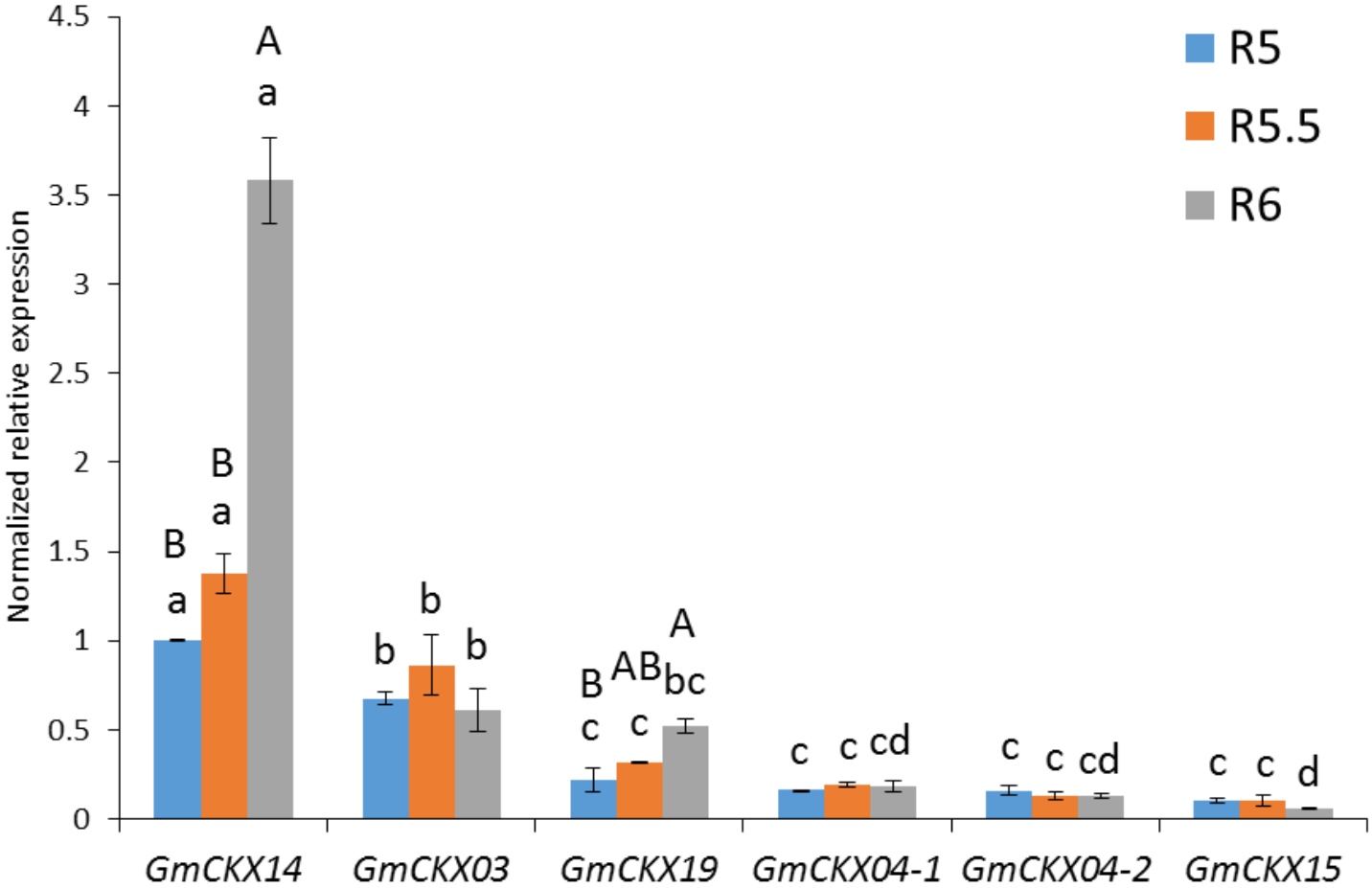
Fig. 2 Homology model of GmCKX14 showing the predicted residue for substrate binding in yellow and predicted residue for FAD binding in blue. The mutant residue position A159G is represented in orange, T69A in red and H105Q in green. The residue alanine at position 159 substituted by glycine shown in an expanded view (image developed using SDM software)

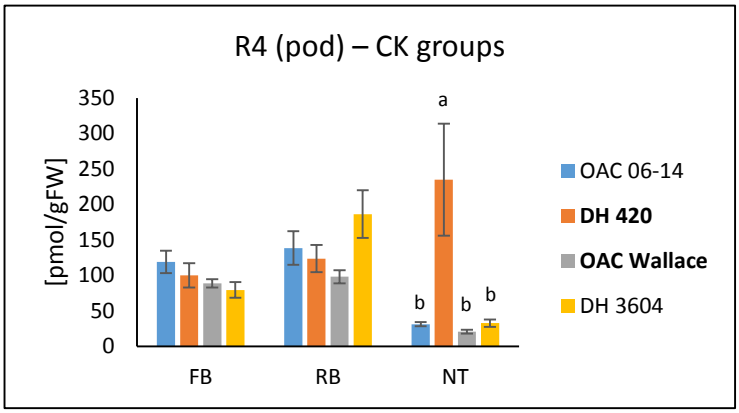
Fig. 3 Quantitative expression profiles of 6 *GmCKX* genes analysed at 1 soybean developmental stages: R4 (pods), R5.5 (seeds) and R6 (seeds) in OAC Wallace. The housekeeping gene *GmGAPDH* was used as an endogenous control for normalization of expression analysis. The relative transcript level is presented using 2 delta CT method (qRT-PCR). Error bars represent standard error of the mean (n=3). Small letters denote significant differences among the genes at each of the growth stages and capital letters denote significant differences among the three stages for particular gene (ANOVA, Duncan's multiple range test, $p<0.05$)

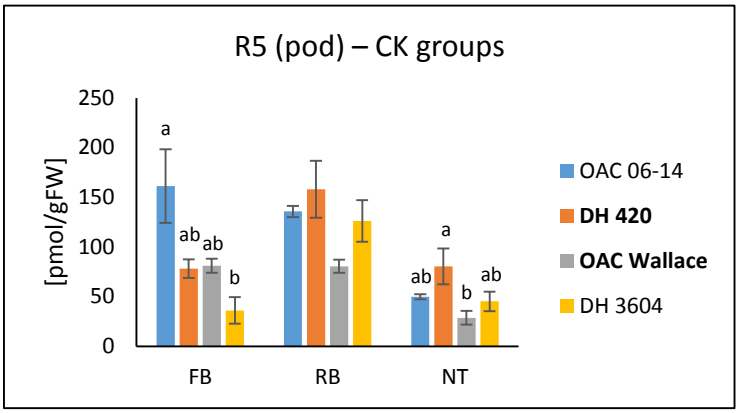
Fig. 4 CK profile data (groups: FB, RB and NT, and types within these groups: tZ-type, cZ-type, DZ-type and iP-type) analysed using HPLC-MS/MS at 4 developmental stages in 4 soybean cultivars. Highlighted bold are the cultivars with SNPs in *GMCKX14* detected. Letters denote significant differences among the cultivars (ANOVA, Duncan's multiple range test, $p<0.05$)

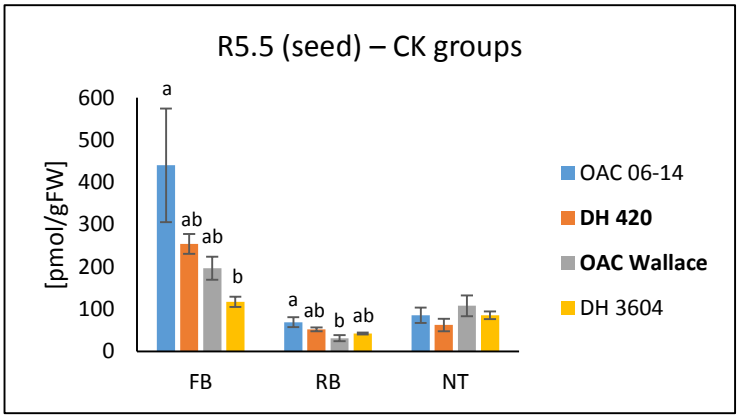

GmCKX09-2
GmCKX15
GmCKX17-1
GmCKX13-1
AtCKX3
GmCKX09-3
GmCKX17-2
GmCKX13-2
AtCKX2
AtCKX4
ZmCKX1
OsCKX1
ZmCKX5
OsCKX2
ZmCKX9
OsCKX10
ZmCKX7
ZmCKX8
OsCKX6
OsCKX7
GmCKX04-1
GmCKX06
AtCKX5
OsCKX5
ZmCKX4
ZmCKX4b
ZmCKX2
ZmCKX3
OsCKX4
OsCKX9
GmCKX19
GmCKX03
AtCKX1
AtCKX6
GmCKX11
GmCKX09-1
GmCKX12
ZmCKX10
OsCKX11
AtCKX7
GmCKX04-2
GmCKX14
GmCKX17-3
ZmCKX6
OsCKX3
OsCKX8
ZmCKX11
ZmCKX12
I
II
III
IV
V
VI

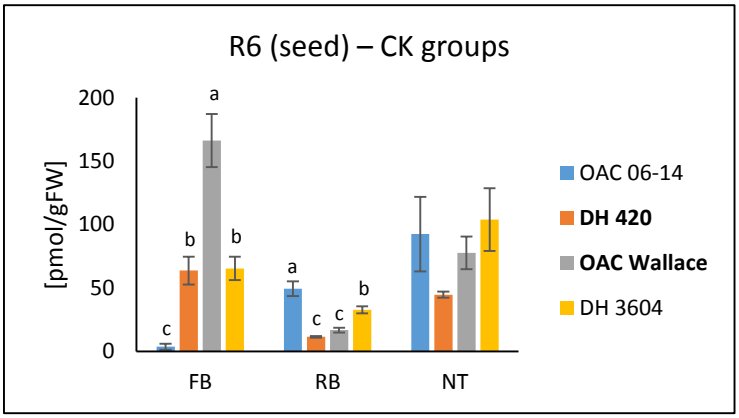


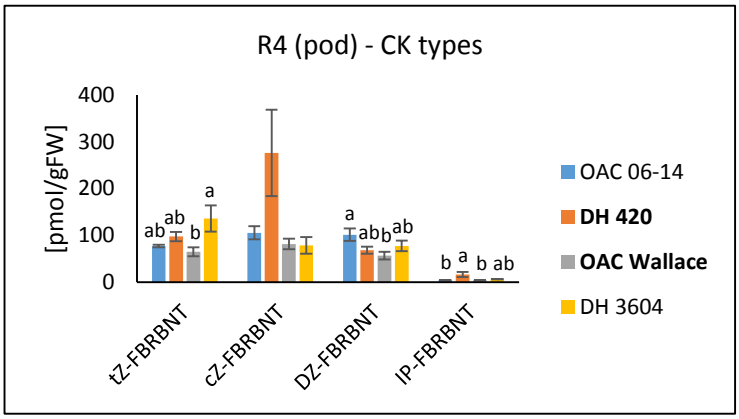

Normalized relative expression
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
R5
R5.5
R6
A
a
B
a
B
a
b
b
b
B
c
AB
c
A
bc
c
c
cd
c
c
cd
c
c
d
GmCKX14
GmCKX03
GmCKX19
GmCKX04-1
GmCKX04-2
GmCKX15

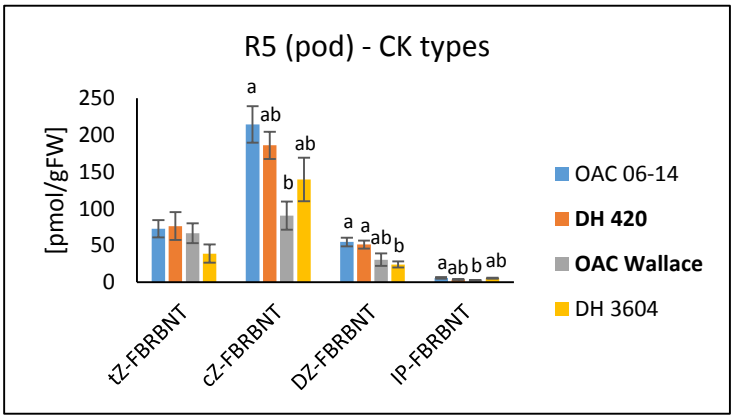

R4 (pod) – CK groups
[pmol/gFW]
350
300
250
200
150
100
50
0
FB
RB
NT
a
b
b
b
OAC 06-14
DH 420
OAC Wallace
DH 3604

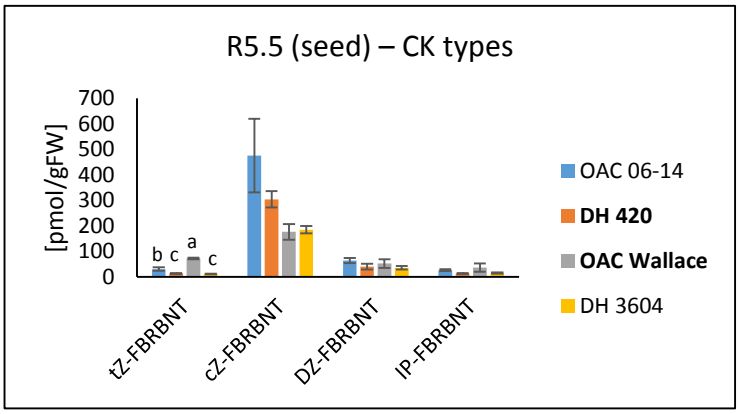

R5 (pod) – CK groups
[pmol/gFW]
250
200
150
100
50
0
FB
RB
NT
a
ab
ab
b
ab
a
b
ab
OAC 06-14
DH 420
OAC Wallace
DH 3604

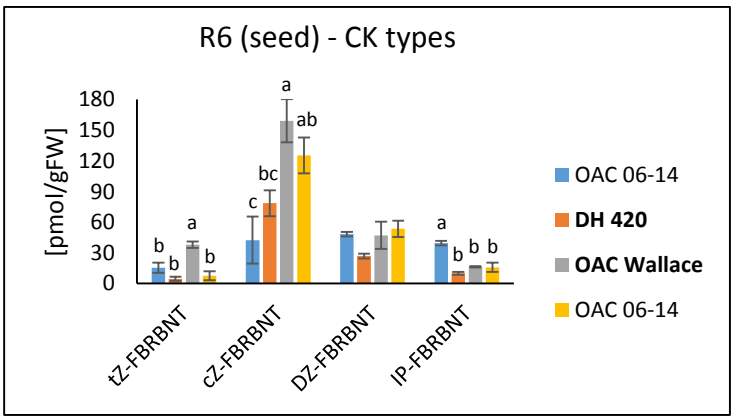

R5.5 (seed) – CK groups
[pmol/gFW]
600
500
400
300
200
100
0
FB
RB
NT
a
ab
ab
b
a
ab
b
ab
OAC 06-14
DH 420
OAC Wallace
DH 3604


R6 (seed) – CK groups
[pmol/gFW]
200
150
100
50
0
FB
RB
NT
c
b
a
b
a
c
c
b
OAC 06-14
DH 420
OAC Wallace
DH 3604


R4 (pod) - CK types
[pmol/gFW]
400
300
200
100
0
tZ-FBRBNT
cZ-FBRBNT
DZ-FBRBNT
IP-FBRBNT
ab
ab
b
a
a
ab
b
ab
b
a
b
ab
OAC 06-14
DH 420
OAC Wallace
DH 3604


R5 (pod) - CK types
[pmol/gFW]
250
200
150
100
50
0
tZ-FBRBNT
cZ-FBRBNT
DZ-FBRBNT
IP-FBRBNT
a
ab
b
ab
a
a
ab
b
ab
b
ab
OAC 06-14
DH 420
OAC Wallace
DH 3604


R5.5 (seed) – CK types
[pmol/gFW]
700
600
500
400
300
200
100
0
tZ-FBRBNT
cZ-FBRBNT
DZ-FBRBNT
IP-FBRBNT
b
c
a
c
OAC 06-14
DH 420
OAC Wallace
DH 3604


R6 (seed) - CK types
[pmol/gFW]
180
150
120
90
60
30
0
tZ-FBRBNT
cZ-FBRBNT
DZ-FBRBNT
IP-FBRBNT
b
b
a
b
c
bc
a
ab
a
b
b
b
OAC 06-14
DH 420
OAC Wallace
OAC 06-14