2S albumin g13 polypeptide, less related to Fag e 2, can be eliminated in common buckwheat (Fagopyrum esculentum Moench) seeds

AUTHOR(S):

Monshi, Fakhrul Islam; Katsube-Tanaka, Tomoyuki

CITATION:

Monshi, Fakhrul Islam ...[et al]. 2S albumin g13 polypeptide, less related to Fag e 2, can be eliminated in common buckwheat (Fagopyrum esculentum Moench) seeds. Food Chemistry: Molecular Sciences 2022, 5: 100138.

ISSUE DATE:

2022-12-30

URL:

http://hdl.handle.net/2433/284586

Food Chemistry: Molecular Sciences 5 (2022) 100138

Contents lists available at ScienceDirect

Food Chemistry: Molecular Sciences

journal homepage: www.sciencedirect.com/journal/food-chemistry-molecular-sciences

2S albumin g13 polypeptide, less related to Fag e 2, can be eliminated in common buckwheat (*Fagopyrum esculentum* Moench) seeds

Fakhrul Islam Monshi [a,b], Tomoyuki Katsube-Tanaka [a,*]

[a] *Graduate School of Agriculture, Kyoto University, Kitashirakawa, Kyoto 606-8502, Japan*
[b] *Department of Genetics and Plant Breeding, Sylhet Agricultural University, Sylhet-3100, Bangladesh*

ARTICLE INFO

Keywords:
2S albumin
Allergen
Common buckwheat
g13
Null allele

ABSTRACT

2S albumin (g11, g13, g14, and g28) is an important allergen in common buckwheat (*Fagopyrum esculentum*). g13 is hydrophobic, rare in seeds, and may show distinct allergenicity from the others; therefore, we tried to eliminate this protein. Phylogenetic and property distance analyses indicated g13 is less related to g14 (Fag e 2) than g11/g28 is related to g14, particularly in the second domain containing the II and III α-helices. A null allele with a 531 bp insertion in the coding region was found for *g13* at an allele frequency of 2 % in natural populations of common buckwheat. The *g13_null* allele homozygote accumulated no g13 protein. A BLAST search for the 531 bp insertion suggested the insert-like sequence resided frequently in the buckwheat genome, including the self-incompatibility responsible gene *ELF3* in *Fagopyrum tataricum*. The *g13_null* insert-like sequence could, therefore, help in producing hypoallergenic cultivars, and expand the genetic diversity of buckwheat.

1. Introduction

Buckwheat is increasing in popularity and is nutritionally valuable as well as consumption being associated with health benefits (Zhang et al., 2012). Despite the benefits of buckwheat seed, its allergenic properties have raised concerns about public health and safety with increased use (Heffler et al., 2014). Buckwheat seeds contain highly potent allergens that trigger immediate-type anaphylactic reactions through an IgE-mediated response in susceptible patients (Park et al., 2016). Thus, there is a need to eliminate or mitigate its allergenicity, as has been performed with certain proteins found in rice (Teramura et al., 2019) and soybean (Sugano et al., 2020), for the further utilization of buckwheat as a food crop.

Common buckwheat allergens are categorized as Fag e 1 (13S globulin), Fag e 2 (2S albumin), Fag e 3 (7S globulin/vicilin), Fag e 4 (antimicrobial peptide), Fag e 5 (vicilin-like protein), Fag e 10 kD (α-amylase inhibitor/trypsin inhibitor), and Fag e TI (trypsin inhibitor). Among these, 13S globulin and 2S albumin, the most and second-most abundant storage proteins in seeds, respectively, have been reported as the main allergens (Katayama et al., 2018).

Khan et al. (2012) identified a different number (0–6) of tandem repeats inserted in α polypeptides of the methionine-poor subunit of 13S globulin. The tandem repeat insertion, composed of a 15 residue unit leads to molecular size variation and increased trypsin digestibility due to its hydrophilic and arginine-rich nature. Using a genomic DNA library, Sano et al. (2014) detected 17 open reading frames (ORFs) encoding 13S globulin, including trypsin-resistant zero-repeat subunits (GlbNA and GlbNB). In a natural common buckwheat population, Katsube-Tanaka et al. (2014) and Monshi et al. (2020) identified six other new alleles for the zero-repeat subunit, including three which had an approximately 200-bp-long miniature inverted-repeat transposable element (MITE)-like sequence. Thus, the tandem repeat and MITE-like sequence introduce large variations in the structure and function of the 13S globulin gene, suggesting that a natural population of common buckwheat shows extensive genetic diversity.

An 8–16 kDa allergenic protein, 2S albumin, which is resistant to pepsin/trypsin, can cause anaphylaxis in buckwheat-sensitive patients (Tanaka et al., 2002). 2S albumin proteins with a distinct molecular weight are also characterized as main allergens in different crops, such as peanut (Ara h 6, Ara h 7 as 15 kDa, and Ara h 2 as 17 kDa), English walnut (Jug r 1, 15–16 kDa), cashew nut (Ana o 3, 14 kDa), pistachio (Pis v 1, 7 kDa), hazelnut (Cor a 14, 10 kDa), and Brazil nut (Ber e1, 9 kDa) (reviewed by Luparelli et al., 2022). Although a substantial amount of basic research, including epitope determination of Fag e 2, has been undertaken (Tanaka et al., 2002), agronomical studies aimed at breeding and production of hypoallergenic buckwheat have been

* Corresponding author at: Graduate School of Agriculture, Kyoto University, Kitashirakawa, Kyoto 606-8502, Japan.
E-mail address: tanakato@kais.kyoto-u.ac.jp (T. Katsube-Tanaka).

https://doi.org/10.1016/j.fochms.2022.100138
Received 14 August 2022; Received in revised form 22 September 2022; Accepted 24 September 2022
Available online 26 September 2022

limited because of a lack of fundamental knowledge of the gene composition, gene copy number, and allelic variations of the 2S albumin gene.

In the first draft genome assembly of common buckwheat developed in 2016, five 2S albumin genes (*g03, g11, g13, g14,* and *g28* abbreviated from *Fes_sc0007211.1.g000003.aua.1*, *Fes_sc0000087.1.g000011.aua.1*, *Fes_sc0000087.1.g000013.aua.1*, *Fes_sc0000087.1.g000014.aua.1,* and *Fes_sc0000087.1.g000028.aua.1*, respectively) were identified (Yasui et al., 2016). In our previous study, we found that *g03* was a possibly pseudo gene owing to 8 and 622 bp deletions in the coding region. Characterization of the other 2S albumin polypeptides (g11, g28, g13, and g14) prepared in recombinant or native forms revealed differences in accumulation levels within seeds and post-translational processing, and the hydrophobic nature between them (Katsube-Tanaka & Monshi, 2022). The results showed that g13 is considerably hydrophobic, scarce in seeds, and may show distinct allergenicity, but it is still unexplored compared to other polypeptides. Therefore, the present study was conducted to further characterize 2S albumin polypeptide g13 and eliminate g13 for the development of hypoallergenic buckwheat. In this study, we hypothesised that the outcrossing nature of common buckwheat, which could help in the conservation of extensive genetic diversity in natural common buckwheat populations, was considered an advantageous characteristic to provide useful alleles.

2. Materials and methods

2.1. Plant materials and preparation of genomic DNA

Four Japanese cultivars of common buckwheat, Harunoibuki, Shinano 1, Botansoba, and Miyazaki Wase Kaori, and four landraces from Hokkaido (Japan), Pakistan, Nepal, and Bangladesh were used as plant materials for the preparation of gDNA. A large number of seeds (>15 g) from all cultivars and landraces were independently ground to produce buckwheat flour to ensure extensive genetic diversity. Genomic DNA was extracted from 4 g of the seed flour (equivalent to >400 grains) from each of the eight cultivars and landraces using the DNAs-ici!-S method (Rizo, Tsukuba, Japan). Individual seed gDNA from the Harunoibuki cultivar was also extracted.

2.2. Isolation and identification of g13_null allele

The coding region of *g13* was amplified by PCR with the forward primer (2SA_g13_N:5′-AGAACTTTCGTTGAAGGAGAAAACG-3′) and reverse primer (TAG_2SA_g13_Bam:5′-gcggatccCTAGTAAGCTATGATCCTATTAC-3′). PCR conditions were as follows: initial denaturing step at 94 °C for 2 min, 30 cycles of denaturing at 98 °C for 10 s, annealing at 60 °C for 30 s, extension at 68 °C for 1 min, and a final extension at 68 °C for 7 min with KOD FX Neo DNA polymerase (TOYOBO, Osaka, Japan). The amplified fragment, with a molecular size 531 bp larger than that of authentic *g13,* was isolated from agarose gel and purified using the FastGene Gel/PCR Extraction Kit (Nippon Genetics, Tokyo, Japan). The purified fragment was cloned into the pTAC-2 vector using a TA cloning kit (BioDynamics Laboratory Inc., Tokyo, Japan). The nucleotide sequence of the 531 bp larger fragment was verified by Sanger sequencing. PCR primers were purchased from Eurofins Genomics ltd. (Tokyo, Japan).

2.3. Estimation of allele frequency for g13_null allele and development of g13_null allele homozygote

The Japanese indigenous cultivar Harunoibuki was used as plant material to examine the frequency of the null allele. Twenty-three seeds were grown in the soil-filled plastic pots. Three- to four-week-old leaves were collected from each plant for genomic DNA isolation. Approximately 100 mg of leaf sample from each plant was ground with Mixer-Mill (QIAGEN K.K., Tokyo, Japan) and used for genomic DNA extraction using DNAs-ici!-P (Rizo, Tsukuba, Japan). PCR with coding region primers for *g13* was conducted using the aforementioned standardized thermal cycler conditions. The PCR products were electrophoresed on a 1 % agarose gel containing ethidium bromide and the gel was visualized and photographed under a UV transilluminator. Subsequently, 415 individual plants of the cultivar Shinano 1 were screened using coding region primers to identify *g13_null* allele-containing plants. Plants containing *g13_null* allele were then grown in the isolated environment of the phytotron to develop the *g13_null* allele homozygote and maintain genetic purity.

2.4. Reverse-transcription PCR

For reverse transcription-polymerase chain reaction (RT-PCR) analysis, first-strand cDNA was used as a template, which was synthesized from 1.0 mg of total RNA isolated from the young buckwheat endosperm (white dough stage) using a total RNA isolation kit (RNAs-ici!-R, Rizo, Tsukuba, Japan) and DNaseI (NIPPON GENE, Tokyo, Japan). The RNA was primed with oligo-dT adaptor primers and reverse-transcribed in a total volume of 10 μL, followed by PCR using a TaKaRa RNA PCR Kit (AMV) Ver.3.0 (Takara Bio Inc., Kusatsu, Japan). The above-mentioned coding region primers for *g13,* along with *g14* primers (2SA_g14_N:5'-ACGTACCCGAGAGATGAAGGCTTC-3' and TAA_2SA_g14_Bam:5'-gcggatccTTACACAAAATACCGATTTCCTC-3') and actin (BW actin Self3 Fwd: 5'-GGCATCACACTTTCTACAATGAGC-3' and BW actin Self3 Rev: 5'-GATATCCACATCACATTTCATGAT-3') were used. The thermal cycling program was set to 95 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s, and extension for 7 min at 68 °C using Ex Taq DNA polymerase (Takara Bio Inc., Kusatsu, Japan). At least three biological replicates were examined using biologically independent seeds with single technical replicate.

2.5. SDS-PAGE and western blot analyses

SDS-PAGE was conducted using 17 % T acrylamide gel at a constant voltage of 200 V for 70 min. Immunoblotting was performed according to a previous report (Khan et al., 2012). Briefly, proteins were separated by SDS-PAGE, and western blotting was performed when the resolved proteins were transferred electrophoretically to a nitrocellulose membrane (Cytiva, Uppsala, Sweden). The membrane was then incubated in a blocking solution (0.15 M NaCl, 5 % (w/v) skimmed milk, 0.05 % (v/v) Tween-20, and 0.04 M sodium-phosphate buffer pH 7.4) for 60 min with gentle shaking at room temperature. The membrane was first probed with primary polyclonal antibodies against g13 and g14, which were produced in rabbits against the partially purified recombinant g13 (Eurofins Genomics ltd., Tokyo, Japan) and against the peptide (CRIGERLIKEGVRDLKE) (Sigma-Aldrich Japan GK., Tokyo, Japan), respectively (Katsube-Tanaka & Monshi, 2022) at room temperature for 60 min. After thoroughly washing three times for 20 min each with TBST (0.02 M Tris pH 7.5, 0.9 % (w/v) NaCl, 0.05 % (v/v) Tween-20), the membrane was incubated with secondary goat anti-rabbit IgG alkaline phosphatase conjugate (Promega, Madison, USA) for 60 min at room temperature. The membrane was then washed three times with TBST for 20 min. After antibody recognition, the membrane was incubated with BCIP (5-bromo-4-chloro-3-indolyl-phosphate) and NBT (nitroblue tetrazolium) for color development. Before scanning the air-dried membrane, detection was stopped using deionized water. Western blotting was repeated at least three times with biologically independent seeds with single technical replicate. All chemical reagents, unless otherwise mentioned, were purchased from NACALAI TESQUE, Inc. (Kyoto, Japan).

2.6. Nucleotide and amino acid sequence analyses

Property distance (E1–E5 Property-Based Peptide Similarity Index, PD) was calculated for all combinations of eight amino acid residues

from g11 or g13 and g14 according to Venkatarajan and Braun (2001) and the Structural Database of Allergenic Proteins (SDAP, https://fermi.utmb.edu/SDAP/index.html). Phylogenetic tree analysis was performed with MEGA X (Kumar et al., 2018) using sequences retrieved from the ALLERGEN NOMENCLATURE database (https://allergen.org/index.php), by limiting the search to food allergens and 2S albumin, as well as from sequences identified in the Buckwheat Genome DataBase (BGDB, *Fagopyrum esculentum*) and Molecular Breeding Knowledgebase (MBKbase, *Fagopyrum tataricum*). A BLASTN search for *g13_null* insert sequence (531 letters) was conducted in the BGDB and with the local BLAST program BioEdit (Hall, 1999) using all chromosomal sequences from the MBKbase. BLASTN (discontiguous megablast) was also conducted with the NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

3. Results

3.1. Similarity between g11, g13, g14, and g28

Phylogenetic tree analysis demonstrated that the 2S albumin polypeptides of buckwheat were near to those of *Pinus pinea* and distant from those of *Ricinus communis* and *Sesamum indicum* (Fig. 1). Buckwheat g11 resembles g28 and is more similar than g13 to g14. Within each polypeptide (g11/g28, g13, and g14), the amino acid sequences were similar between *F. esculentum* and *F. tataricum*. Analyses of property distance (PD), whereby a low value indicates resemblance between two peptides, showed a clearer diagonal line between g11 and g14 than that between g13 and g14 (Fig. 2A). The diagonal lines between g11 and g13, and between Cuc ma 5 (winter squash 2S albumin) and g13 were unclear (Fig. S1). The diagonal line between g11 and g14 was divided into three regions, corresponding to the three domains of 2S albumin. From more detailed analyses, the average PD values ranged from 16 to 20 for g11 against g14 and from 17 to 20 for g13 against g14 (Fig. 2B). However, the minimum values of PD ranged from 1 to 14 for g11 against g14, and from 5 to 13 for g13 against g14, implying that g11 is more related to g14 than g13 is related to g14. In particular, the minimum values for g11 against g14 became very low (<6) within the start positions of eight amino acid residues in g14 at 57 and 86, corresponding to the second domain containing the II and III α-helices (Moreno & Clemente, 2008). Note that in the third domain containing the epitope of Fag e 2, the minimum values were lower for g11 against g14 than for g13 against g14. These results imply that g11 is more similar to g14 than g13 is similar to g14.

3.2. Variation in the coding region of g13 and the allele frequency of g13_null

PCR amplification of the coding region of *g13* produced an additional large band along with a 443 bp authentic band for all genotypes and landraces except Botansoba (Fig. 3A). The amplified fragment with a greater molecular weight was isolated and cloned into the pTAC-2 vector. Sequencing of the large band revealed that the extra sequence (531 bp) was inserted into the coding region (Fig. 3B), which produced a

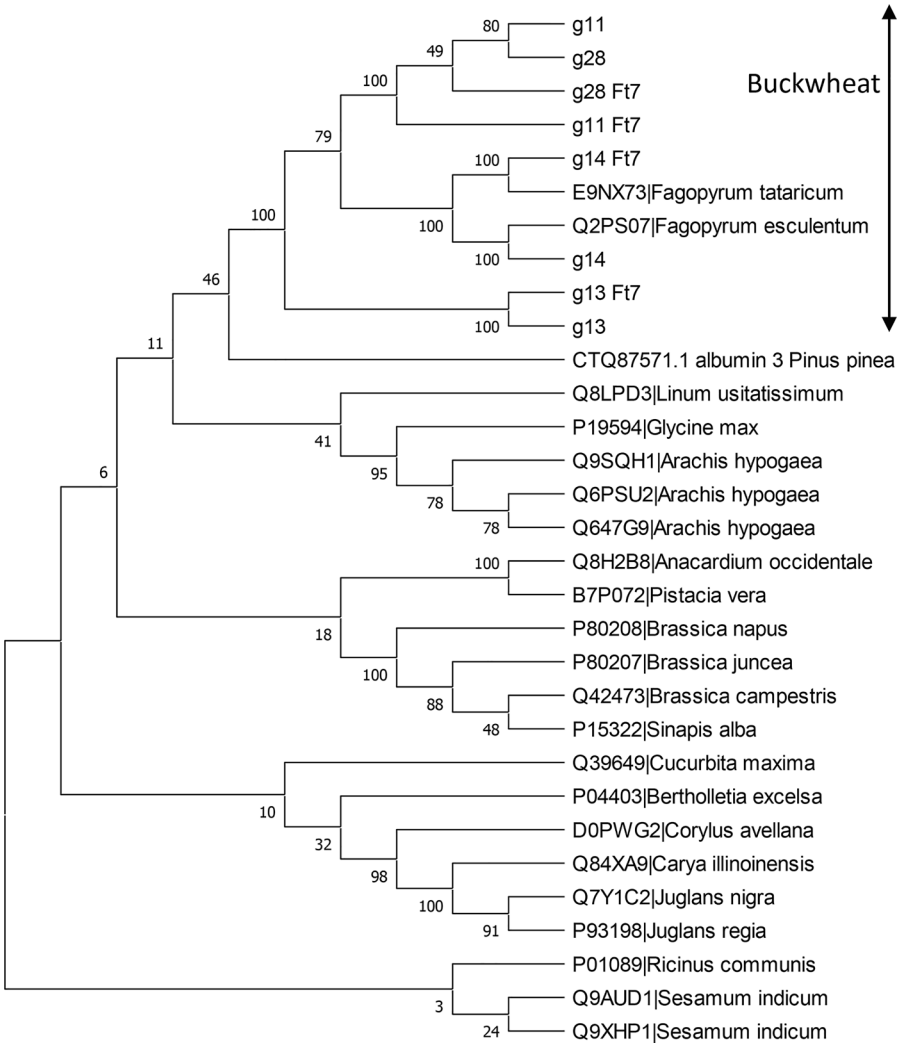

Fig. 1. Phylogenetic tree for 2S albumins of buckwheat and its homologues. The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analyzed. Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. This analysis involved 31 amino acid sequences. The sequences were retrieved from ALLERGEN NOMENCLATURE (http://allergen.org/index.php) database limiting search to food allergens and with a query of 2S albumin. 2S albumin genes from Buckwheat Genome DataBase (*Fagopyrum esculentum*, http://buckwheat.kazusa.or.jp/) and Molecular Breeding Knowledgebase (*Fagopyrum tataricum*, Ft7, http://mbkbase.org/). All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 332 positions in the final dataset. Evolutionary analyses were conducted in MEGA X (Kumar et al., 2018).

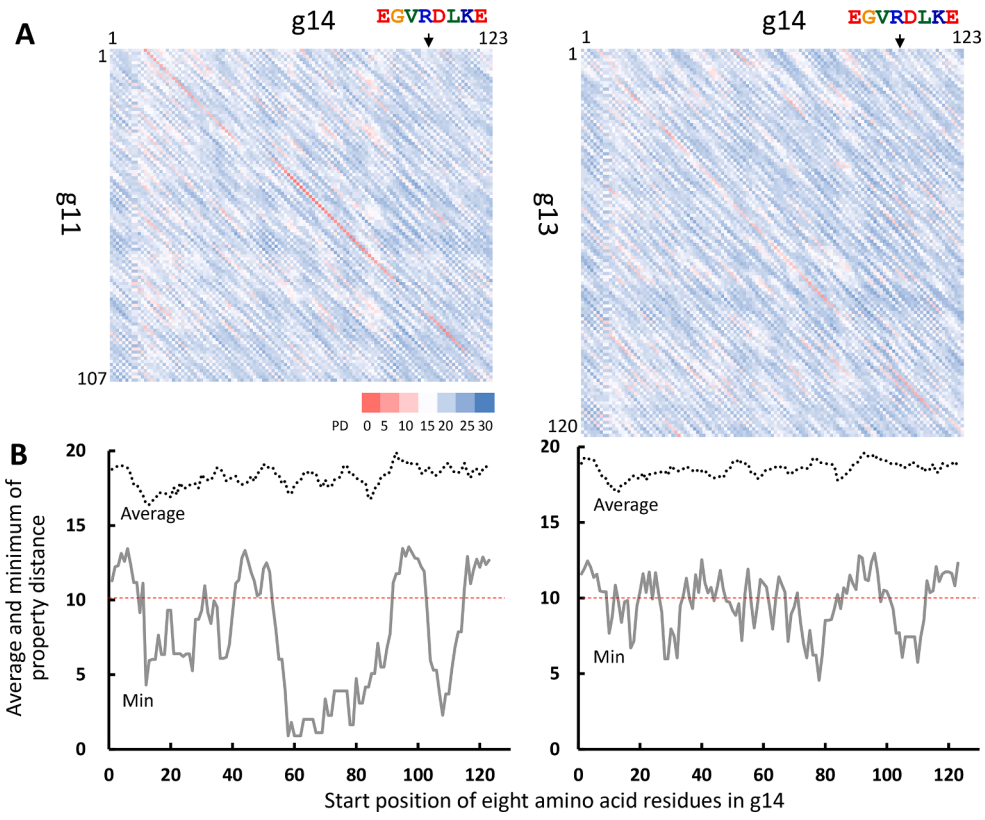

Fig. 2. Property distance for all combinations of eight amino acid residues between g11, g13 and g14. (A) Heatmap of property distance (PD). PD was calculated for all combinations of eight amino acid residues from g11 (left, vertical) or g13 (right, vertical) and g14 (horizontal). The numerical values at the left and top of figures indicate the start position of eight amino acid residues. The position of Fag e 2 epitope (EGVRDLKE) is shown by arrows. (B) Average and minimum of PD were calculated at each start position of eight amino acid residues in g14 (left, g11; right, g13).

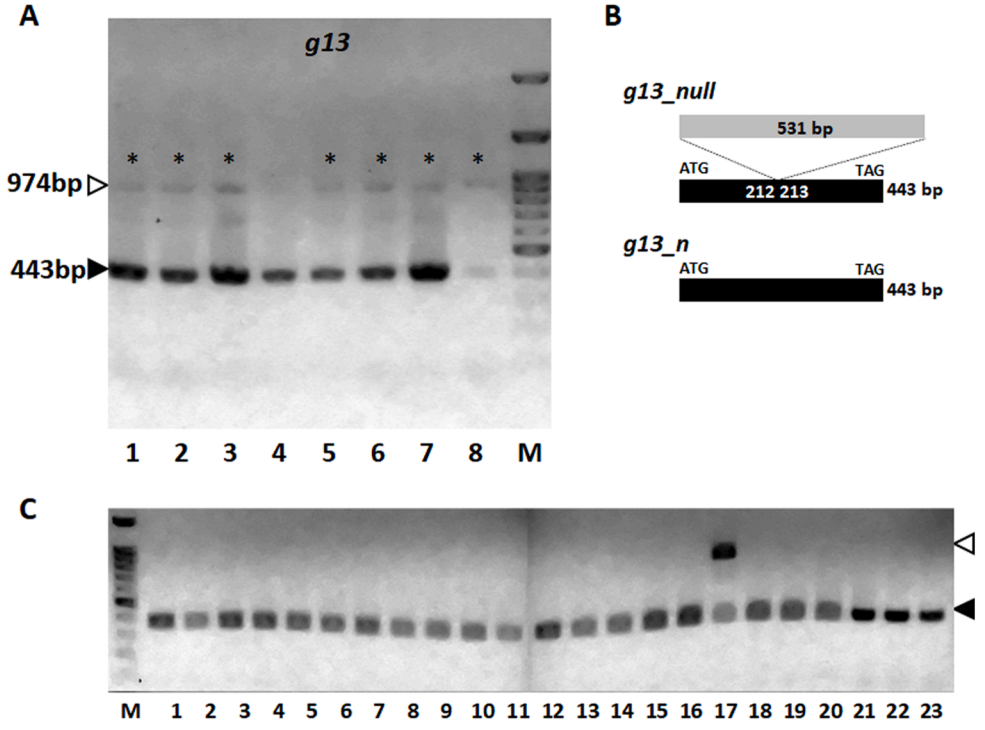

Fig. 3. Amplification and gene structure of *g13* and *g13_null*. (A) PCR amplification of *g13* coding regions. Lanes 1 and 6 to 8 are landraces of Pakistan, Hokkaido (Japan), Nepal, and Bangladesh, respectively and lanes 2 to 5 are cultivars of Harunoibuki, Shinano 1, Botansoba, and Miyazaki-wase-kaori, respectively. Asterisks indicate a band derived from null alleles. (B) Gene structures of *g13* normal allele (*g13_n*) and *g13* null allele (*g13_null*). (C) Allele frequency of *g13_null*. Twenty-three plants of cultivar Harunoibuki were genotyped for *g13_null*. Black and white arrowheads indicate PCR fragments derived from the normal and *g13_null* allele, respectively. M indicates a size marker.

translation termination within the insert, suggesting that the large band (974 bp) was derived from the *g13* null allele (hereafter, *g13_null*) (Fig. 4).

The Harunoibuki cultivar was examined for the allele frequency of *g13_null*. A total of 23 individual plants were analyzed, and all plants produced the 443 bp authentic band for *g13*, while only one produced the 974 bp additional band (Fig. 3C). Therefore, the results suggest that the allele frequency of *g13_null* was relatively high at approximately 2 %. We screened a vast number (415) of Shinano 1 plants and identified 17 carrying *g13_null* (data not shown), indicating that the feasible allele frequency of *g13_null* would be 2.0 %. The *g13_null*-containing plants were placed in an isolated area, where genetic purity was strictly

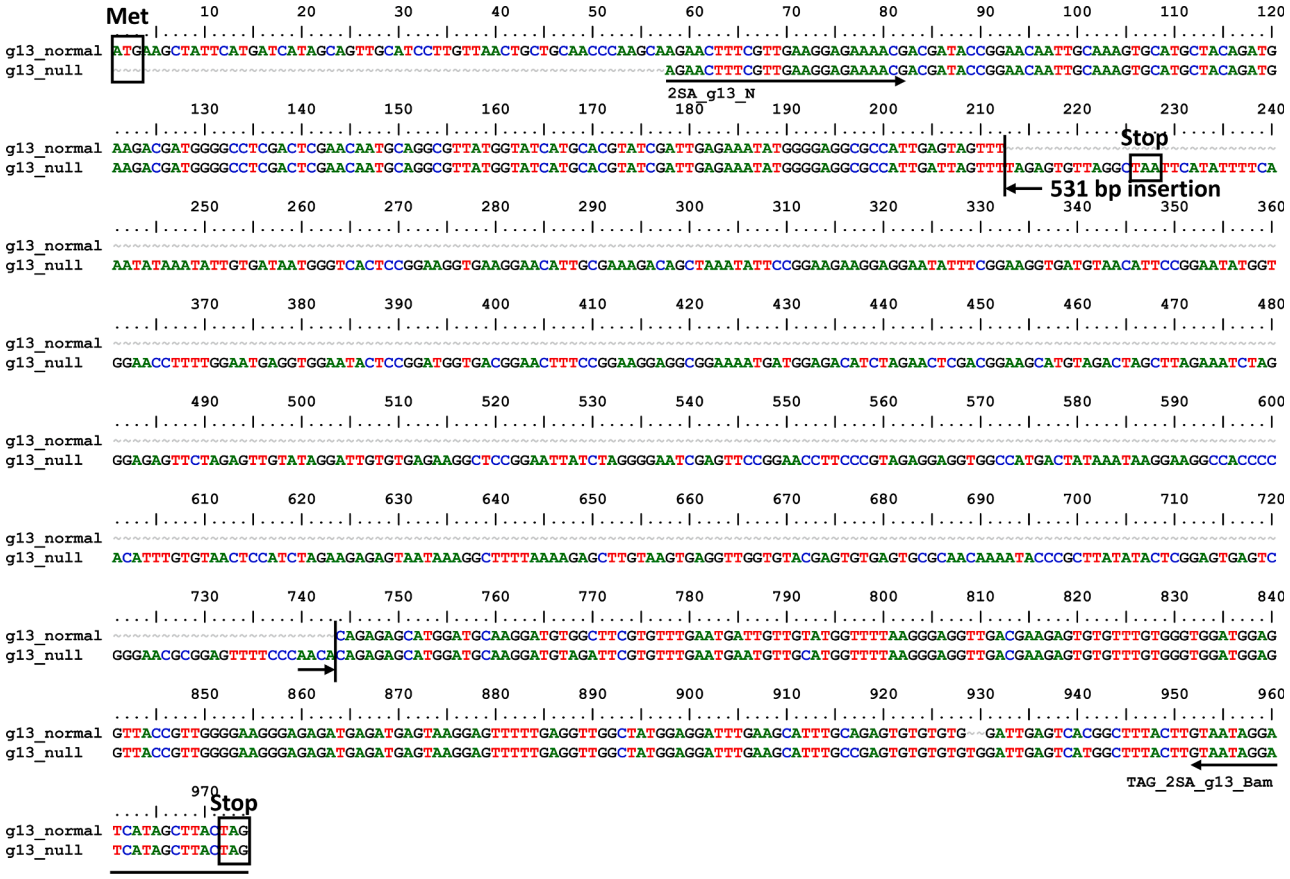

Fig. 4. Alignment of *g13* normal and *g13* null alleles. The coding region of *g13* normal allele (g13_normal) and *g13* null allele (g13_null) were compared. Translation initiation (ATG) and predicted termination (TAG) were highlighted with a black box. The position of 531 bp insertion was also indicated. Primers for genotyping were shown by horizontal arrows with the primer names. The first 57 bp of g13 null allele was not shown.

maintained in subsequent generations to obtain homozygous plants for further analysis.

3.3. Development of g13_null homozygotes and expression of g13 and g14 genes

In subsequent generations, *g13_null* homozygous plants were identified (Fig. 5A). The RT-PCR results showed no transcript of the *g13* gene, whereas wild-type Harunoibuki produced transcripts of *g13* (Fig. 5B). Both wild-type and *g13_null* homozygous plants produced transcripts of *g14* and actin, and that of *g14* was increased in *g13_null* homozygous plants (Fig. 5B). The accumulation level of g13 detected by immunoreaction with anti-g13 antibody indicated that g13 was detected only in the wild-type Harunoibuki along with recombinant g13, while anti-g14 antibody reacted with all samples except recombinant g13 (Fig. 5C). Thus, the lack of accumulation of g13 in the *g13_null* homozygous plants was not compensated for by homologous g14, suggesting that the *g13_null* property can be useful in the development of hypoallergenic buckwheat.

3.4. BLAST search of g13_null insert in BGDB, MBKBase, and NCBI

A Basic Local Alignment Search Tool (BLAST) program was employed in the Buckwheat Genome Database (BGDB) for *Fagopyrum esculentum* as well as the Molecular Breeding Knowledge Base (MBKbase) for *Fagopyrum tataricum* to detect similarities in the nucleotide sequences of the *g13_null* insert (Fig. 6A). The *g13_null* insert-like sequence with a score >44 was found to be fewer in number in BGDB (218) than in MBKBase (1,076). However, the *g13_null* insert-like sequence with a high score >300 was detected more frequently in the BGDB (113) than in the MBKBase (56). The highest BLAST hit scores for *g13_null* insert were 785 and 361 in the BGDB and MBKBase, respectively.

4. Discussion

4.1. Novel characteristics of g13

The property distance (PD) index has been demonstrated to be a good predictor of epitope cross-reactivity or peptide structure-IgE binding relationships (Ivanciuc et al., 2009). The PD index was calculated using five amino acid descriptors based on the multidimensional scaling of 237 physicochemical properties of amino acids. These five descriptors correlate with hydrophobicity, size, amino acid preference to occur in α-helices, degenerate triplet codon number, and occurrence frequency of amino acid residues in β-strands (Venkatarajan & Braun, 2001). In the present study, the algorithm for the PD index was employed not only for specific pairs of peptides (epitopes), but also for all regions of the two proteins to evaluate the similarity between g14 and g11/g28 or g13. The results, coupled with phylogenetic analyses, showed that buckwheat 2S albumin g13 polypeptide was less related to g14 than g11/g28 was related to g14. In our previous study, we found that g13 content is scarce in seeds (Katsube-Tanaka & Monshi, 2022), which appears to be a reason for the gene encoding g13 having not been identified until recently (Yasui et al., 2016). In addition, g13 is hydrophobic compared to the other 2S albumin polypeptides, particularly in the first and third domains (Katsube-Tanaka & Monshi, 2022), possibly producing more heat-induced aggregates that are resistant to gastric digestion (Ma et al., 2021). Low digestibility is one of important characteristics for food allergens and is significant as well as abundance of

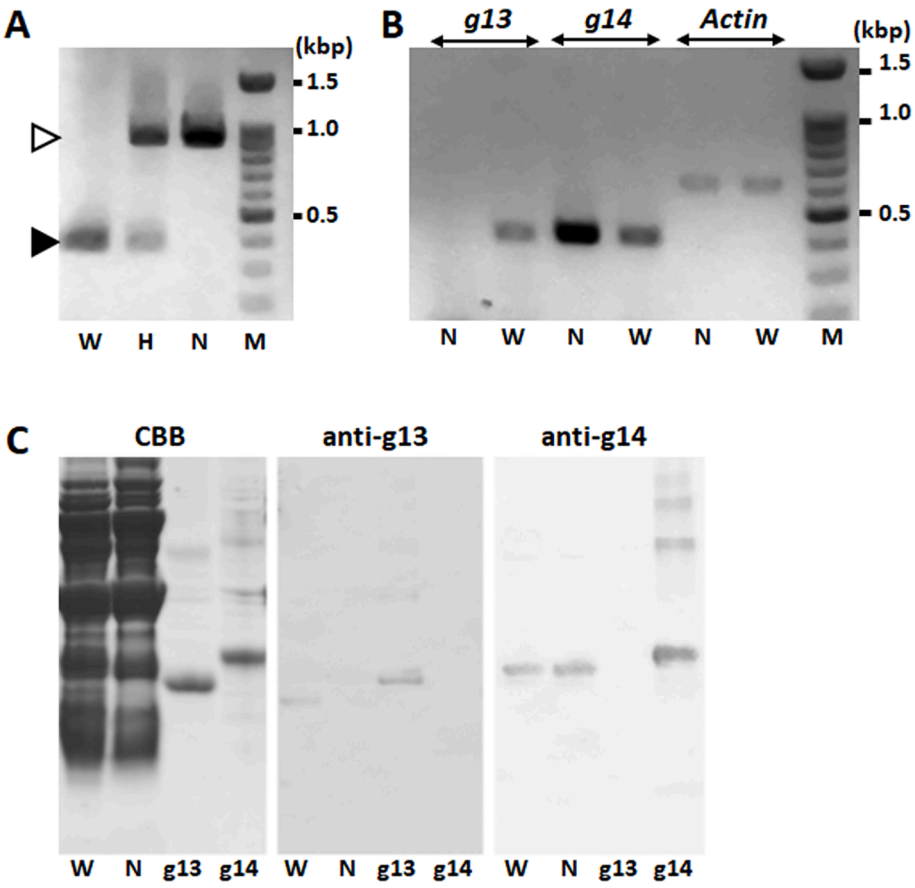

Fig. 5. Development and characterization of *g13_null* homozygote plants. (A) Genotypes for *g13* wild type (W), *g13_null* heterozygote (H) and homozygote (N) plants were confirmed. White and black arrowheads indicate PCR fragments derived from *g13* with and without a 531 bp insertion, respectively. (B)Transcription levels of *g13*, *g14*, and *actin* were compared with RT-PCR between the *g13_null* homozygote (N) and the wild type (Harunoibuki, W) seeds. (C) The translational products of *g13* and *g14* in seeds were analyzed with SDS-PAGE followed by CBB staining and western blotting with anti-g13 antibody prepared against recombinant g13 and with anti-g14 antibody against peptide sequence of CRIGERLIKEGVRDLKE. W and N indicate seed extract of wild type and *g13_null* homozygote, respectively. g13 and g14 indicate recombinant protein of *g13* and *g14*, respectively. M indicates a size marker.

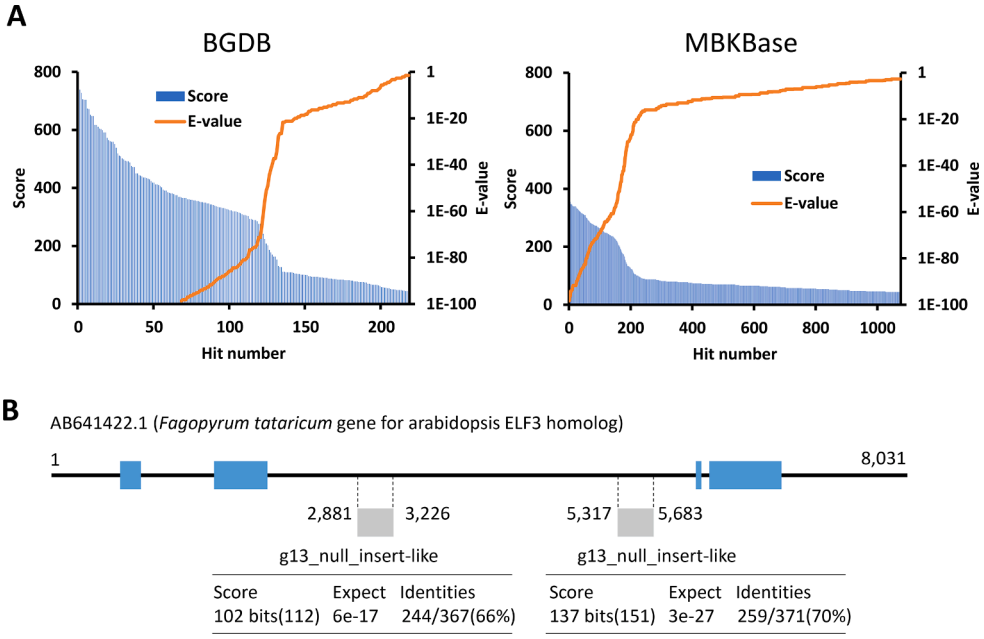

Fig. 6. BLAST search for the insert sequence of *g13_null* allele. (A) BLASTN search was conducted at Buckwheat Genome DataBase (BGDB, *Fagopyrum esculentum,* left) and Molecular Breeding Knowledgebase (MBKbase, *Fagopyrum tataricum,* right) with a query of g13_null insert sequence (531 letters). Hits with a score >44 were shown. The E-value line was not shown under 1E-100. (B) The sole hit detected by BLASTN (discontiguous megablast) at NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with the above-mentioned query was shown. Blue and gray rectangulars indicate a coding region and the *g13_null* insert-like region, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

allergens. Thus, g13 is a noteworthy polypeptide in efforts to lower allergenicity in buckwheat.

4.2. Large insertion or deletion and genetic diversity

Insertions and deletions in a particular locus generally arise due to cellular mechanisms such as unequal crossing-over, replication slippage, and transposable element movement within the genome (Britten et al., 2003). Of the two classes of transposable elements, class 1 LTR-retrotransposons are largely intergenic, constituting plant genome size differences, making up to 75 % of nuclear DNA (Baucom et al., 2009). The majority of the common buckwheat genome (64.8 %) is occupied by

retrotransposons (RTEs, class I repeats), which is two times greater than that of tartary buckwheat (30.7 %) (Penin et al., 2021) and more than that in the genomes of other cross-pollinated (sunflower ~60 %, oil palm ~57 %, and apple ~42.9 %) and self-pollinated (peanut ~64.74 %, tomato ~60 %, and soybean ~42 %) crops (Bennetzen & Park, 2018; Kirov et al., 2020; Zhuang et al., 2019). Therefore, the genome size of common buckwheat (1.2 Gbp) is three times larger than that of tartary buckwheat (0.48 Gbp) (Penin et al., 2021), although both species have the same chromosome number (Neethirajan et al., 2011).

The other transposable elements, DNA transposon type II elements (1.98 %), are present in the common buckwheat genome (Penin et al., 2021) and can enhance the allelic diversity in the entire genome. MITE-like insertions (generally 50–800 bp) frequently occur near (<2 kb upstream or downstream) or within the coding sequence of a gene (Dubin et al., 2018), accelerating the disruption of a particular gene function.

Shirasawa et al. (2012) identified 504 MITE-like insertions (AhMITE1s) with a mean length of 205.5 bp which disrupted the function of the fatty-acid desaturase-encoding gene *ahFAD2B* in peanut. MITE-like sequences in the hAT family have been examined in many allergenic crops such as rice, wheat, rye, maize, and *Brassica* species (Lu et al., 2012; Muehlbauer et al., 2006; Perumal et al., 2020). Thus, such MITE-like transposon elements in plant genomes are useful for generating wide genetic diversity, which has been previously noted in the genomes of rice (up to 10 %), *Brassica* (0.7–4.5 %), potato (0.72 %), and bread wheat (0.16 %) (Chen et al., 2013; Crescente et al., 2018; Perumal et al., 2020; Zavallo et al., 2020).

In our previous studies, novel alleles for the trypsin-resistant zero-repeat subunit of 13S globulin were identified in a natural common buckwheat population (Katsube-Tanaka et al., 2014; Monshi et al., 2020). The zero-repeat subunit is expected to show higher allergenicity than the other 1–6 repeat subunits because of its low digestibility (Khan et al., 2012). Three novel alleles, *GlbNA2, GlbNA3,* and *GlbNC* have a ~ 200 bp MITE-like sequence near or immediately after the stop codon, possibly affecting their expression (Monshi et al., 2020). A BLAST search of these MITE-like sequences yielded 50, 30, and 152 hits with E-values less than 1.0e-01 in BGDB for *GlbNA2, GlbNA3,* and *GlbNC*, respectively. Thus, *g13_null* insert-like sequences are more prevalent than the MITE-like sequence of zero-repeat alleles in the common buckwheat genome.

Both retrotransposons and MITE-like sequences are widespread in the buckwheat genome, and the buckwheat population tolerated this spreading, which could be a reason for the extensive genetic diversity of buckwheat, along with the wider nucleotide diversity than other cross-pollinated crops, as demonstrated by Mizuno and Yasui (2019). Such a genetic diversity in common buckwheat may enable us to find other useful alleles for the other allergens in this species.

4.3. Development of hypoallergenic buckwheat

Accumulation of useful alleles from existing plant genetic resources is an effective strategy for improving crop yield and quality. Mutagenesis has been frequently used to create allelic variations, especially in self-pollinating crops (reviewed by Jankowicz-Cieslak et al., 2017). However, it requires the use of cancer-causing substances or large facilities to irradiate. Genome editing is also important strategies to create useful alleles (Sugano et al., 2020) but it is still difficult in buckwheat. The outcrossing nature of cross-pollinated crops is one of the most important characteristics for the distribution of genetic diversity (Bhandari et al., 2017), which makes it difficult to fix useful traits, but can allow for the conservation of wide genetic diversity. Such mating systems allow common buckwheat to conserve this diversity with higher nucleotide diversity than the other allogamous (maize and sunflower) and autogamous (rice and soybean) crops (Mizuno & Yasui, 2019). Therefore, this feature in common buckwheat aids the present study in identifying novel alleles from natural common buckwheat populations that can contribute to the reduction of allergenic proteins.

Allergenic 2S albumin in common buckwheat is resistant to trypsin/pepsin digestion. It induces immediate hypersensitivity reactions, including anaphylaxis (Tanaka et al., 2002) when consumed by allergic patients. To reduce the incidence of fatal allergic reactions, the development of hypoallergenic buckwheat could be helpful in enhancing the quality of life of sensitive individuals. Hypoallergenic lines of peanuts and soybeans have previously been identified by lowering the content of specific allergenic proteins (Pandey et al., 2019; Watanabe et al., 2017). However, the present study is the first attempt to develop a less allergenic common buckwheat. Similar approaches to identify a useful allele from natural common buckwheat populations are now occurring for the 2S albumin polypeptides g11, g14, and g28. Next generation sequencing with AI system predicting a protein 3D structure from its amino acid sequence might be promising for extensive analysis of genetic diversity in common buckwheat.

Further immunoreactivity analysis using the blood serum of buckwheat allergy patients is necessary to confirm its lower allergenicity. In the near future, the identified null allele homozygous line will be combined with all useful alleles to produce hypoallergenic buckwheat.

5. Conclusion

2S albumin g13 is hydrophobic and rare in seeds, and its characteristics are largely unknown. In this study, we analyzed g13 structure and explored useful alleles of g13 to develop a hypoallergenic buckwheat. The g13 was less related to g14 than g11/g28 was related to g14 in terms of property distance. A null allele with a 531 bp insertion for g13 was found in natural populations of common buckwheat. The *g13_null* allele homozygote expressed no g13 transcript or protein. The *g13_null* insert-like sequence resided at a high frequency in the buckwheat genome. Thus, the spontaneous *g13_null* insert-like sequence could lower allergenicity and expand the genetic diversity in common buckwheat.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Tomoyuki Katsube-Tanaka reports financial support was provided by Kyoto University. Tomoyuki Katsube-Tanaka reports financial support was provided by Japan Society for the Promotion of Science. Fakhrul Islam MONSHI reports financial support was provided by Tojuro Iijima Foundation for Food Science and Technology.

Data availability

Data will be made available on request.

Acknowledgements

This study was supported by JSPS Research Fellowships for Young Scientists and a research fund for foreign students from the Tojuro Iijima Foundation for Food Science and Technology, Japan to FIM, and JSPS KAKENHI Grant Number 20H02967 to TKT.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochms.2022.100138.

References

Baucom, R. S., Estill, J. C., Chaparro, C., Upshaw, N., Jogi, A., Deragon, J. M., … Bennetzen, J. L. (2009). Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genetics, 5*(11), e1000732.

Bennetzen, J. L., & Park, M. (2018). Distinguishing friends, foes, and freeloaders in giant genomes. *Current Opinion in Genetics & Development, 49*, 49–55.

Bhandari, H. R., Bhanu, A. N., Srivastava, K., Singh, M. N., Shreya, & Hemantaranjan, A. (2017). Assessment of genetic diversity in crop plants -an overview. *Advances in Plants & Agriculture Research, 7*(3), 279–286. https://doi.org/10.15406/apar.2017.07.00255

Britten, R. J., Rowen, L., Williams, J., & Cameron, R. A. (2003). Majority of divergence between closely related DNA samples is due to indels. *Proceedings of the National academy of Sciences of the United States of America, 100*, 4661–4665. https://doi.org/10.1073/pnas.0330964100

Chen, J., Hu, Q., Zhang, Y., Lu, C., & Kuang, H. P. (2013). MITE: A database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Research, 42*, D1176–D1181.

Crescente, J. M., Zavallo, D., Helguera, M., & Vanzetti, L. S. (2018). MITE Tracker: An accurate approach to identify miniature inverted-repeat transposable elements in large genomes. *BMC Bioinformatics, 19*, 348.

Dubin, M. J., Scheid, O. M., & Becker, C. (2018). Transposons: A blessing curse. *Current Opinion in Plant Biology, 42*, 23–29.

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95–98.

Heffler, E., Pizziment, S., Badiu, I., Guida, G., & Rolla, G. (2014). Buckwheat allergy: An emerging clinical problem in Europe. *Journal of Allergy Therapy, 5*, 1–4.

Ivanciuc, O., Midoro-Horiuti, T., Schein, C. H., Xie, L., Hillman, G. R., Goldblum, R. M., & Braun, W. (2009). The property distance index PD predicts peptides that cross-react with IgE antibodies. *Molecular Immunology, 46*(5), 873–883.

Jankowicz-Cieslak, J., Mba, C., & Till, B. J. (2017). Mutagenesis for crop breeding and functional genomics. In J. Jankowicz-Cieslak, T. Tai, J. Kumlehn, & B. Till (Eds.), *Biotechnologies for plant mutation breeding*. Cham: Springer. https://doi.org/10.1007/978-3-319-45021-6_1.

Katayama, S., Yamaguchi, D., Suzuki, Y., Athamneh, A. M. A., Mitani, T., Satoh, R., … Nakamura, S. (2018). Oral immunotherapy with a phosphorylated hypoallergenic allergen ameliorates allergic responses more effectively than intact allergen in a murine model of buckwheat allergy. *Molecular Nutrition & Food Research, 62*, e1800303.

Katsube-Tanaka, T., & Monshi, F. I. (2022). Characterization of 2S albumin allergenic proteins for anaphylaxis in common buckwheat. *Food Chemistry: Molecular Sciences, 100127*. https://doi.org/10.1016/j.fochms.2022.100127

Katsube-Tanaka, T., Nakagawa, M., Sano, M., & Yasui, Y. (2014). Development of novel common buckwheat (*Fagopyrum esculentum* M.) plants with lowered contents of tandem repeat-less 13S globulin-Discrimination methods of the tandem repeat-less genes. *Journal of Crop Research, 59*, 31–35.

Khan, N., Takahashi, Y., & Katsube-Tanaka, T. (2012). Tandem repeat inserts in 13S globulin subunits, the major allergenic storage protein of common buckwheat (*Fagopyrum esculentum* Moench) seeds. *Food Chemistry, 133*, 29–37.

Kirov, I., Omarov, M., Merkulov, P., Dudnikov, M., Gvaramiya, S., Kolganova, E., … Soloviev, A. (2020). Genomic and transcriptomic survey provides new insight into the organization and transposition activity of highly expressed LTR retrotransposons of sunflower (*Helianthus annuus* L.). *International Journal of Molecular Sciences, 21*(23), 9331.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution, 35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096. PMID: 29722887, PMCID: PMC5967553.

Lu, C., Chen, J., Zhang, Y., Hu, Q., Su, W., & Kuang, H. (2012). Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Molecular Biology and Evolution, 29*(3), 1005–1017.

Luparelli, A., Losito, I., De Angelis, E., Pilolli, R., Lambertini, F., & Monaci, L. (2022). Tree nuts and peanuts as a source of beneficial compounds and a threat for allergic consumers: Overview on methods for their detection in complex food products. *Foods, 11*(5), 728–768.

Ma, R., Lin, Z., Wu, Y., Gao, Z., Hu, B., Xu, L., … Nishinari, K. (2021). Modulating the in vitro gastric digestion of heat-induced beta-lactoglobulin aggregates: Incorporation with polysaccharide. *Food Chemistry, 354*, 129506. https://doi.org/10.1016/j.foodchem.2021.129506

Mizuno, N., & Yasui, Y. (2019). Gene flow signature in the S-allele region of cultivated buckwheat. *BMC Plant Biology, 19*, 125.

Monshi, F. I., Khan, N., Kimura, K., Seita, S., Yamamoto, Y., & Katsube-Tanaka, T. (2020). Structure and diversity of 13S globulin zero-repeat subunit, the trypsin-resistant storage protein of common buckwheat (*Fagopyrum esculentum* M.) seeds. *Breeding Science, 70*(1), 118–127.

Moreno, F. J., & Clemente, A. (2008). 2S albumin storage proteins: What makes them food allergens? *The Open Biochemistry Journal, 2*, 16–28.

Muehlbauer, G. J., Bhau, B. S., Syed, N. H., Heinen, S., Cho, S., Marshall, D., … Flavell, A. J. (2006). A hAT superfamily transposase recruited by the cereal grass genome. *Molecular Genetics and Genomics, 275*(6), 553–563.

Neethirajan, S., Hirose, T., Wakayama, J., Tsukamoto, K., Kanahara, H., & Sugiyama, S. (2011). Karyotype analysis of buckwheat using atomic force microscopy. *Microscopy and Microanalysis, 17*(4), 572–577.

Pandey, A. K., Sudini, H. K., Upadhyaya, H. D., Varshney, R. K., & Pandey, M. K. (2019). Hypoallergen peanut lines identified through large-scale phenotyping of global diversity panel: Providing hope toward addressing one of the major global food safety concerns. *Frontiers in Genetics, 10*, 1177.

Park, K., Jeong, K., & Lee, S. (2016). Clinical and laboratory findings of childhood buckwheat allergy in a single tertiary hospital. *Korean Journal of Pediatrics, 59*(10), 402–407.

Penin, A. A., Kasianov, A. S., Klepikova, A. V., Kirov, I. V., Gerasimov, E. S., Fesenko, A. N., & Logacheva, M. D. (2021). High-resolution transcriptome atlas and improved genome assembly of common buckwheat, *Fagopyrum esculentum*. *Frontiers in Plant Science, 12*, Article 612382.

Perumal, S., James, B., Tang, L., Kagale, S., Robinson, S. J., Yang, T. J., & Parkin, I. A. P. (2020). Characterization of B-genome specific high copy hAT MITE families in *Brassica nigra* genome. *Frontiers in Plant Science, 11*, 1104.

Sano, M., Nakagawa, M., Oishi, A., Yasui, Y., & Katsube-Tanaka, T. (2014). Diversification of 13S globulins, allergenic seed storage proteins, of common buckwheat. *Food Chemistry, 155*, 192–198.

Shirasawa, K., Hirakawa, H., Tabata, S., Hasegawa, M., Kiyoshima, H., Suzuki, S., … Isobe, S. (2012). Characterization of active miniature inverted-repeat transposable elements in the peanut genome. *TAG. Theoretical and Applied Genetics, 124*, 1429–1438.

Sugano, S., Hirose, A., Kanazashi, Y., Adachi, K., Hibara, M., Itoh, T., … Yamada, T. (2020). Simultaneous induction of mutant alleles of two allergenic genes in soybean by using site-directed mutagenesis. *BMC Plant Biology, 20*, 513. https://doi.org/10.1186/s12870-020-02708-6

Tanaka, K., Matsumoto, K., Akasawa, A., Nakajima, T., Nagasu, T., Iikura, Y., & Saito, H. (2002). Pepsin-resistant 16-kD buckwheat protein is associated with immediate hypersensitivity reaction in patients with buckwheat allergy. *International Archives of Allergy and Immunology, 129*(1), 49–56.

Teramura, H., Kondo, K., Suzuki, M., Kobayashi, H., Furukawa, T., Kusano, H., & Shimada, H. (2019). Aberrant endosperm formation caused by reduced production of major allergen proteins in a rice flo2 mutant that confers low-protein accumulation in grains. *Plant Biotechnology, 36*(2), 85–90. https://doi.org/10.5511/plantbiotechnology.19.0312a

Venkatarajan, M., & Braun, W. (2001). New quantitative descriptors of amino acids based on multidimensional scaling of a large number of physical-chemical properties. *Journal of Molecular Modeling, 7*, 445–453. https://doi.org/10.1007/s00894-001-0058-5

Watanabe, D., Adanyi, N., Takacs, K., Maczo, A., Nagy, A., Gelencser, E., … Vollmann, J. (2017). Development of soybeans with low P34 allergen protein concentration for reduced allergenicity of soy foods. *Journal of the Science of Food and Agriculture, 97*(3), 1010–1017.

Yasui, Y., Hirakawa, H., Ueno, M., Matsui, K., Katsube-Tanaka, T., Yang, S. J., … Mori, M. (2016). Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes. *DNA Research, 23*(3), 215–224.

Zavallo, D., Crescente, J. M., Gantuz, M., Leone, M., Vanzetti, L. S., Masuelli, R. W., & Asurmendi, S. (2020). Genomic re-assessment of the transposable element landscape of the potato genome. *Plant Cell Reports, 39*(9), 1161–1174.

Zhang, Z., Zhou, M., Tang, Y., Li, F., Tang, Y., Shao, J., … Wu, Y. (2012). Bioactive compounds in functional buckwheat food. *Food Research International, 49*(1), 389–395.

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., … Varshney, R. K. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature Genetics, 51*, 865–876.