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

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 undertaken (Tanaka et al., 2002), agronomical studies aimed at breeding and production of hypoallergenic buckwheat have 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

2666-5662/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
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.aa1.1, Fes_sc0000087.1.g000011.aa1.1, Fes_sc0000087.1.g000013.aa1.1, Fes_sc0000087.1.g000014.aa1.1, and Fes_sc0000087.1.g000028.aa1.1, 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 hydrophilic 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 hypothesized 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 materials for the preparation of gDNA. A large number of seeds (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!-S method (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!-S-R, Rizo, Tsukuba, Japan) and DNasel (NIPFON GENE, Tokyo, Japan). The RNA was primed with oligo-dT adaptors 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 TAA2SA_g14.BamS:5’-gggcttctTATCAATACTGGTTCCTC-3’ and actin (BW actin Self3 Fwd: 5’-GGGCATCACTCCTTCTCATAATGC-3’ and BW actin Self3 Rev: 5’-GATACTCACTACTTCTGTGTTG-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, and 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 (CRGERLIKEGFRLDE) (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
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
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
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

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 rectangles 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.)
buckwheat, along with the wider nucleotide diversity than other cross-
spreading, which could be a reason for the extensive genetic diversity of
the buckwheat genome, and the buckwheat population tolerated this
less than 1.0e-01 in BGDB for (AhMITE1s) with a mean length of 205.5 bp which disrupted the func-
is an effective strategy for improving crop yield and quality. Mutagen
esis has been frequently used to create allelic variations, especially in
self-pollinating crops (reviewed by Jankowicz-Cieslak et al., 2017).
3. Development of hypoallergenic buckwheat

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, GbN2A, GbN3A, and GbNC 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 GbN2A, GbN3A, and GbNC, 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 (Shandari 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 buck-
weed 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 character-
istics 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 re-
lationships which may be considered as potential competing interests: Tomoyoaki Katsube-Tanaka reports financial support was provided by Kyoto University. Tomoyoaki Katsube-Tanaka reports financial support was provided by Japan Society for the Promotion of Science. Fakhru Islam MONSEH 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.
