Characterization of 2S albumin allergenic proteins for anaphylaxis in common buckwheat

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

[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

ABSTRACT

2S albumin (2SA) is responsible for anaphylaxis following consumption of buckwheat in allergic individuals. To reduce allergen incidents, characterization of 2SA polypeptides is prerequisite, thus was analyzed in this study. Of the five 2S albumin genes (*g03, g11, g13, g14,* and *g28*), *g03* was seemingly non-functional. The g14 content was 3- and 40-fold higher than that of g11/g28 and g13, respectively. The g11/g28 were more processed to a ~8 kDa band from a 16 kDa band than g14 in seeds, agreeing with that g11/g28 have high similarity with Fag e 8kD. Meanwhile, anti-g13 produced only a single ~10 kDa band. Modification of g13 and domain exchange between g13 and g14 suggested that the hydrophobicity of the first domain and the nature of some amino acids in g13 contributed, at least in part, to the lower apparent molecular weight of g13 than expected. Thus, g13 might be an unexplored and noteworthy allergen.

1. Introduction

Common buckwheat (*Fagopyrum esculentum* Moench) is an important food crop in less fertile and mountainous areas where world's major crops such as wheat, maize, and rice cannot be grown owing to short summers, low temperatures, and scarce rainfall. The crop is not only traditionally grown in east and south Asia and middle and eastern European countries but is also utilized in western countries because of its gluten-free nature, attractiveness as an ethnically diverse diet (Heffler et al., 2014), and nutritional and health-promoting benefits (Katsube-Tanaka, 2016). Because buckwheat can be found in various foods as a hidden ingredient as well as in pillow fillings (Guzmán et al., 2020), it is of great concern that the increase in buckwheat consumption in developed countries leads to an increase in allergenic incidents as food allergy, occupational allergy, or domestic allergy (Norbäck and Wieslander, 2021).

Previous studies have identified and characterized buckwheat-allergenic proteins with diverse apparent molecular weights (MWas) (24, 19, 18, 16, 9, and 8 kDa). The 24 kDa allergen (Fag e 1) is a β polypeptide of 13S globulin (Nair and Adachi, 1999) and it is found as the allergen of most patients with buckwheat allergy. The 16 kDa protein (Fag e 2) has been identified as a 2S albumin protein (Choi et al., 2007a; Koyano et al., 2006; Tanaka et al., 2002) and the 19 kDa protein

(Fag e 3) has been recognized as a vicilin-like allergen (Choi et al., 2007b). Recently, Geiselhart et al. (2018) identified 55 kDa vicilin-like and 3.9 kDa hevein-like antimicrobial polypeptides, recognized as Fag e 5 and Fag e 4, respectively, by the International Union of Immunological Societies (IUIS) allergen nomenclature sub-committee. Fag e 10 kDa (α-amylase inhibitor/trypsin inhibitor, 10 kDa/ 16 kDa) and Fag e TI (Trypsin inhibitor, 7 kDa/ 9 kDa) were also reported (Satoh et al., 2020; Heffler et al., 2014).

One of the most dominant buckwheat seed proteins, 2S albumin, consists of 8–16 kDa water-soluble polypeptides (Radovic et al., 1999) that are resistant to pepsin/trypsin and cause immediate hypersensitivity reactions in buckwheat-sensitive patients (Tanaka et al., 2002). Fag e 2 is composed of 127 amino acids (Choi et al., 2007a) and exhibits 47 % identity with the reported buckwheat 8 kDa allergenic 2S albumin (Fag e 8kD). Fag e 2 also exhibits homology with other known allergens such as 2S albumins of peanut (Viquez et al., 2001; Ara h 2, 30 % and Ara h 6, 21 % identity), English nut (Teuber et al., 1998; Jug r 1 2 r, 26 % identity), and Brazil nut (Altenbach et al., 1987; Ber e 1, 26 % identity).

Recently, Yasui et al. (2016) identified all 2S albumin genes using the buckwheat genome database (BGDB), which was constructed with a draft assembly of the buckwheat genome, with a total length of 1.18 Gbp sequence. A Basic Local Alignment Search Tool (BLAST) search in the BGDB generated one identical gene, *Fes_sc0000087.1.g000014.aua.1*

(hereinafter abbreviated as *g14*) to Fag e 2 and four homologues, (*Fes_sc0000087.1.g000011.aua.1* (*g11*), *Fes_sc0000087.1.g000013.aua.1* (*g13*), *Fes_sc0000087.1.g000028.aua.1* (*g28*), and *Fes_sc0007211.1. g000003.aua.1* (*g03*). The amino acid sequence of g14 had the same peptide as the epitope (EGVRDLKE) for IgE of buckwheat-sensitive patients at amino acid numbers in the range of 122–130, whereas g13 exhibited 79 % similarity to g14. The other genes, *g11*, *g28*, and *g03*, were observed to encode proteins similar to buckwheat 8 kDa allergen with 97 %, 98 %, and 43 % similarity to BW8KDa (GenBank Accession: AB055892), respectively. The predicted molecular weights from the nucleotide sequences are slightly different from each other (13.6, 13.7, 14.4, and 14.9 kDa for *g11*, *g28*, *g13*, and *g14*, respectively) except for *g03*. However, the reason why some are more related to the 8 kDa and others to the 16 kDa polypeptides and the difference in fundamental characteristics between the five 2S albumin proteins is still not fully understood. We assumed there is a difference in allergenicity between polypeptides even within a same protein family. In this study, the nucleotide sequences, accumulation, and MWa of the gene products were analyzed to deepen our knowledge for safe foods containing buckwheat and minimizing allergenic problems. Particularly, newly identified polypeptide g13 was further characterized.

2. Materials and methods

2.1. Plant materials and preparation of genomic DNA

The Japanese cultivars of common buckwheat, Harunoibuki and Shinano 1, were used as plant materials. Genomic DNA was extracted from the seeds through the DNAs-ici!S method (Rizo, Tsukuba, Japan).

2.2. Isolation and identification of g03 genes

The coding, upstream, and downstream regions of *g03* were amplified using polymerase chain reaction (PCR) with a forward primer (KAZg03-60Kpn_FWD: 5′- tacggTACCCACATCGTTTCAACACTC -3′) and reverse primer (KAZg03 + 520Bam_REV: 5′- cgcgGATcCAAGTT-CATTGTCCAATG -3′). PCR conditions were as follows: initial denaturation at 94 °C for 2 min, 30 cycles of denaturation 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 using KOD FX Neo DNA polymerase (TOYOBO, Osaka, Japan). The amplified fragments were isolated from the agarose gel and purified using a FastGene Gel/PCR Extraction Kit (Nippon Genetics, Tokyo, Japan). The purified fragments were cloned into the pTAC-2 vector using a TA cloning kit (BioDynamics Laboratory, Inc., Tokyo, Japan) and the TArget Clone kit (TOYOBO, Osaka, Japan), and sequenced. Allele frequency examination of *g03* was performed on 20 plants each of the cultivars Harunoibuki and Shinano 1 (single technical replicate) using the aforementioned specific primers. PCR primers were purchased from Eurofins Genomics Ltd. (Tokyo, Japan).

2.3. Constructions of E. coli expression plasmids

Leaf genomic DNA was isolated from the cultivar Harunoibuki and used as a template for the amplification of 2S albumin genes (*g11*, *g28*, *g13*, and *g14*) with the specific primers listed in Table S1. The signal peptide was predicted using SignalP – 5.0 and omitted. The *E. coli* expression vector pET21d (Merck KGaA, Darmstadt, Germany) was cut with restriction enzyme *Nco*I, treated with Klenow polymerase to create blunt ends, and digested with *Bam*HI. The amplified PCR products were purified, digested with *Bam*HI, and cloned into the *Bam*HI/*Nco*I sites of the vector pET21d. Restriction enzymes and Klenow polymerase were from NIPPON GENE (Tokyo, Japan).

The modified g13 expression plasmid was prepared with a forward primer (g13_I14gL19gM21g_F: 5′- ATGggACAaggGAAGACGATGGG -3′) and reverse primer (g13_I14gL19gM21g_R: 5′- aCACTTTG-CAccTGCTCCGGTATC -3′) and the g13 expression plasmid was used as a

PCR template.

For domain exchange between g13 and g14, the HindIII site was first incorporated into the g13 expression plasmid with a forward primer (g13_Hind3-CCA_F: 5′- GGGAaGCtTTGAGTAGTTTCAGAGAGC -3′) and reverse primer (g13_Hind3_R: 5′- CATATTTCTCAATCGAAACGTGCATG -3′) (Table S2). The g14 coding region originally contained a HindIII site. The HindIII-HindIII (on the vector) fragment was thereafter exchanged between the g13 (with HindIII) and g14 expression plasmids to produce domain exchange series Nos.4 and 5. Thereafter, the *Kpn*I site was incorporated into the g13 (with HindIII) and g14 expression plasmids using the forward primer (g13_KpnI_F: 5′-GTTTTTGAGGTTGGCTATG-GAGGATTTG -3′) and reverse primer (g13_KpnI_R: 5′-TCggTACc-CATCTCATCTCTCCCTTCC) and forward primer (g14_KpnI_F: 5′-TGAAAGGTTGATTAAGGAGGGGGTTAG -3′) and reverse primer (g14_KpnI_R: 5′-CCggTACcCCCTTTTTGATTCTCAACC -3′) to produce series Nos.1 and 8, respectively. The *Xba*I (on the vector)-*Kpn*I short fragment was thereafter amplified from series Nos.1, 4, 5, and 8. After digestion of the short fragments and series Nos.1 and 8 (vector) with *Xba*I and *Kpn*I, the short and long fragments from the vector were ligated to produce series Nos.2, 3, 6, and 7.

All plasmids were verified through Sanger sequencing. *E coli* Rosetta-gami B (DE3) cells (Merck KGaA, Darmstadt, Germany) were transformed by expression plasmids for each gene.

2.4. Expression and purification of recombinant 2SA from E. coli

E coli Rosetta-gami B (DE3) cells harboring the expression plasmids were incubated in Luria-Bertani (LB) medium containing ampicillin (50 ug/ml), tetracycline (12.5 ug/ml), kanamycin (15 ug/ml), chloramphenicol (34 ug/ml), and 5 μM riboflavin at 37 °C and 160 rpm until the optical density at 600 nm reached 0.5. The expression of recombinant proteins was thereafter induced by adding isopropylthio-β-ᴅ-galactoside (1 mM) for 3 h at 37 °C and 160 rpm. All expressed proteins were separated into soluble and insoluble fractions through centrifugation at 14,000 rpm for 10 min at 4 °C after sonication. The recombinant proteins in the insoluble fraction were solubilized in 3 M urea and densitometrically quantified using bovine serum albumin (BSA). A total of 40, 10, and 2.5 ng of g13, 750, 188, and 47 ng of g14, and 630, 158, and 39 ng of g28 were used as calibration standards (STD 1–3) to quantify their accumulation levels in seeds.

2.5. Antibodies for 2SA

Polyclonal antibodies were purchased from Eurofins Genomics Ltd. (Tokyo, Japan), which were produced in rabbits against the partially purified recombinant proteins g11, g28, and g13 (anti-g11, anti-g28, and anti-g13, respectively), and purchased from Sigma-Aldrich Japan GK. (Tokyo, Japan), which were produced against the peptide (CRI-GERLIKEGVRDLKE) containing the epitope sequence of Fag e 2 for anti-g14.

2.6. SDS-PAGE and western blot analyses

Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) was performed as described by Laemmli (1970) with minor modifications using 17 % T acrylamide gel at a constant voltage of 200 V for 70 min. Immunoblotting was performed as described by Khan et al. (2012). Briefly, proteins were separated by SDS-PAGE and transferred electrophoretically to a nitrocellulose membrane (Cytiva, Uppsala, Sweden). Thereafter, the membrane was 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 membranes were first probed with primary antibodies against g11, g28, g13, and g14 at room temperature for 60 min. After three thorough washes 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 thereafter washed three times with TBST for 20 min. After antibody recognition, the membrane was incubated with 5-Bromo-4-chloro-3-indolyl-phosphate and nitroblue tetrazolium for color development. Detection was stopped using deionized water. SDS-PAGE for molecular weight estimation was repeated three times and average and standard error were calculated. Western blotting was repeated at least three times. All chemical reagents, unless otherwise indicated, were purchased from NACALAI TESQUE, Inc. (Kyoto, Japan).

2.7. Statistical analyses

Tukey's tests were used ($p < 0.05$) to compare the mean values of apparent molecular weight between the recombinant 2S albumins (n = 3). The analyses were performed using the statistical software BellCurve for Excel version 3.21 (Social Survey Research Information Co. Ltd., Tokyo, Japan, https://bellcurve.jp/ex/).

3. Results

3.1. 2S albumin genes in common buckwheat

The draft genome sequence of common buckwheat (BGDB) released in 2016 successfully identified all 2S albumin genes (*g03, g11, g13, g14,* and *g28*) (Yasui et al., 2016). Fig. 1 shows the positional relationship among the five genes of 2S albumin located in the two scaffolds (Fes_sc0000087.1 and Fes_sc0007211.1) on a buckwheat genome. In the former scaffold, the 2S albumin genes *g11, g13,* and *g14* exist within a 17 kbp region successively in the same direction, whereas *g28* exists in the inverse direction, 108 kbp away from *g11,* whereas *g03* exists in a separate scaffold (Yasui et al., 2016).

In the BGDB database, translation initiation (ATG) was mispredicted for *g03.* After correction, a deletion of 8 bp after 111 bp from ATG and a deletion of 622 bp at 88 bp upstream from the stop codon were detected, suggesting that the gene *g03* is non-functional (Fig. 2). PCR-based genotyping for *g03* using 20 plants from the cultivars Harunoibuki and Shinano 1 illustrated that all the plants exhibited approximately 600 bp authentic band, whereas several plants exhibited an extra band of 1200 bp (Fig. S1). Sequencing of the 1200 bp extra band demonstrated that the band was derived from two new alleles (*g03_s* and *g03_h*). Premature translation termination was observed in either of the two alleles (Fig. 2), suggesting that *g03* is a pseudo gene.

3.2. Expression and purification of recombinant 2S albumin

The recombinant 2S albumin (g11r, g28r, g13r, and g14r) highly expressed in *E. coli* cells was separated into soluble and insoluble fractions after disruption of cells by sonication. Because the induction of the

recombinant proteins was rapid at a high temperature (37 °C), approximately-two-thirds of the recombinant proteins were insoluble in g11r, g28r, and g14r (Fig. 3A). Meanwhile, SDS-PAGE demonstrated that most of the recombinant protein was insoluble in g13r and it migrated faster than the other 2S albumin proteins. The migration of g11r, g28r, and g14r was similar to that of the major band (16 kDa) of the 2S albumin fraction extracted from common buckwheat seeds (data not shown). The expression level and purity of recombinant proteins in the insoluble fraction were relatively high. Thus, insoluble 2S albumin was solubilized and quantified with BSA and subsequently used as a calibration standard for western blotting.

3.3. Reactivity of anti-g11, anti-g28, anti-g13, and anti-g14 antibodies

The reactivity of anti 2S albumin antibodies was examined against recombinant 2S albumin prepared in *E. coli* cells (g11r, g28r, g13r, and g14r) (Fig. 3B). The anti-g11 and anti-g28 antibodies used in this study cross-reacted equally to both g11r and g28r because the similarity of the g11 and g28 amino acid sequences to buckwheat 8 kDa allergen (a member of 2S albumin) is as high as 97 % and 98 %, respectively (Yasui et al., 2016). Meanwhile, anti-g11 and anti-g28 antibodies were neither reactive nor remarkably less reactive to g13r and g14r. Anti-g13 and anti-g14 antibodies were almost exclusively specific to g13r and g14r, respectively.

3.4. Detection and quantification of 2S albumin in buckwheat seeds

Albumin fractions prepared from bulked seed flour of the cultivars 'Shinano 1' and 'Harunoibuki' were detected using Coomassie Brilliant Blue (CBB) and antibodies against g11, g28, g13, and g14. Smaller bands of ~8 kDa and major bands of 16 kDa were found at varying ratios in the bulked seed flour with anti-g11, anti-g28, and anti-g14 (Fig. 4A). The semi-quantified ratios of the 8 kDa band to the total (8 and 16 kDa bands) were ~0.5, ~0.4–0.5, and <0.2 for g11, g28, and g14, respectively. Anti-g11 and anti-g28 were reactive with g11 and g28, and vice versa, implying that these antibodies cannot distinguish g11 and g28 (Fig. 3B). Therefore, we quantified g11 and g28 simultaneously as g11/g28. Anti-g14 produced an extra faint band of ~14 kDa. Anti-g13 produced only a single band of ~10 kDa.

The accumulation levels of the major bands of 2S albumin, g11/g28, g13, and g14, were compared using appropriately diluted recombinant 2S albumin (Fig. 4B). Three randomly selected seeds from the wild-type cultivar 'Harunoibuki' (WT) demonstrated that the major band of g14 accumulated at the highest level (~60 μg/seed) among the four types of 2S albumin and g11/g28 at ~20 μg/seed and g13 at ~1.5 μg/seed on average. The migration of g13 was faster than that of g13r, suggesting that almost all g13 in seeds was post-translationally processed (i.e. processing to a ~10 kDa band from a 16 kDa band). In addition, the variation in the accumulation level of g13 was relatively high in the wild-type seeds.
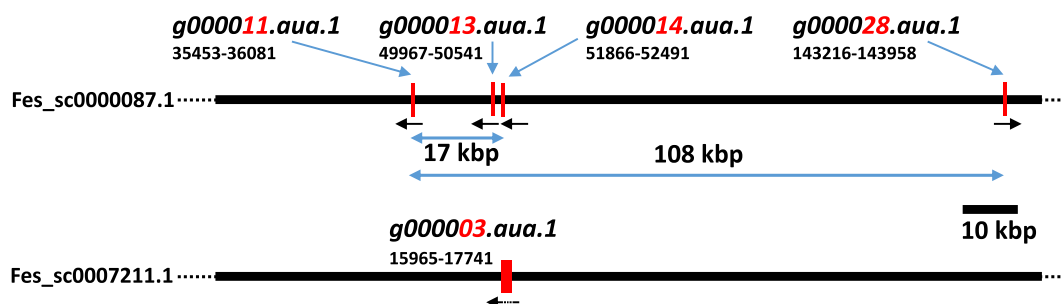

Fig. 1. 2S albumin genes identified in Buckwheat Genome DataBase (BGDB). 2S albumin genes (*g000003.aua.1, g03; g000011.aua.1, g11; g000013.aua.1, g13; g000014.aua.1, g14; g000028.aua.1, g28*) were indicated with red filled box. Arrows indicate the direction of the genes. Approximate distances between the genes were shown.
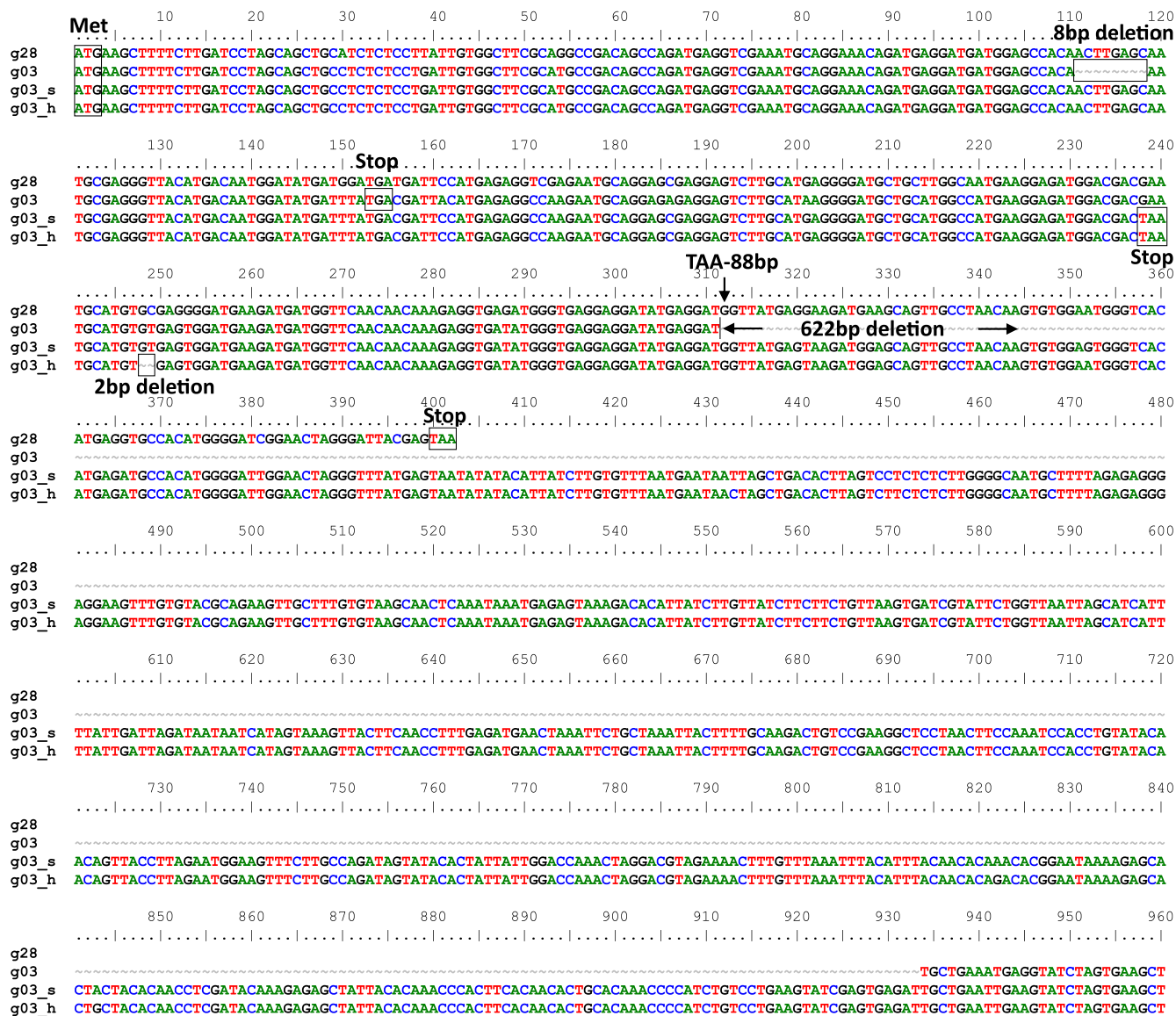
Fig. 2. Alignment for new alleles of the 2S albumin gene *g03*. Newly identified alleles of *g03* (*g03_s* and *g03_h*) were compared with the registered alleles of *g03* and *g28* in the BGDB database. Translation initiation (ATG) of the *g03* in the BGDB was corrected. Translation initiation, predicted termination (TGA, TAA), and deletion were highlighted with a black box. The position of 622 bp deletion was also indicated.

3.5. Hydrophobicity of 2S albumin and modification and domain exchange of g13

The hydrophobicity profile calculated by Kyte & Doolittle (BioEdit version 7.2.5) with a scan window size of 13 demonstrated that g13 is more hydrophobic than the other 2S albumin polypeptides (g14, g11, and g28) in almost the entire region, particularly in the *N*-terminus region (Fig. S2A). The modification of some amino acids (I14G, L19G, and M21G) (Fig. S3A) in the *N*-terminus region resulted in markedly lower hydrophobicity than that of the original g13 (Fig. S2C). Meanwhile, the three-dimensional (3D) structure of g13 predicted by Phyre2 (https://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) suggests that g13 has five alpha-helices (Fig. S4), reported as Ia, Ib, II, III, and IV (Moreno and Clemente, 2008). Thus, HindIII and *Kpn*I sites were introduced between the 2nd and 3rd and between the 4th and 5th alpha helices, respectively, and the three domains were exchanged between g13 and g14 (Fig. S3). The alteration of hydrophobicity by the introduction of HindIII and *Kpn*I sites and domain exchange was confirmed (Fig. S2B, S2D, S2E, and S2F).

3.6. MWa of 2S albumin and modified- and domain exchanged- g13

The four recombinant 2S albumin (g11r, g28r, g13r, and g14r), *N*-terminus-modified g13 (g13′), and g13/g14 domain exchange series from Nos.1 to 8 were electrophoresed (Fig. 5A), and the migration distance of each band was quantified using ImageJ (U. S. National Institutes of Health, https://imagej.nih.gov/ij/) (Table S3). Comparison of molecular weight estimated from SDS-PAGE migration (MWa, n = 3) and that calculated from amino acid sequences demonstrated that the migration variance between g11r, g28r, and g14r was almost entirely explained by the difference in their molecular weights (calculated); however, the relationship between the two MWs in the domain exchange series and the modified g13 (g13′) including g14r was different from the regression line y = x (Fig. 5B). The MWa of g13r was considerably less than the calculated MW. The effect of the 1st and 3rd domains were additive and more relevant than the 2nd domain in the migration variance between g13 and g14. Note that g13 prepared from seeds migrated greatly faster than g13r, as described (Fig. 4B), probably because of the post-translational processing. However, the processing enzyme responsible for post-translational processing does not exist in
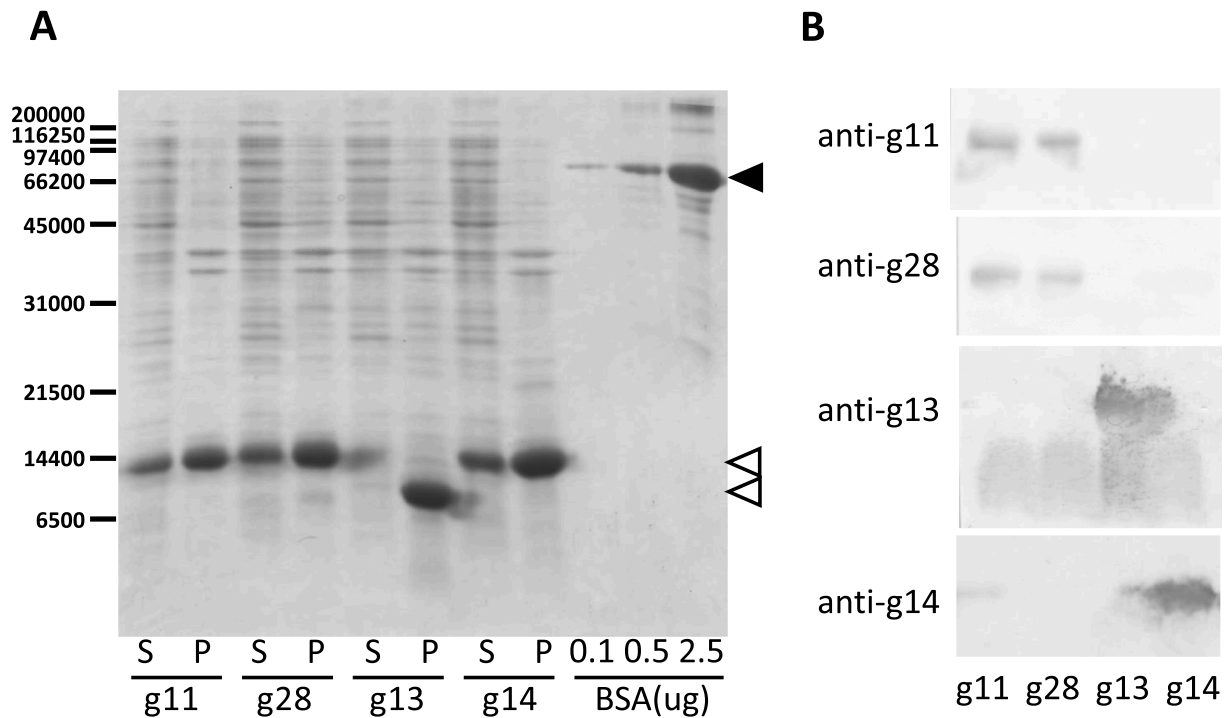
Fig. 3. Immunoreactivity of the representative recombinant 2S albumin prepared in *E. coli* cells. (A) The representative alleles of 2S albumin genes (g11, g13, g14, g28) were expressed in *E. coli* cells. After disrupting the cells, soluble (S) and insoluble (P) proteins were separated and detected using SDS-PAGE with CBB staining. For a quantification, designated amounts of BSA were also applied. White and black triangles indicate the position of 2S albumin and BSA, respectively. (B) The reactivity of anti 2S albumin antibodies were examined against the recombinant 2S albumin prepared in *E. coli* cells (g11, g28, g13, g14).
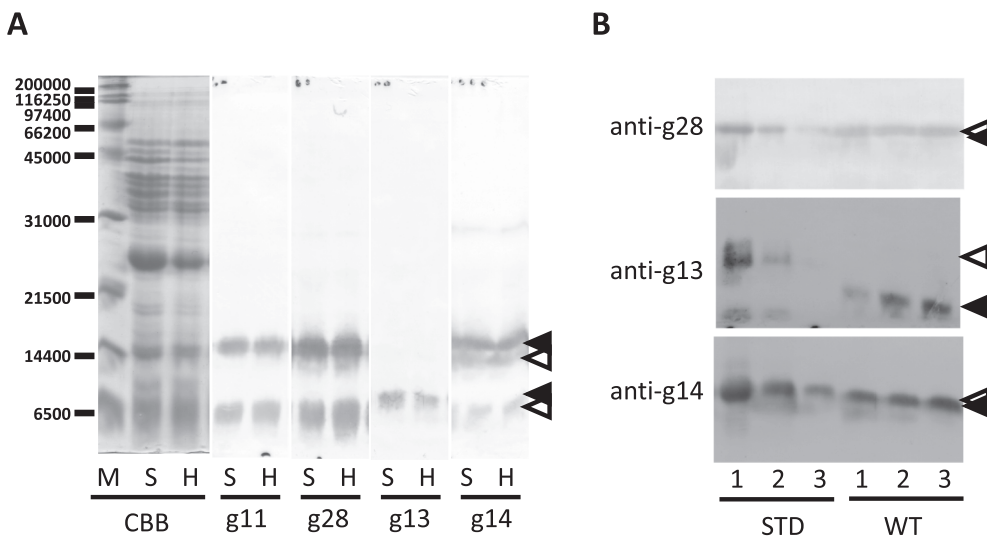

Fig. 4. Detection and quantification of 2S albumin in common buckwheat seeds. (A) Albumin fractions prepared from bulked seeds flour of cultivars 'Shinano 1' (S) and 'Harunoibuki' (H) were detected with CBB and antibodies against g11, g28, g13, and g14. Black and white triangles indicate the position of major and minor processed 2S albumin, respectively. Note that anti-g11 and anti-g28 were reactive with g11 and g28 and vice versa. (B) Accumulation levels of 2S albumin g11/g28, g13, and g14 were analyzed using appropriately diluted recombinant 2S albumin (STD 1–3, see Materials and Methods for the dilution ratio). Three randomly selected seeds from the wild type cultivar 'Harunoibuki' (WT) were examined. White and black triangles indicate the position of recombinant and native 2S albumin, respectively.

E. coli cells.

4. Discussion

4.1. First measurement for accumulation of individual 2S albumin in buckwheat seeds

The measurement of allergen content in seed crops is crucial for developing reliable assessment methods for allergenic materials or for producing hypoallergenic seed crops. In buckwheat seeds, it has been suggested that only four (g11, g28, g13, and g14) out of the five 2S albumin (responsible for anaphylaxis) genes are functional. However, the differences or similarities between the four gene products are not well known. Thus, we analyzed gene products prepared from buckwheat seeds and *E. coli* cells (recombinant, r) individually for the first time. Our results clearly demonstrated that the content of the major bands of ~16 kDa of g14, equivalent to known allergen Fag e 2, was the highest (~60 µg/seed), followed by g11/g28 (~20 µg/seed), and g13 (~1.5 µg/seed) in average. The representative allele of the four genes suggested that all four gene products contained the eight conserved cysteine residues responsible for the four disulfide bonds and three conserved regions A, B, and C, which are typical characteristics of 2S albumin (Shewry et al., 1995). The four disulfide bonds are supposed to contribute to the maintenance of the tertiary structure (José-Estanyol et al., 2004) and the rigid and proteinase-resistant structure of 2S albumin, resulting in highly similar alpha-helical structures that are stable to both thermal
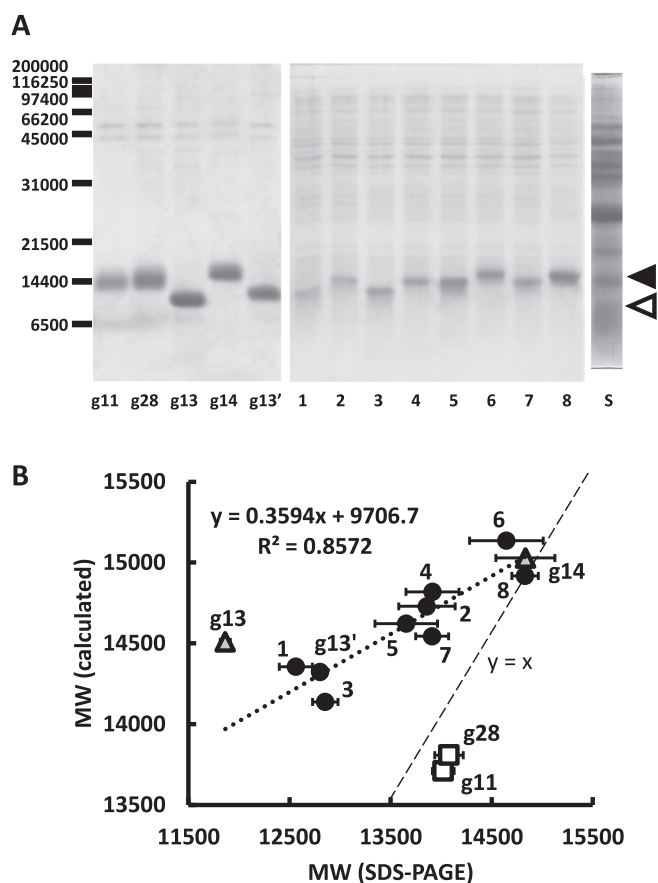
A

B

Fig. 5. Comparison of apparent and calculated molecular weights of modified and domain-exchanged 2S albumin. (A) The four recombinant 2S albumin (g11r, g28r, g13r, g14r), *N*-terminus modified g13 (g13'), and g13/g14 domain exchange series from No.1 to No.8 were prepared in *E. coli* expression system and electrophoresed with buckwheat seed proteins (S) in SDS-PAGE. Black and white triangles indicate the position of 16 kDa and 8–10 kDa 2S albumin, respectively. (B) Molecular weight estimated from SDS-PAGE migration (n = 3) and that calculated from amino acid sequence were compared. The regression line was drawn for all the symbols except g13, g11, and g28 with a dotted line. Dashed line indicates y = x.

and proteolytic denaturation (Breiteneder and Mills, 2005).Therefore, all four gene products have the potential to be allergens, but the most abundant polypeptide, g14, might have the greatest impact on inducing an allergenic response. More detailed analyses, not only on g14 (Fag e 2) but also on the other three polypeptides (g11, g28, and g13) in allergenicity will be required.

4.2. Difference and similarity between 8 and 16 kDa 2S albumin

2S albumin is a member of the prolamin superfamily and has a conserved cysteine skeleton (Radauer and Breiteneder, 2007). 2S albumin is widely distributed in both mono- and di-cotyledonous plants (Moreno and Clemente, 2008). The proteins are synthesized as a single larger precursor polypeptide in the range of ~18–21 kDa and the precursor is processed to a polypeptide in the range of ~12–14 kDa in the vacuole and eventually to the large subunit in the range of ~8–10 kDa and the small subunit in the range of ~3–4 kDa (Shewry et al., 1995). In our study, g11/g28 ranging from 40 to 50 % and less than 20 % of g14 were processed into large and small subunits. Such differences in the processing ratio as well as in the accumulation level could be a possible reason why g11/g28 is historically known as 8/10 kDa and g14 is known as 16 kDa 2S albumin. In the case of g13, we could not estimate the ratio

at which the processing occurred because the recombinant g13 (without a signal peptide, but no further processing) exhibited an unusually low MWa of ~12 kDa, possibly because of its hydrophobicity and specific amino acid nature. However, the MWa of g13 in seeds was less than that of g13r, implying that almost all g13 in seeds was processed. The *N*-terminal amino acid sequencing of g13 in seeds would be useful for clarifying the factors underlying these unique characteristics.

Alterations in apparent molecular size were also observed between reducing and non-reducing conditions for 2S albumins (Sharma, Mundoma, Seavy, Roux, & Sathe, 2010). Sen et al. (2002) demonstrated that Ara h 2, a peanut 17 kDa allergen (2S albumin) possessing eight cysteine residues that could form four disulfide bonds, exhibited doublet bands with a mean MWa of ~12 kDa under native (non-reducing) and ~17 kDa under reducing conditions in SDS-PAGE. Thus, it cannot be excluded that the different disulfide bridge formation of g13 from the other 2S albumin is responsible for the low MWa.

4.3. Development of hypoallergenic buckwheat plants

2S albumin is pepsin-resistant. Tanaka et al. (2002) demonstrated that various sizes of IgE positive bands from less than 6.5 kDa to greater than 16.5 kDa were detected in immunoblot of pepsin-digested and undigested buckwheat proteins from the sera of patients exhibiting immediate hypersensitivity reaction (anaphylaxis). The diversity of the detected bands, varying in size and intensity depending on the patient, suggests that all 2S albumin polypeptides (g11, g28, g13, and g14), regardless of whether they are processed or unprocessed, can become allergens. Large allelic variations preserved by the heterogeneous nature of outcrossing buckwheat and other post-translational modification may also produce bands of various sizes, complicating the development of hypoallergenic buckwheat. The accumulation level of g13 highly varied by seed (Fig. 4B), suggesting that hypoallergenic plants with reduced accumulation of g13 can be developed.

5. Conclusion

Characterization of allergens is prerequisite for developing assessment methods for allergenic materials and for producing hypoallergenic crops. Four functional 2S albumin genes and their polypeptides of common buckwheat were preliminarily analyzed to show differences in structure, accumulation level, and apparent molecular size. However, further detailed analyses on the individual polypeptides are necessary to overcome this allergenic problem.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

FIM and TKT conceived, designed, and conducted the experiments. FIM and TKT also analyzed the data and wrote the manuscript. All authors have read and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.

org/10.1016/j.fochms.2022.100127.

References

Altenbach, S. B., Pearson, K. W., Leung, F. W., & Sun, S. S. (1987). Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. *Plant Molecular Biology, 8*, 239–250. https://doi.org/10.1007/BF00015032

Breiteneder, H., & Mills, E. N. C. (2005). Molecular properties of food allergens. *Journal of Allergy and Clinical Immunology, 115*, 14–23. https://doi.org/10.1016/j.jaci.2004.10.022. ISSN 0091-6749.

Choi, S. Y., Sohn, J. H., Lee, Y. W., Lee, E. K., Hong, C. S., & Park, J. W. (2007a). Application of the 16-kDa buckwheat 2S storage albumin protein for diagnosis of clinical reactivity. *Annals of Allergy, Asthma & Immunology, 99*, 254–260.

Choi, S. Y., Sohn, J. H., Lee, Y. W., Lee, E. K., Hong, C. S., & Park, J. W. (2007b). Characterization of buckwheat 19-kD allergen and its application for diagnosing clinical reactivity. *International Archives of Allergy and Immunology, 144*, 267–274.

Geiselhart, S., Nagl, C., Dubiela, P., Pedersen, A. C., Bublin, M., Radauer, C., ... Mortz, C. G. (2018). Concomitant sensitization to legumin, Fag e 2 and Fag e 5 predicts buckwheat allergy. *Clinical and Experimental Allergy, 48*, 217–224.

Guzmán, R., Lemus, J. A., Sastre, B., Del Pozo, V., & Senent, C. J. (2020). Buckwheat: An emerging allergen. *Journal of Investigational Allergology and Clinical Immunology, 30*, 74–75. https://doi.org/10.18176/jiaci.0452

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.

José-Estanyol, M., Gomis-Rüth, F. X., & Puigdomènech, P. (2004). The eight-cysteine motif, a versatile structure in plant proteins. *Plant Physiology and Biochemistry, 42*, 355–365. https://doi.org/10.1016/j.plaphy.2004.03.009. PMID: 15191737.

Katsube-Tanaka, T. (2016). Chapter five - Buckwheat production, consumption, and genetic resources in Japan. In M. Zhou, I. Kreft, S.-H. Woo, N. Chrungoo, G. Wieslander (Editors), Molecular breeding and nutritional aspects of buckwheat, Academic Press, 2016, Pp. 61-80, ISBN 9780128036921, https://doi.org/10.1016/B978-0-12-803692-1.00005-5.

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.

Koyano, S., Takagi, K., Teshima, R., & Sawada, J. (2006). Molecular cloning of cDNA, recombinant protein expression and characterization of a buckwheat 16-kDa major allergen. *International Archives of Allergy and Immunology, 140*, 73–81.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*, 680–685.

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

Nair, A., & Adachi, T. (1999). Immunodetection and characterization of allergenic proteins in common buckwheat (*Fagopyrum esculentum*). *Plant Biotechnology, 16*, 219–224.

Norbäck, D., & Wieslander, G. (2021). A review on epidemiological and clinical studies on buckwheat allergy. *Plants, 10*, 607. https://doi.org/10.3390/plants10030607

Radauer, C., & Breiteneder, H. (2007). Evolutionary biology of plant food allergens. *The Journal of Allergy and Clinical Immunology, 120*, 518–525. https://doi.org/10.1016/j.jaci.2007.07.024

Radovic, R. S., Maksimovic, R. V., Brkljacic, M. J., Varkonji Gasic, I. E., & Savic, P. A. (1999). 2S albumin from buckwheat (*Fagopyrum esculentum* Moench) seeds. *Journal of Agricultural and Food Chemistry, 47*, 1467–1470.

Satoh, R., Jensen-Jarolim, E., & Teshima, R. (2020). Understanding buckwheat allergies for the management of allergic reactions in humans and animals. *Breeding Science, 70*, 85–92. https://doi.org/10.1270/jsbbs.19051

Sen, M., Kopper, R., Pons, L., Abraham, E. C., Burks, A. W., & Bannon, G. A. (2002). Protein structure plays a critical role in peanut allergen stability and may determine immunodominant IgE-binding epitopes. *Journal of Immunology, 169*, 882–887. https://doi.org/10.4049/jimmunol.169.2.882

Sharma, GM, Mundoma, C, Seavy, M, Roux, KH, & Sathe, SK (2010). Purification and biochemical characterization of Brazil nut (*Bertholletia excelsa* L.) seed storage proteins. *J Agric Food Chem, 58*(9), 5714–5723. https://doi.org/10.1021/jf9036326. PMID: 20405841.

Shewry, P. R., Napier, J. A., & Tatham, A. S. (1995). Seed storage proteins: Structures and biosynthesis. *The Plant Cell, 7*, 945–956. https://doi.org/10.1105/tpc.7.7.945

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*, 49–56.

Teuber, S., Dandekar, A., Peterson, W., & Sellers, C. (1998). Cloning and sequencing of a gene encoding a 2S albumin seed storage protein precursor from English walnut (*Juglans regia*), a major food allergen. *Journal of Allergy and Clinical Immunology, 101*, 807–814.

Viquez, O. M., Summer, C. G., & Dodo, H. W. (2001). Isolation and molecular characterization of the first genomic clone of a major peanut allergen, Ara h 2. *Journal of Allergy and Clinical Immunology, 107*, 713–717.

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*, 215–224.