Insertion of ten amino acids into 13S globulin zero-repeat subunit improves trypsin digestibility in common buckwheat (Fagopyrum esculentum Moench) seeds

AUTHOR(S):

Okada, Takeyuki; Monshi, Fakhrul Islam; Kudo, Syuto; Katsube-Tanaka, Tomoyuki

CITATION:

Okada, Takeyuki ...[et al]. Insertion of ten amino acids into 13S globulin zero-repeat subunit improves trypsin digestibility in common buckwheat (Fagopyrum esculentum Moench) seeds. Food Chemistry: Molecular Sciences 2023, 6: 100159.

ISSUE DATE:

2023-07

URL:

http://hdl.handle.net/2433/285518

Food Chemistry: Molecular Sciences 6 (2023) 100159

Contents lists available at ScienceDirect

Food Chemistry: Molecular Sciences

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

Insertion of ten amino acids into 13S globulin zero-repeat subunit improves trypsin digestibility in common buckwheat (*Fagopyrum esculentum* Moench) seeds

Takeyuki Okada [a], Fakhrul Islam Monshi [a,b], Syuto Kudo [a], 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:
13S globulin
Allergen
Common buckwheat
Zero-repeat subunit
10 amino acids insert

ABSTRACT

The 13S globulin zero-repeat subunit is resistant to trypsin and may have higher allergenicity than the 1–6 tandem repeat subunits in common buckwheat (*Fagopyrum esculentum* Moench). To explore alleles useful for lowering allergenicity, amplicon deep sequencing targeting the zero-repeat subunit gene was conducted in bulked genomic DNA from eight cultivars and landraces. The analysis identified a unique allele encoding a zero-repeat subunit with 10 amino acid insertion (10aa) at a position equivalent to the tandem repeat insertion. Prediction of its 3-D structure suggested that 10aa changes the β-hairpin structure in the non-10aa (native) subunit to a random coil, which is also found in 1- and 3- repeat subunits. Homozygotes of the 10aa allele were developed and showed that the 10aa subunit was more digestible than the native subunit. However, the 10aa subunit was still less digestible than the 1–6 repeat subunits, suggesting needs to explore unfunctional alleles.

1. Introduction

Buckwheat is a historically important crop in not only the eastern and southern Asia but also the central and eastern Europe, owing to its beneficial features, such as rapid growth rates, low agronomical costs, and high adaptability to infertile lands (Katsube-Tanaka, 2016). Buckwheat is traditionally eaten as noodles, jellies, pancakes, porridge, and pasta. It is also consumed in western countries for gluten-free diets (Heffler et al., 2014; Wieslander & Norbäck, 2001). Although buckwheat has nutritionally and physiologically beneficial properties, such, anti-hypercholesteremia, anti-hypertension, anti-inflammation, anti-carcinogenesis, and a balanced amino acid composition (Giménez-Bastida & Zieliński, 2015; Huda et al., 2020; Zhang et al., 2012), allergy issues remain to be resolved (Katsube-Tanaka & Monshi, 2022).

The most significant allergens in common buckwheat are Fag e 1 and Fag e 2, which are major seed storage proteins, 13S globulin and 2S albumin, respectively (Katayama et al., 2018). 13S globulin belongs to the 11S globulin superfamily, which is typically a hexameric protein that is widely found in seed plants (Adachi et al., 2003). Like 11S globulins, 13S globulin subunit consists of two polypeptides, α (acidic) and β (basic) (Radović et al., 1996) which are translated as a single pre-proprotein containing a signal peptide (Fig. 1). The signal peptide is removed during translocation across the endoplasmic reticulum membrane, and then the proprotein is processed into α- and β-polypeptides, which are disulfide-bonded in a mature form. The α polypeptides have larger variations in size and charge than β polypeptides (Khan et al., 2008a, 2008b). 13S globulin contains methionine (Met)-rich and Met-poor subunits. The Met-poor subunits have different number (zero and 1–6) of tandem repeats inserted in α polypeptides, forming zero-repeat and 1–6-repeat subunits. The tandemly repeated insert, consisting of a unit of 15 amino acid residues, is hydrophilic and arginine-rich, and thus, susceptible to trypsin digestion (Khan et al., 2012).

In previous studies, 17 open reading frames for 13S globulin, including *GlbNA* and *GlbNB* encoding trypsin-resistant zero-repeat subunits were isolated (Sano et al., 2014), using the genomic DNA (gDNA) library developed from a cross between common buckwheat strain numbers XIF1999 and *dw*E1999 by Yasui et al. (2008). The other open reading frames included 12 types of 1–4-repeat subunit genes and 3 types of methionine-rich subunit genes. Furthermore, for the zero-repeat subunit, six new alleles (*GlbNA1*, *GlbNC1*, *GlbND*, *GlbNA2*, *GlbNA3*, and *GlbNC*) were identified in elite Japanese cultivars and Pakistani landraces (Katsube-Tanaka et al., 2014; Monshi et al., 2020). Notably, the identified alleles, *GlbNA2*, *GlbNA3*, and *GlbNC* contain an about 200-bp-long sequence like a miniature inverted-repeat transposable element

* Corresponding author.
E-mail address: tanakato@kais.kyoto-u.ac.jp (T. Katsube-Tanaka).

https://doi.org/10.1016/j.fochms.2022.100159
Received 2 November 2022; Received in revised form 12 December 2022; Accepted 24 December 2022
Available online 26 December 2022

(MITE). Tandem repeats and MITE-like insertion can change the function of the 13S globulin zero-repeat subunit. However, useful alleles of 13S globulin for hypoallergenic buckwheat production have not been identified yet.

Recently, unfunctional alleles were isolated in standard cultivars of common buckwheat for 2S albumin g03 and g13 (Katsube-Tanaka & Monshi, 2022; Monshi & Katsube-Tanaka, 2022), suggesting that common buckwheat could help conserve extensive genetic diversity because of its outcrossing nature. In the present study, we hypothesized that novel allele of the 13S globulin zero-repeat subunit, that is useful for developing hypoallergenic buckwheat, can be found in natural common buckwheat populations. Therefore, we employed amplicon deep sequencing to evaluate the genetic diversity and analyzed the structure and function of isolated zero-repeat subunits by comparing them with the higher-order structures of 11S globulin.

2. Materials and methods

2.1. Buckwheat flours and DNA extraction

Four common buckwheat varieties from Japan, Miyazaki Wase Kaori, Botansoba, Shinano 1, and Harunoibuki and four common buckwheat landraces from Bangladesh, Nepal, Pakistan, and Hokkaido (Japan) were subjected to the extraction of gDNA (Monshi & Katsube-Tanaka, 2022). Buckwheat flours were made by independently grinding matured seeds (>15 g). Seed flour (4 g each of the eight cultivars and landraces) was used for gDNA extraction with the DNAs-ici!-S reagent (Rizo, Tsukuba, Japan). Because common buckwheat plant has an outcrossing nature and is heterogeneous within a single cultivar, large amount of seed flour was processed to retain the genetic diversity.

2.2. Amplicon deep sequencing

The coding region of the zero-repeat subunit gene was divided into four regions, 0rep-1 to 0rep-4 (Fig. 1). Each gDNA region was amplified using PCR primers (Table 1) designed to overlap the terminals of the amplified fragments. Pre-PCR was conducted with KOD FX neo DNA polymerase (TOYOBO, Osaka, Japan) and the following PCR conditions: one cycle of 94 °C for 2 min, 30 cycles of 98 °C for 10 s, 63 °C for 30 s, 68 °C for 30 s, and one cycle of 68 °C for 7 min. The first and second PCRs were performed with ExTaq DNA polymerase (TaKaRa, Kusatsu, Japan) and the following PCR conditions: one cycle of 94 °C for 2 min, 25 cycles (1^{st} PCR) or 10 cycles (2^{nd} PCR) of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s, and one cycle of 72 °C for 5 min. The concentrations of the 1^{st} and 2^{nd} PCR products were measured using a BioTek Synergy H1 microplate reader (Agilent, Santa Clara, CA, USA) and QuantiFluordsDNA System (Promega, Madison, WI, USA). Library quality was measured using a Fragment Analyzer and dsDNA 915 Reagent Kit [Advanced Analytical Technologies (Agilent), Santa Clara, CA, USA]. Amplicon deep sequencing was performed using MiSeq (2 × 300 bp, Illumina, San Diego, CA, USA). Quality filtering and merging of reads were conducted with the FASTQ/A Barcode splitter of FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html), sickle tools, and FLASH (http://www.cbcb.umd.edu/software/flash). The amplicon deep sequencing was performed only once without biological or technical replicates. The oligonucleotides for PCR analysis were column-purified grade and obtained from a DNA synthesis company (Eurofins Genomics, Tokyo, Japan).

2.3. Development of 10aa homozygote line

The 10 amino acid insertion (10aa) allele-containing plants were screened out in Harunoibuki using PCR. Crude gDNA was prepared from a leaf disk punched with a 2-cc microtube lid and ground using MixerMill (QIAGEN K.K., Tokyo, Japan) in 100 μL of DNA extraction buffer (0.1 M Tris pH 9.5, 1 M KCl, and 10 mM EDTA pH 8.0). The genotypes were determined using the following PCR primers: 0rep_137Ins10_F, 5'-TCACCACGACAAACAATCTTCGGG-3' and M13R_0rep_4R_new, 5'-caggaaacagctatgacAACGACGTCGTATCTCTC-3' for the 10aa allele and GlbNAB_N, 5'-TACCTCCGCCACCACGAAGGC-3' and 0rep_137normal-R, 5'-GGTGCTCATCGCTTATCGATCTTG-3' for the native allele. The primer 0rep_137Ins10_F anneals to the nucleotide sequence encoding 10aa and the primer 0rep_137normal-R anneals to the nucleotide sequence before and after that encoding 10aa. The 10aa allele-containing plants were naturally pollinated in isolated areas to enrich the frequency of the 10aa allele. Successive generations were also genotyped and screened for 10aa allele homozygotes.

2.4. Extraction of protein and gDNA from a single seed

A single seed was crushed using MixerMill (QIAGEN K.K., Tokyo, Japan); albumin and globulin fractions were extracted using 300 μL of protein extraction buffer (35 mM potassium phosphate buffer pH 7.6, 0.4 M NaCl). After protein extraction, gDNA was extracted from the residue using 300 μL of DNAs-ici!S (Rizo, Tsukuba, Japan).

2.5. Digestion of extracted protein with trypsin

The extracted protein samples [50 μL; 10aa heterozygote and a mixture of 10aa homozygote and non-10aa (native) homozygote] were digested in vitro with 2.5 μL of 50 mM NH_4HCO_3, 40.5 μL of sterilized distilled water, and 6.4 μL of trypsin (2.5 g/L-Trypsin Solution #35555-54, Nacalai, Kyoto, Japan) at 37 °C for up to 8 h. During and after the digestion, 25 μL of the sample solution was mixed with 25 μL of 2 × SDS buffer [100 mM Tris pH 6.8, 4% (w/v) SDS, 20% (w/v) glycerol], 2.5 μL of 2-mercaptoethanol, and 2.0 μL of 0.1% (w/v) bromophenol blue, followed by the digestion termination by boiling (100 °C, 5 min). Trypsin digestion was repeated using at least six biologically

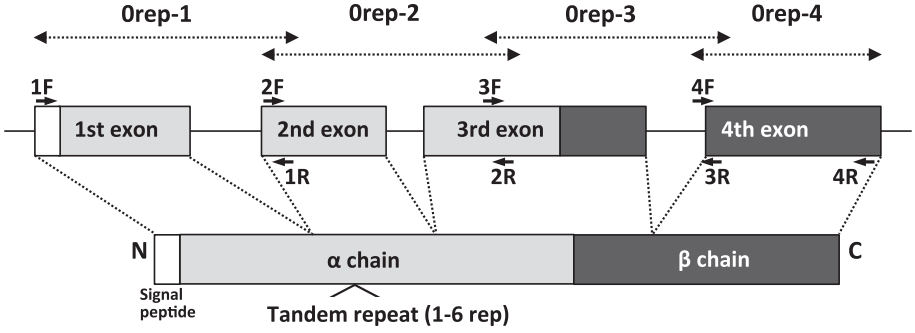


Fig. 1. Schematic representation of 13S globulin zero-repeat subunit and its genomic DNA. Boxes represent exons for gDNA and polypeptides for protein. Short horizontal arrows indicate PCR primers for amplicon deep sequence. Dotted horizontal arrows indicate four amplified regions. The equivalent position for the tandem repeat insertion of 1–6 repeat subunits is indicated in the α polypeptide chain.

Table 1
PCR primers used in amplicon deep sequencing.

Pre-PCR	M13F_0rep_1F	gtaaaacgacggccagtCACTAAGCCACCAACATG	M13R_0rep_1R	caggaaacagctatgacGAAACTCCGATTGGAAGG
	M13F_0rep_2F	gtaaaacgacggccagtGTTATACCAGGATGTCCG	M13R_0rep_2R	caggaaacagctatgacTCCGAATCTTGCTGATCG
	M13F_0rep_3F	gtaaaacgacggccagtGTATTCAGAGGTGCTGAC	M13R_0rep_3R	caggaaacagctatgacCACGTTCCATTTAGGTCC
	M13F_0rep_4F	gtaaaacgacggccagtAACCGAGTGAAkGATCGG	M13R_0rep_4R_new	caggaaacagctatgacAACGACGTCGTATCTCTC
1st PCR	1st-FagopyrumF	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTAAAACGACGGCCAGT	1st-FagopyrumR	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGGAAACAGCTATGAC
2nd PCR*	2ndF	AATGATACGGCGACCACCGAGATCTACAC-Index2-ACACTCTTTCCCTACACGACGC	2ndR	CAAGCAGAAGACGGCATACGAGAT-Index1-GTGACTGGAGTTCAGACGTGTG

*Index1: TGCCTCTT. Index2: AGGCGAAG for 0rep-1, TAATCTTA for 0rep-2, CAGGACGT for 0rep-3, and GTACTGAC for 0rep-4.

independent materials in a single technical replicate. The band intensity was measured using ImageJ software (U. S. National Institutes of Health, https://imagej.nih.gov/ij/).

2.6. SDS-PAGE and western blotting

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was conducted according to Laemmli (1970) using a 14% T acrylamide gel at 200 V for 60 min. Western blotting and immuno-detection were performed according to Khan et al. (2012). The primary antibodies against the α-polypeptide of zero-repeat subunit (epitope: CSSKHQESPEGFIGG, GlbNA, GenBank: AB828117.1) and against that of 1–6 repeat subunits (epitope: CEEGSDRQSRESDDDEALLE, 5rep, GenBank: D87980.1) were used (Sigma-Aldrich Japan, Tokyo, Japan). The secondary antibody was goat anti-rabbit IgG alkaline phosphatase conjugate (Promega, Madison, WI, USA). Unless otherwise mentioned, all chemical reagents were analytical grade and obtained from a chemical reagent manufacturer (Nacalai Tesque, Kyoto, Japan).

2.7. Statistical analyses

Tukey's test was employed to evaluate the difference in mean values of the relative content between the 10aa and non-10aa (native) bands after trypsin digestion. Six biologically independent samples were analyzed with a single technical replicate. The statistical analysis software Bell Curve for Excel (ver. 3.21, Social Survey Research Information, Tokyo, Japan) was used.

3. Results

3.1. Amplicon deep sequence

The gene of 13S globulin zero-repeat subunit consists of four exons (Fig. 1), spanning 1722 bp (*GlbNA*, GenBank accession no. AB828117). Four regions (0rep-1 to 0rep-4) were amplified for amplicon deep sequencing to examine the allelic diversity of the gene in four elite Japanese cultivars adapted to the northern (Botansoba), central (Shinano 1), and southern (Miyazaki Wase Kaori and Harunoibuki) regions in Japan, as well as diversified landraces obtained from Bangladesh, Nepal, Pakistan, and Hokkaido (Japan). A large number of raw reads (37,000–45,000), more than 86% of which had >99.9% accuracy [Q30 (%)], were obtained. The deduced amino acid sequences showed that the most dominant allele accounted for 42%, 29%, 28%, and 35% of 0rep-1, 0rep-2, 0rep-3, and 0rep-4, respectively (Fig. 2). The top three alleles collectively shared 62–70%. In the second region (0rep-2), which covers the tandem repeat insertion site of 1–6 repeat subunits (Fig. 1), novel alleles with an extra 10 amino acid sequence were found at 8% frequency (Fig. 2). The 10 amino acids sequence was DHHHDKQSSG (Fig. 3A), the first six amino acids of which are charged, hydrophilic, and contain a lysine residue (target of trypsin).

The 3-D structure predicted by the Phyre2 program (Kelley et al., 2015) (https://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) showed that the β-hairpin at the position equivalent to the tandem repeat insertion in the native zero-repeat subunit was changed to random coils in the 10aa subunit, which was also observed in the 1-repeat and 3-repeat subunits (Fig. 3B).

3.2. Development of 10aa homozygote

PCR amplification using specific primers (Fig. 4A) for genotyping of zero-repeat subunit alleles showed ~1200 bp and ~430 bp bands for the 10aa and non-10aa (native) alleles, respectively (Fig. 4B). From the 10aa allele-enriched population, 12 and 10 seeds were obtained for the 10aa allele homozygote and heterozygote, respectively. Nine native allele homozygote seeds were obtained from Harunoibuki. Western blotting of the corresponding seed proteins with anti-GlbNA (zero-repeat subunit)

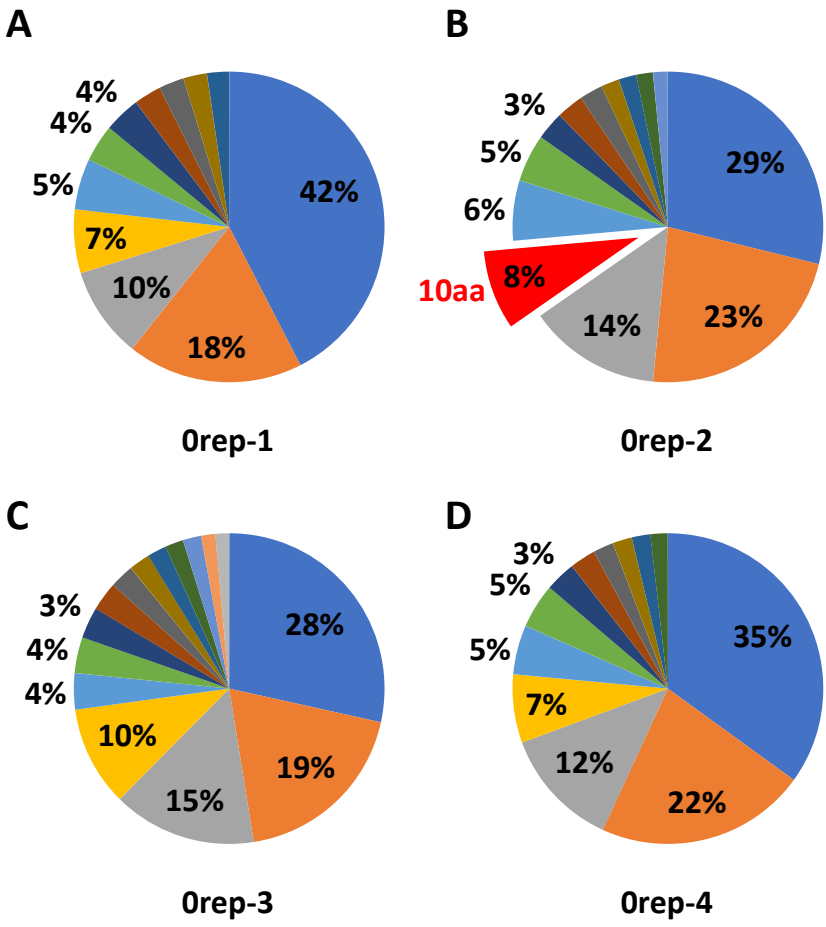


Fig. 2. Allele frequency of zero-repeat subunit gene within the four regions. Alleles with allele frequency of more than 5%, relative to that of the top allele, were compared in each of the four regions (A: 0rep-1, B: 0rep-2, C: 0rep-3, and D: 0rep-4). 10aa represents the allele containing an insertion of 10 amino acids.

demonstrated that 10aa homozygotes showed a single band with a uniform size and intensity, while native seeds showed a single band with slightly different sizes and intensities (Fig. 4C). Most heterozygotes showed two bands: a larger band derived from the 10aa subunit and a smaller band derived from the native subunit. Taken together, 10aa homozygotes were successfully developed.

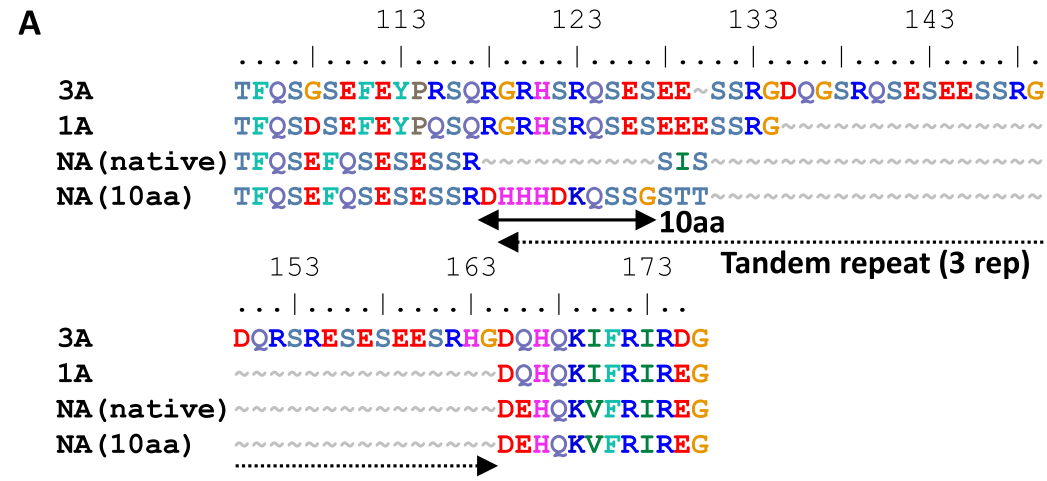
3.3. Digestibility of 10aa homozygotes and heterozygotes

To compare the trypsin digestibility of the zero-repeat subunit between the 10aa and native subunits, globulin fractions extracted from 10aa heterozygote seeds were first digested (Fig. 5A). The results showed that α polypeptides of 1–5 repeat subunits were rapidly (≤2 h) digested compared to that of the zero-repeat subunit (10aa and native), which remained undigested for 4 h. No significant difference was observed between the digestibility of the 10aa and native subunits in the 10aa heterozygotes. Meanwhile, the preparation and digestion of mixtures of the 10aa homozygotes and non-10aa (native) homozygotes showed difference in the digestibility of both the proteins (Fig. 5B). The relative content against the initial content at 0 h incubation demonstrated that ~43% of native and ~71% of 10aa bands were digested within 1 h. Approximately 64% and 84% of native and 92% and 96% of 10aa were digested within 3 and 5 h, respectively, indicating that 10aa was digested more rapidly than the native subunit (Fig. 5C).

4. Discussion

4.1. Variable / disordered regions and a β-hairpin structure in 11S globulins

The most significant difference in the predicted 3-D structure between the zero-repeat subunit [GlbNA(native)] and the 1-repeat (Glb1A) or 3-repeat (Glb3A) subunits was the presence or absence of a β-hairpin structure in the tandem repeat region (arrowheads, Fig. 3B). Utsumi et al. (1997) evaluated the modification tolerability of five variable regions (I–V) by protein engineering of soybean glycinin, a member of the 11S globulin family, to which buckwheat 13S globulin belongs. As all five variable regions exist in hydrophilic regions (i.e., the surface of protein molecules), they were considered to have little function in forming and maintaining the protein higher-order structure and to tolerate modification (Utsumi et al., 1997). They demonstrated that variable regions I, IV, and V could tolerate the deletion of ~10 amino acid residues, whereas variable regions II, III, IV, and V could not tolerate the deletion of 27–39 amino acid residues. Tandem repeat inserts of 13S globulin exist in variable region II, which corresponds to the disordered region dI of the peanut major allergen Ara h 3 (Piersma et al., 2005). Piersma et al. (2005) showed that the disordered region dI accepts proteolytic processing, as do the other disordered regions dII and dIII, yielding multiple proteolyzed bands in peanuts. Consequently, the variable region II would not tolerate a large deletion but was flexible



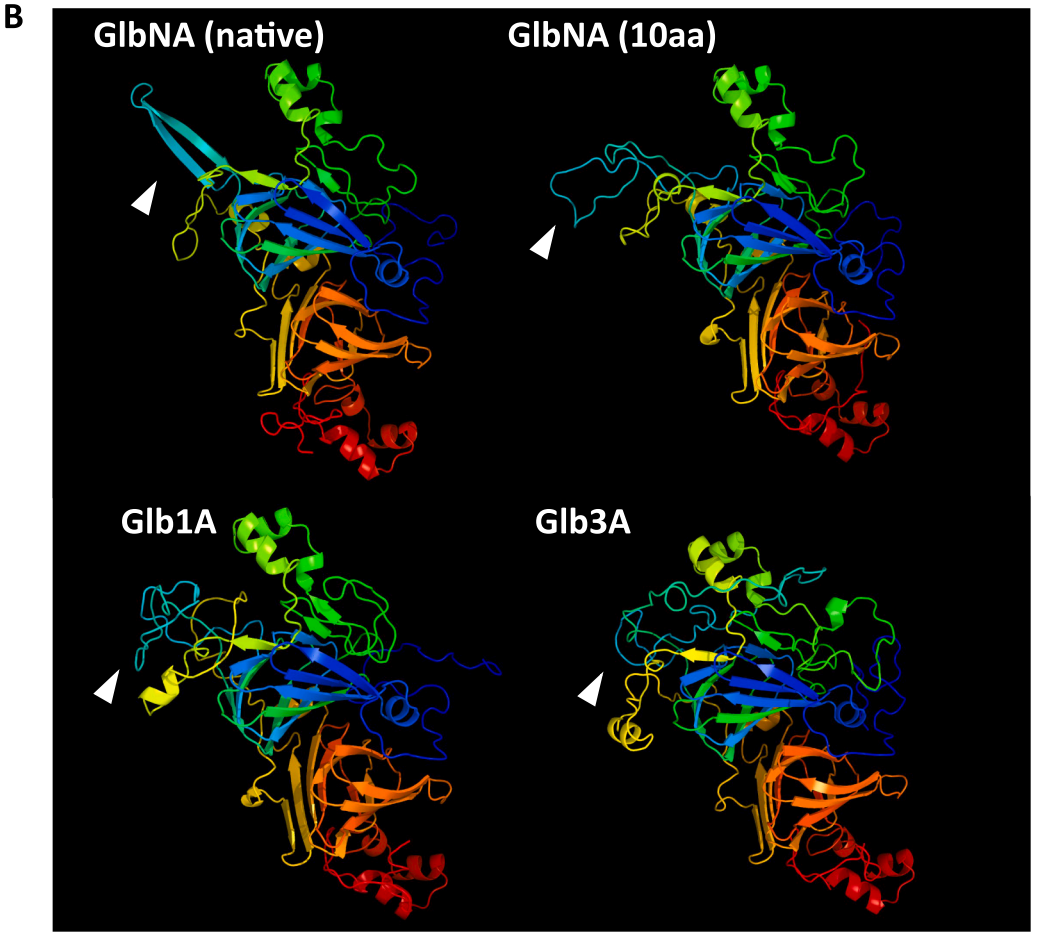


Fig. 3. Predicted 3-D structure and alignment of amino acid sequences at 10aa insertion. (A) Alignment of the α-polypeptide chain (partial). Numbers above the sequences indicate the amino acid position from the N-terminus of the mature Glb3A. Solid and dotted horizontal arrows indicate the position of the 10aa and tandem repeat insertion (3 rep), respectively. (B) The 3-D structure was predicted using the Phyre2 (Protein Homology/analogY Recognition Engine V 2.0) program using UniProtKB - W6JLU9 [GlbNA, NA(native)], W6JLL7 [Glb1A, 1A], W6JNG2 [Glb3A, 3A], and zero-repeat subunit containing 10aa [NA(10aa)] sequences.

enough to tolerate an insertion of up to 86 amino acid residues of six repeats and to accept proteolytic processing, drastically changing the apparent characteristics of protein molecules. Jin et al. (2009) analyzed the crystal structure of Ara h 3 and found a β-hairpin structure (I" and J") next to flexible loop 1 (corresponding to variable region II), suggesting that the β-hairpin structure may affect trypsin digestibility and allergenicity in 11S globulin-type allergens.

4.2. *β-hairpin structure and other factors affecting the digestibility of zero-repeat subunit*

The predicted 3-D structure of GlbNA(10aa) had no β-hairpin structure, as found in GlbNA(native) in variable region II, and formed random coils like that of 1-repeat (Glb1A) and 3-repeat (Glb3A) subunits (arrowheads, Fig. 3B). In accordance with these structural changes, the trypsin digestibility of GlbNA(10aa) was improved compared to that of GlbNA(native). However, the trypsin digestibility of zero-repeat subunits (both the 10aa and non-10aa) was still much lower than that of 1–5 repeat subunits (Fig. 5A), suggesting that factors other than the β-hairpin structure are also important for determining trypsin digestibility. Katsube et al. (1994) protein-engineered variable region IV of soybean glycinin called a hyper-variable region and demonstrated that some modifications [deletion of a polyglutamic acid sequence; IV (ΔGlu) and its substitution with polyglutamine, IV(Gln);

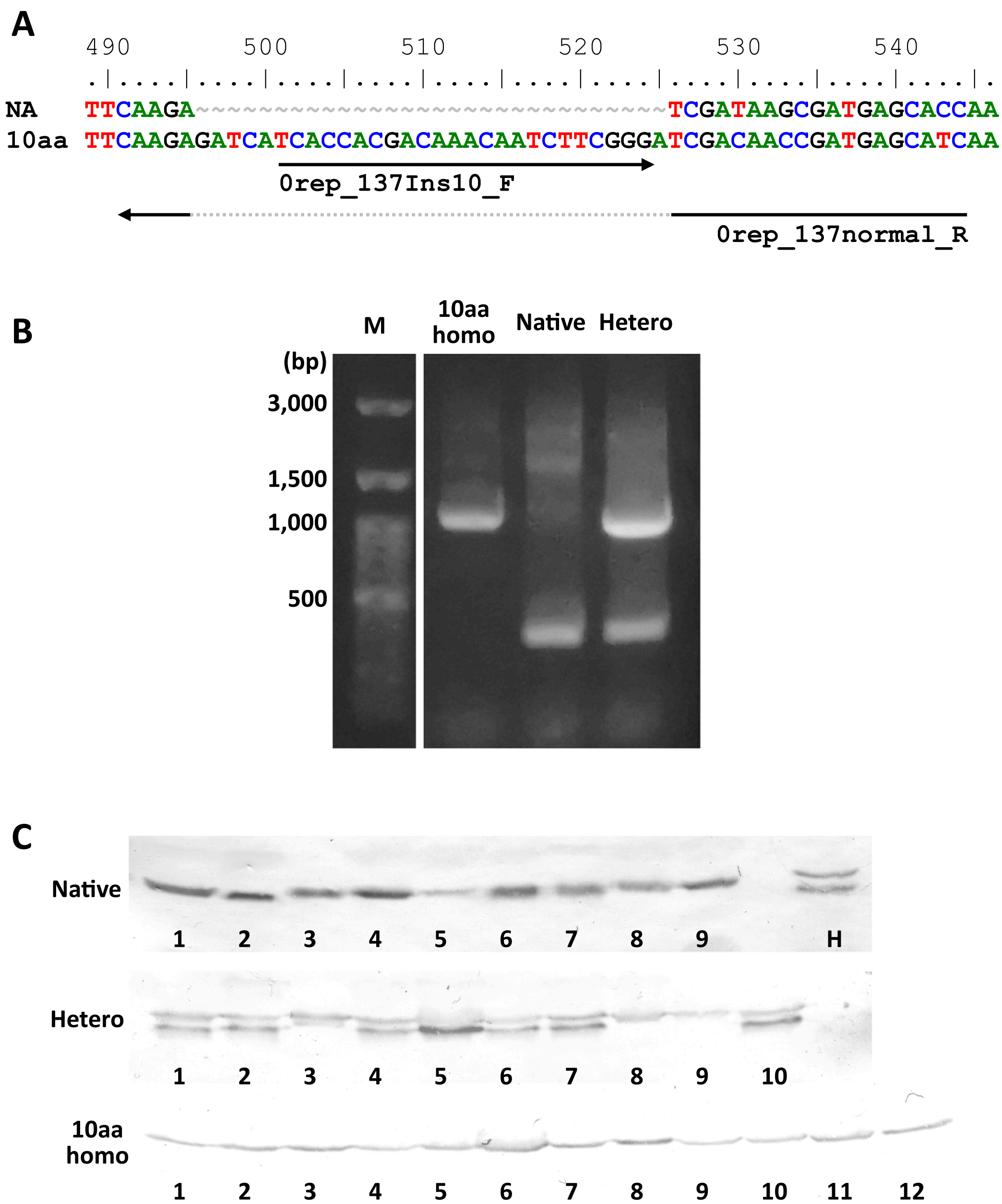


Fig. 4. Genotyping for native and 10aa alleles in zero-repeat subunit genes. (A) Nucleotide sequences of zero-repeat genes in the insertion region of 10 amino acids are compared. NA represents a native (non-10aa) genotype. Primer positions were indicated by horizontal arrows. (B) Exemplified results of genotyping for 10aa homozygotes (10aa homo), native, and heterozygotes (hetero) of zero-repeat subunit genes are shown. M indicates a size marker. (C) Western blot of native, 10aa heterozygote, and 10 aa homozygote using anti-GlbNA antibody. Numbers represent seed number. H represents 10aa heterozygote as a size marker.

polymethionine, IV(Met); and polylysine, IV(Lys) sequences] did not affect self-assembly to a trimer structure and crystallization even though limited proteolysis occurred in the region. Among the modified glycinins, only IV(Lys) crystals suffered considerably limited proteolysis. Note that the estimated pI of IV(ΔGlu), IV(Gln), IV(Met), and IV(Lys) were 7.3, 7.1, 7.0, and 8.0, respectively, suggesting that a high pI might affect limited proteolysis. Sano et al. (2014) found 17 genes encoding 13S globulin 1–4 repeat, Met-rich, and zero-repeat subunits in a buckwheat gDNA library, including one pseudogene. Among the active 16 gene subunits, the average pI of α polypeptides in 1–4 repeat, Met-rich, and zero-repeat subunits were 5.1, 4.9, and 5.9, respectively. The difference in pI, one of the important characteristics determining the physicochemical properties of a protein, might directly or indirectly affect trypsin digestibility and allergenicity.

4.3. Trypsin digestibility in heterogenous or homogenous protein molecules

In buckwheat seeds, 13S globulin subunits assemble into hexamers (Choi et al., 2006). It is unclear whether the large variation in molecular size of the tandem repeat region affects the subunit composition of the hexamer. However, the difference in trypsin digestibility between 10aa heterozygote seeds and mixtures of 10aa homozygote seeds and non-10aa (native) homozygote seeds (Fig. 5) suggests that the 10aa subunit and the non-10aa (native) subunit may be able to assemble together into a single hexamer, and the 10aa subunit in the heterologous hexamer is less digestible than that in the homozygous hexamer. Adachi et al. (2003) determined the high-resolution crystal structure of the soybean glycinin A3B4 homohexamer, but found it difficult to obtain good crystals from heterologous glycinin hexamers composed of various subunits such as A1aB1b, A2B1a, A1bB2, A3B4, and A5A4B3. Thus, the capacity of 11S globulin to assemble into a hexamer is different from that for crystallization, and likely differs from that for trypsin digestibility. Further analyses of protein structure and function and breeding focusing on zero-repeat subunits would be necessary to better understand buckwheat allergens and to create hypoallergenic buckwheat.

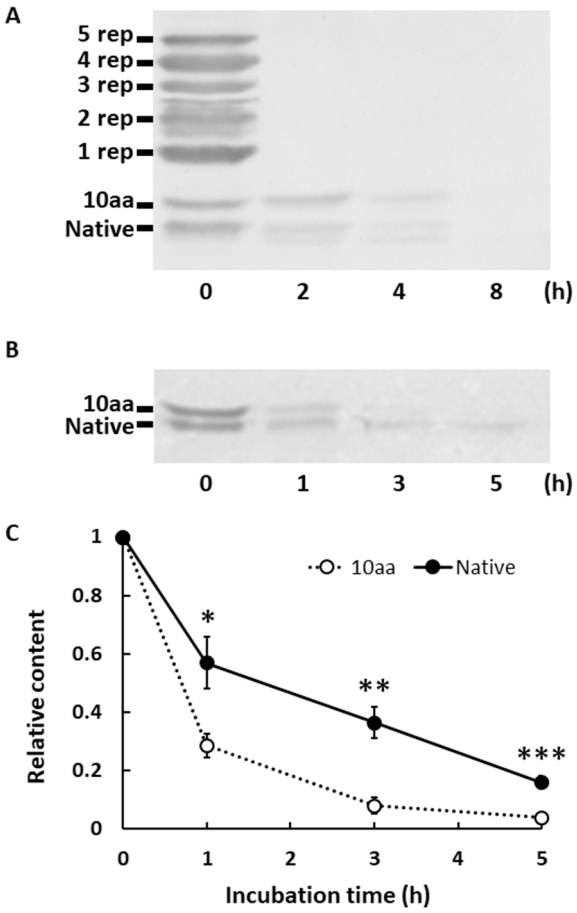


Fig. 5. Comparison in trypsin digestibility between the α polypeptides of native and 10aa zero-repeat subunits. Globulin fractions extracted from 10aa heterozygotes (A) and a mixture of 10aa homozygotes (10aa) and native subunits (Native) (B) were incubated with trypsin for 8 h. The digested samples were analyzed by SDS-PAGE and western blotting, using anti-5rep and anti-GlbNA antibodies. (A, B) Representative images of western blot data (α polypeptide of 13S globulin). Estimated molecular size is 28.3, 29.4, 33.5, 35.9, 37.4, 39.4, and 41.3 kDa for Native, 10aa, and 1–5 rep, respectively. (C) Time course of trypsin digestion in the mixture of 10aa and native subunits. Asterisks indicate significant differences between 10aa and Native at each incubation time (Student's t-test: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). n=6; error bars represent SE.

5. Conclusion

13S globulin of common buckwheat is an important allergen in buckwheat-sensitized patients. There is large variation in the molecular size of the protein caused by the insertion of tandem repeats of various lengths. As the tandem repeat is hydrophilic and contains many arginine residues, 1–6 repeat subunits containing the tandem repeat are more digestible with trypsin than zero-repeat subunits containing no tandem repeats. In this study, we found a new zero-repeat subunit, which contains additionally inserted ten amino acid residues (10aa) at the equivalent position to the tandem repeat region by amplicon deep sequencing. The 10aa subunit was predicted to possess no β-hairpin structure, which was found in the zero-repeat (non-10aa, native) subunit but not in the 1–6 repeat subunits. The 10aa subunit was more digestible with trypsin than the non-10aa (native) subunit. However, the trypsin digestibility of zero-repeat subunits, even the 10aa allele homozygote, was still much lower than that of the 1–5 repeat subunits. Thus, another new zero-repeat allele, for example, a non-functional allele, would be required to develop hypoallergenic 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 Japan Society for the Promotion of Science. Fakhrul Islam Monshi 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

The authors thank Mr. Seita Suzuki for his technical assistance. This research was partially supported by a research fund for foreign students from the Tojuro Iijima Foundation for Food Science and Technology, Japan and JSPS Research Fellowships for Young Scientists to FIM, and JSPS KAKENHI Grant Number 20H02967 to TKT.

References

Adachi, M., Kanamori, J., Masuda, T., Yagasaki, K., Kitamura, K., Mikami, B., & Utsumi, S. (2003). Crystal structure of soybean 11S globulin: glycinin A3B4 homohexamer. *Proceedings of the National Academy of Sciences of the United States of America,* 100 (12), 7395-7400. doi: 10.1073/pnas.0832158100.

Choi, S. M., Mine, Y., & Ma, C. Y. (2006). Characterization of heat-induced aggregates of globulin from common buckwheat (*Fagopyrum esculentum* Moench). *International Journal of Biological Macromolecules, 39*, 201–209. https://doi.org/10.1016/j.ijbiomac.2006.03.025

Giménez-Bastida, J. A., & Zieliński, H. (2015). Buckwheat as a functional food and its effects on health. *Journal of Agricultural Food Chemistry, 63*, 7896–7913.

Heffler, E., Pizzimenti, S., Badiu, I., Guida, G., & Rolla, G. (2014). Buckwheat allergy: An emerging clinical problem in Europe. *Journal of Allergy & Therapy, 5*, 2.

Huda, M. N., Lu, S., Jahan, T., Ding, M., Jha, R., Zhang, K., … Zhou, M. (2020). Treasure from garden: Bioactive compounds of buckwheat. *Food Chemistry, 335*, Article 127653. https://doi.org/10.1016/j.foodchem.2020.127653

Jin, T., Guo, F., Chen, Y., Howard, A., & Zhang, Y. Z. (2009). Crystal structure of Ara h 3, a major allergen in peanut. *Molecular Immunology, 46*, 1796–1804. https://doi.org/10.1016/j.molimm.2009.01.023

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, T., Gidamis, A. B., Kanamori, J., Kang, I. J., Utsumi, S., & Kito, M. (1994). Modification tolerability of the hypervariable region of soybean proglycinin. *Journal of Agricultural and Food Chemistry, 42*, 2639–2645.

Katsube-Tanaka, T. (2016). Buckwheat production, consumption, and genetic resources in Japan. In M. Zhou, I. Kreft, S. H. Woo, N. Chrungoo, & G. Wieslander (Eds.), *Molecular breeding and nutritional aspects of buckwheat* (pp. 61–80). Amsterdam: Elsevier.

Katsube-Tanaka, T., & Monshi, F. I. (2022). Characterization of 2S albumin allergenic proteins for anaphylaxis in common buckwheat. *Food Chemistry: Molecular Sciences*, 100127, 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. In Japanese with English abstract https://doi.org/10.18964/jcr.59.0_31.

Kelley, L., Mezulis, S., Yates, C., Wass, M., & Sternberg, M. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols, 10*, 845–858. https://doi.org/10.1038/nprot.2015.053

Khan, N., Katsube-Tanaka, T., Iida, S., Yamaguchi, T., Nakano, J., & Tsujimoto, H. (2008a). Diversity of rice glutelin polypeptides in wild species assessed by the higher-temperature sodium dodecyl sulfate-polyacrylamide gel electrophoresis and subunit-specific antibodies. *Electrophoresis, 29*, 1308–1316.

Khan, N., Katsube-Tanaka, T., Iida, S., Yamaguchi, T., Nakano, J., & Tsujimoto, H. (2008b). Identification and variation of glutelin α polypeptides in the genus *Oryza* assessed by two-dimensional electrophoresis and step-by-step immuno detection. *Journal of Agricultural and Food Chemistry, 56*, 4955–4961.

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.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature, 227*(5259), 680–685.

Monshi, F. I., & Katsube-Tanaka, T. (2022). 2S albumin g13 polypeptide, less related to Fag e 2, can be eliminated in common buckwheat (*Fagopyrum esculentum* Moench) seeds. *Food Chemistry: Molecular Sciences, 5*, Article 100138. https://doi.org/10.1016/j.fochms.2022.100138

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.

Piersma, S. R., Gaspari, M., Hefle, S. L., & Koppelman, S. J. (2005). Proteolytic processing of the peanut allergen Ara h 3. *Molecular nutrition & food research, 49*(8), 744–755. https://doi.org/10.1002/mnfr.200500020

Radović, S. R., Maksimović, V. R., & Varkonji-Gašić, E. I. (1996). Characterization of buckwheat seed storage proteins. *Journal of Agricultural and Food Chemistry, 44*, 972–974.

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.

Utsumi, S., Katsube, T., Ishige, T., & Takaiwa, F. (1997). Molecular design of soybean glycinins with enhanced food qualities and development of crops producing such glycinins. In S. Damodaran, & A. Paraf (Eds.), *Food proteins and lipids 1st edition* (pp. 1–15). New York: Plenum Press.

Wieslander, G., & Norbäck, D. (2001). Buckwheat allergy. *Allergy, 56*, 703–704.

Yasui, Y., Mori, M., Matsumoto, D., Ohnishi, O., Campbell, C. G., & Ota, T. (2008). Construction of a BAC library for buckwheat genome research - an application to positional cloning of agriculturally valuable traits. *Genes & Genetic Systems, 83*, 393–401.

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.