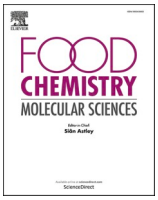
Food Chemistry: Molecular Sciences 8 (2024) 100205

Contents lists available at ScienceDirect

Food Chemistry: Molecular Sciences

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



Elimination of zero-repeat subunit in allergenic seed protein 13S globulin using the novel allele *GlbNB2* in common buckwheat (*Fagopyrum esculentum* Moench)

Takeyuki Okada, Kohtaro Kimura, Naruha Goto, Tomoyuki Katsube-Tanaka *

Graduate School of Agriculture, Kyoto University, Kitashirakawa, Kyoto 606-8502, Japan

ARTICLE INFO

Keywords:
13S globulin
Allergen
Common buckwheat
Zero-repeat subunit
GlbNB2

ABSTRACT

Common buckwheat (*Fagopyrum esculentum* Moench) seeds contain 13S globulin, the zero-repeat subunit of which is trypsin-resistant and allergenic. Here, its two novel alleles were analyzed for development of hypoallergenic plants. The *GlbNC* allele has a Miniature Inverted-repeat Transposable Element (MITE)-like insertion in the 4th exon. However, most of the insertion was spliced-out, resulting in accumulation of zero-repeat subunit in *GlbNC* homozygotes. Meanwhile, the *GlbNB2* has a 164-bp insertion in the 3rd exon, resulting in no accumulation of zero-repeat subunit in *GlbNB2* homozygotes (NB2_homo). Both the insertion sequences were predicted to form a hairpin-like structure, and that of *GlbNB2* was more rigid than that of *GlbNC*. Trypsin digestion in NB2_homo showed that the α polypeptide of Met-rich subunit is also hard to digest, that is a next target to eliminate for hypoallergenic buckwheat development.

1. Introduction

Buckwheat has been conventionally consumed in Southern and Eastern Asia, as well as Eastern and Central Europe, where it is used in various culinary forms, including noodles, porridge, pancakes, jellies, and pasta. This extensive utilization can be attributed to its rapid growth and remarkable adaptability to less fertile environments where major crops cannot be grown (Katsube-Tanaka, 2016). In recent times, it has garnered attention in Western countries due to its alignment with gluten-free dietary preferences (Wieslander & Norbäck, 2001; Heffler et al., 2014). Even though buckwheat has exceptional nutritional and health-promoting characteristics (Zhang et al., 2012; Giménez-Bastida & Zieliński, 2015; Zhang et al., 2017), it is imperative to address allergy concerns (Katsube-Tanaka & Monshi, 2022).

Common buckwheat (*Fagopyrum esculentum* Moench) and Tartary buckwheat (*Fagopyrum tataricum* (L.) Gaertn.) are two cultivated species in the genus *Fagopyrum*. The latter species contains more rutin, a health-promoting flavonoid, than the former. Recently new allergen Fag t 6 (18 kDa oleosin-type protein) was found in Tartary buckwheat (Chen et al., 2021), in addition to Fag t 1 (11S globulin), Fag t 10kD (alpha-amylase/trypsin inhibitor), and Fag t 2 (2S albumin) (ALLERGOME, Mari et al., 2009). Common buckwheat (referred as just buckwheat, hereafter) has two major seed storage proteins 13S globulin and 2S albumin, and they are recognized as crucial allergens Fag e 1 and Fag e 2, respectively (Katayama et al., 2018), along with Fag e 10kD (alpha-amylase/trypsin inhibitor), Fag e 3 (7S vicilin), Fag e 4 (anti-microbial peptide), Fag e 5 (8S globulin), and Fag e TI (trypsin inhibitor) (ALLERGOME, Mari et al., 2009).

The 13S globulin belongs to the 11S globulin superfamily. The 11S globulin is ubiquitously present in seed plants and typically undergoes self-assembly, resulting in the formation of a hexameric protein structure (Adachi et al., 2003; Choi et al., 2006). Similar to the behavior of 11S globulins, the proprotein of 13S globulin subunit experiences cleavage into two distinct polypeptides, namely the acidic (α) and basic (β) polypeptides (Radović et al., 1996). In its mature form, these polypeptides are covalently linked through a disulfide bond.

The 13S globulin consists of methionine (Met)-rich and Met-poor subunits. In contrast to other 11S globulins, such as the rice (*Oryza sativa* L.) glutelin (Khan et al., 2008a; 2008b) and the soybean [*Glycine* max (L.) Merr.] glycinin (Katsube et al., 1999), the α polypeptide within the Met-poor subunits exhibits significant size variations due to the presence of a varying number of tandem repeats, ranging from zero to 1–6, giving rise to zero- and 1–6-repeat subunits (Khan et al., 2012). Each unit of tandem repeat insertion is composed of 15 amino acid

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

https://doi.org/10.1016/j.fochms.2024.100205
Received 8 January 2024; Received in revised form 20 April 2024; Accepted 20 April 2024
Available online 22 April 2024

residues that are hydrophilic and arginine-rich. Thus, 1–6-repeat subunits are susceptible to trypsin digestion while zero-repeat subunit is resistant, which may affect the difference in allergenicity (Khan et al., 2012).

With regard to the trypsin-resistant zero-repeat subunit, the initial identification of two alleles, *GlbNA* and *GlbNB*, was reported by Sano et al. (2014), employing a genomic DNA library established by Yasui et al. (2008). Subsequently, six novel alleles, namely *GlbNA1*, *GlbNA2*, *GlbNA3*, *GlbNC*, *GlbNC1*, and *GlbND*, were isolated from Pakistani landraces and Japanese elite cultivars (Monshi et al., 2020; Katsube-Tanaka et al., 2014). In more recent research by Okada et al. (2023), a novel allele, *GlbNA(10aa)*, was identified. This allele encodes a zero-repeat subunit that is more susceptible to trypsin digestion compared to the native zero-repeat subunit. This heightened susceptibility is attributed to the insertion of ten additional amino acids at a position equivalent to that of the tandem repeat insert. However, it is worth noting that no non-functional allele conducive to the development of hypoallergenic buckwheat has been identified as of yet.

It is suggested that common buckwheat has a large genetic diversity because of its outcrossing nature (Katsube-Tanaka, 2016) and the existence of a large portion of transposable elements in its genome (Penin et al., 2021). The tandem repeat insertion into Met-poor subunits of 13S globulin is structurally exceptional in 11S globulin members, since other 11S globulins, for example soybean [*Glycine* max (L.) Merrill] glycinin accepted only limited structural modifications (Katsube et al., 1994; Utsumi et al., 1997). Possibly from these reasons, Monshi & Katsube-Tanaka (2022) could successfully find the unfunctional 2S albumin allele *g13_null*, that contains extra 531 bp sequence in its coding region in natural populations. Additionally, it is noteworthy that the zero-repeat subunit alleles of 13S globulin, namely *GlbNA2*, *GlbNA3*, and *GlbNC*, encompass a Miniature Inverted-repeat Transposable Element (MITE)-like sequence spanning approximately 200 base pairs (Monshi et al., 2020). In the context of this investigation, we postulate that new non-functional alleles pertaining to the zero-repeat subunit can be discerned within natural common buckwheat populations. Here, we report the novel zero-repeat subunit allele *GlbNB2*, which has a MITE-like sequence in its coding region. We compare the structure and function of *GlbNC* and *GlbNB2* alleles to deepen our knowledge on the allergenicity of seed storage proteins and genetic diversity of common buckwheat. This is the first biochemical and molecular report to remove zero-repeat subunit in common buckwheat seeds, that would be beneficial for hypoallergenic buckwheat breeding.

2. Materials and methods

2.1. Common buckwheat varieties and the extraction of genomic DNA and proteins

A single seed of common buckwheat variety ‘Harunoibuki’ and *GlbNC*-homozygous line (Monshi et al., 2020) was ground in a 2-mL tube using a crusher machine (MixerMill, QIAGEN K.K., Tokyo, Japan) and used for the extraction of saline-soluble proteins using sample buffer (35 mM potassium phosphate, pH 7.6, 300 μL) supplemented with 0.4 M sodium chloride. From the residue after extracting the protein, genomic DNA was prepared with the DNAs-ici!-S reagent (Rizo, Tsukuba, Japan) (Monshi & Katsube-Tanaka, 2022).

2.2. Genotyping and development of GlbNB2-homozygous line

The *GlbNB2*-homozygous line was developed from ‘Harunoibuki’ using PCR. Crude genomic DNA was extracted from a leaf disk detached with a 2-mL tube lid and crushed with MixerMill in 100 μL of DNA sample buffer (0.1 M Tris pH 9.5, 1 M potassium chloride, and 10 mM EDTA pH 8.0). The genotyping was conducted using the forward primer; 0rep_TAA-200 (5F): 5′- AtTGGAGTGGGTtGAGTTGAgGAcC −3′ and the reverse primers; NA_C306_rev: 5′- GAGACATGAATACGACGGGTGTTG −3′ for the *GlbNA* type, NB_C406_rev: 5′- GCTTAACATCATTCCGTTACCGG −3′ for the *GlbNB* type, NC_C349_rev: 5′- TGCTGTTTCGGACTTTTTCCTCC −3′ for the *GlbNC* type, and ND-C340_rev: 5′- GCTTCCGAACGATCCCTTAATGCAAG −3′ for the *GlbND* type.

PCR was performed as a single (for *GlbNB* type) or a multiplex (for *GlbNA, GlbNC*, and *GlbND* types) PCR using thermal cycler (GeneAtlas G02, ASTEC, Fukuoka, Japan) and conditions described by Monshi et al. (2020). The novel allele *GlbNB2* was distinguished from the authentic *GlbNB* allele by failed or successful amplification with the forward primer; M13F_0rep_1F (SP): 5′- gtaaaacgacggccagtCACTAAGCCACCAACATG −3′ and the reverse primer; NB_C406_rev, respectively.

The *GlbNB2* allele-containing plants were grown in isolated fields or glasshouse to enrich the *GlbNB2* allele frequency. Progenies were also genotyped to screen for *GlbNB2* allele homozygotes. The oligonucleotides for PCR were purchased from a DNA synthesis company (Eurofins Genomics, Tokyo, Japan). Genotyping for *GlbNB* type (authentic *GlbNB* and *GlbNB2*) and *GlbNB2* was also conducted in the cultivar ‘Shinano 1’ to estimate allele frequency (n = 92).

2.3. Cloning and sequencing

The insertion sequence of *GlbNB2* was amplified using a nested PCR method with 1st PCR primers of M13F_0rep_2F (2F): 5′- gtaaaacgacggccagtGTTATACCAGGATGTCCG −3′ and M13R_0rep_3R (3R): 5′- caggaaacagctatgacCACGTTCCATTTAGGTCC −3′ and 2nd PCR primers of M13F_0rep_3F (3F): 5′- gtaaaacgacggccagtGTATTCAGAGGTGCTGAC −3′ and M13R_0rep_4R (4R): 5′- caggaaacagctatgacGCATTCTGCACAyCATAC −3′. The thermal condition of the 1st PCR was one cycle of 94 °C 2 min, 40 cycles of 98 °C 20 s, 65 °C 30 s, 68 °C 20 min, and one cycle of 68 °C 7 min and that of the 2nd PCR was one cycle of 94 °C 2 min, 35 cycles of 98 °C 20 s, 65 °C 30 s, 68 °C 3 min, and one cycle of 68 °C 7 min. KOD FX neo DNA polymerase (TOYOBO, Osaka, Japan) was used with the manufacturer’s recommended components supplemented with $MgCl_2$ (final concentration: 4 mM) for the 1st and 2nd PCRs and with 4 % (v/v) DMSO and 0.5 M betaine for the 2nd PCR. The amplified fragment (~800 bp) was recovered and cloned into pTAC vector (BioDynamics Laboratory Inc., Tokyo, Japan). Sanger sequencing of the insert was conducted by Azenta Life Sciences (Tokyo, Japan) and Eurofins Genomics (Tokyo, Japan) with a difficult template protocol.

2.4. RT-PCR

The reverse transcription-polymerase chain reaction (RT-PCR) analysis for the zero-repeat subunit gene in *GlbNC* and *GlbNB2* homozygotes was conducted using RNAs-ici!-R (Rizo, Tsukuba, Japan) with DNaseI (NIPPON GENE, Tokyo, Japan) and the TaKaRa RNA PCR Kit (AMV) Ver.3.0 (Takara Bio Inc., Kusatsu, Japan) with oligo dT-adapter (*GlbNC*) and random 9-mers (*GlbNB2*) primers. Second strand cDNA was synthesized with M13F_0rep_3F (3F) and TAA_3′_0rep_rc: 5′- TTAAACGACGTCGTATCTsyCCC −3′ for *GlbNC* and with M13F_0rep_2F (2F) and M13R_0rep_2R (2R) for *GlbNB2*, and BW actin Self3 Fwd: 5′- GGCATCACACTTTCTACAATGAGC −3′ and BW actin Self3 Rev: 5′- GATATCCACATCACATTTCATGAT −3′ for actin. RT-PCR product of *GlbNC* was cloned into pTAC vector (BioDynamics Laboratory Inc., Tokyo, Japan) and sequenced. The thermal cycling program and PCR components were followed to the manufacturer’s instruction. At least three biological replicates were examined using biologically independent seeds with single technical replicate.

2.5. Recombinant α-polypeptide of GlbNA and Met-rich subunit

Immature grains of common buckwheat cultivar ‘Shinano 1’ were used for total RNA extraction using RNeasy Plant Mini Kit (QIAGEN, Tokyo, Japan). Complimentary DNA (cDNA) of *GlbNA* and *GlbRA* (Met-rich subunit) was synthesized using the TaKaRa RNA PCR Kit (AMV)

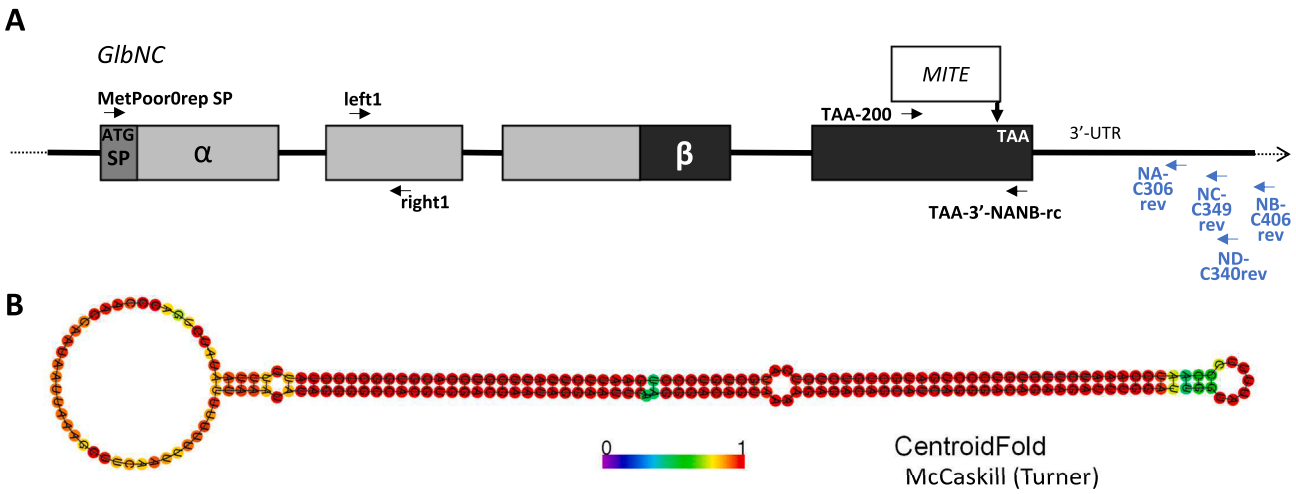


Fig. 1. Schematic representation of *GlbNC* allele and its MITE-like sequence. (A) *GlbNC* allele structure of 13S globulin zero-repeat subunit gene. Shaded boxes represent exons and open box represents MITE-like insertion. Short black and blue horizontal arrows indicate common and specific PCR primers for each zero-repeat subunit gene, respectively. Regions for signal peptide (SP), α and β polypeptide chains are indicated. (B) Predicted hairpin structure of the MITE-like sequence with the CentroidFold software [McCaskill (Turner)]. Warm and cool colors imply strong and weak hydrogen bonds, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Ver.3.0 (Takara Bio Inc., Kusatsu, Japan) with oligo dT-adapter primer and MetPoor 0 repeat SP: 5′- ATGTCTACGAAGCTCAATCTCTTCATCT −3′ or BW MetRich SP: 5′-ATGCTTCATGGGGTGCTTCTAT-3′. Insert DNA fragment coding for the GlbNA and GlbRA mature polypeptide was amplified with GlbNAB_N: 5′- TACCTCCGCCACCACGAAGGC −3′ and TAA 3′ NANB rc: 5′- CGGAGCTCTTAAACGACGTCGTATCTCTC −3′ or GlbRAB_N: 5′-GAAACCCGCAGTCGAGGTTCCTC-3′ and TAA 3′ Met rich rc: 5′-GCGAGCTCCATAAAAGAAGCCCTAGTTC-3′, respectively. The fragments were digested by SacI and ligated to the NcoI/Klenow – SacI sites of pET21d *Escherichia coli* (*E. coli*) expression vector (Merck KGaA, Darmstadt, Germany) to produce pET21d::GlbNA and pET21d::GlbRA plasmids. Insert DNA fragment coding for the GlbNA α polypeptide was amplified with T7 Promoter: 5′- TAATACGACTCACTATAGGG −3′ and NAalphaC TAA_Sac_rev: 5′- acGAGCTCTTAATTGGCACCTCTCCTTC −3′ and the pET21d::GlbNA plasmid. The fragment was digested with XbaI and SacI and ligated to the XbaI – SacI sites of pET21d to produce pET21d::GlbNAα. The expression plasmid pET21d::GlbRAα to produce α-polypeptide of Met-rich subunit was prepared by amplification with RA_beta_N_F: 5′-GGGTTAGAAGAATCCTTCTGCAAC-3′ and Met-rich_alpha C_TAG_Sac rev: 5′-ACGAGCTCTAGTTTGTCTCGTCCTCATC-3′ and the pET21d::GlbRA plasmid, followed by SacI digestion and self-ligation. The recombinant α-polypeptide of GlbNA (NAα) and GlbRA (RAα) was expressed in *E. coli* [BL21(DE3)] cells and recovered in an insoluble fraction after sonication as described previously by Katsube-Tanaka & Monshi (2022).

2.6. Dot blot, SDS-PAGE, and western blot

Eighty grains each of ‘Harunoibuki’ cultivar and *GlbNC*-homozygous line were used for dot blot analysis. Saline-soluble proteins were extracted using the potassium phosphate buffer (35 mM, pH 7.6, 500 μL) supplemented with 0.4 M NaCl and 2 μL of extract was applied on a nitrocellulose membrane (Amersham Protran Premium 0.2 μm NC, Cytiva, Uppsala, Sweden) with a diluted series of recombinant NAα. After immunoreaction using the first antibodies for the α-polypeptide of zero-repeat subunit (Okada et al., 2023), the intensity of each spot was quantified using ImageJ software (U. S. National Institutes of Health, https://imagej.nih.gov/ij/) according to Katsube-Tanaka et al. (2004). The protein concentration of the extract was measured by Bradford method with the bovine serum albumin as a calibration standard.

Another saline-soluble extract of 99 grains in ‘Harunoibuki’ was analyzed with SDS-polyacrylamide gel electrophoresis (SDS-PAGE) following to Laemmli (1970) and western blotting according to Khan et al. (2012) using anti-GlbNA antibody. The genotype of each grain was confirmed using crude gDNA extracted from the residue with genotyping primers described in the section 2.2.

Western blot for zero-repeat subunit accumulation in the *GlbNB2* homozygote (*NB2/NB2*), *GlbNB*-type (*NB/NB* or *NB/NB2*), *GlbNB2* heterozygote (*NB2/NX*), and non-*GlbNB*-type (*NX/NX*) seeds was conducted.

2.7. Digestion of extracted protein with trypsin

The seven dehulled seeds of ‘Shinano1’ and *GlbNB2* homozygote were ground with a zirconia bead in 2-mL microtube with the QIAGEN MixerMill at 30 rps for 30 s. The 100 mg of buckwheat flour was treated with 1 mL of extraction buffer (35 mM potassium phosphate pH 7.6 supplemented with 0.4 M NaCl) and shaken for 10 min. The supernatant was then separated by centrifugation (4 °C, 14,000 rpm, 10 min) twice, and the second supernatant was used for trypsin digestion. The extracted protein (20 μL) was mixed with 3.0 μL of 1 M Tris pH 9.8 and 5.0 μL of Trypsin (2.5 μg/μL, Nacalai, Kyoto, Japan) to make a total volume of 28 μL. Trypsin digestion was carried out in an incubator at 37 °C for up to 4 h. At 0, 0.5, 1, 2, and 4 h after the start of the reaction, 5.6 μL of the incubated mixture was taken and mixed with 5.6 μL of 2 × SDS buffer [100 mM Tris pH 6.8, 20 % (w/v) glycerol, 4 % (w/v) SDS], 0.6 μL of 2-mercaptoethanol, and 0.5 μL of 0.1 % (w/v) bromophenol blue, followed by boiling for 10 min to prepare the samples for electrophoresis. A control sample was prepared by adding an equal volume of sterilized distilled water instead of trypsin. Trypsin digestion was tested at least three times using biologically different samples with a single technical replicate.

2.8. BLAST search and sequence analyses

BLAST (Megablast) search was conducted with default algorithm parameters at the NCBI DataBase (https://www.ncbi.nlm.nih.gov/) against the GenBank assembly of genomic/3617/GCA_004303065.2 (Cultivar: Dasha) with a query sequence of *GlbNC*_insert (208 letters) and *GlbNB2*_insert (164 letters). The secondary structure prediction of nucleotide sequence was conducted with the web server program CENTROIDFOLD (https://rtools.cbrc.jp/centroidfold/) by Sato et al. (2009).

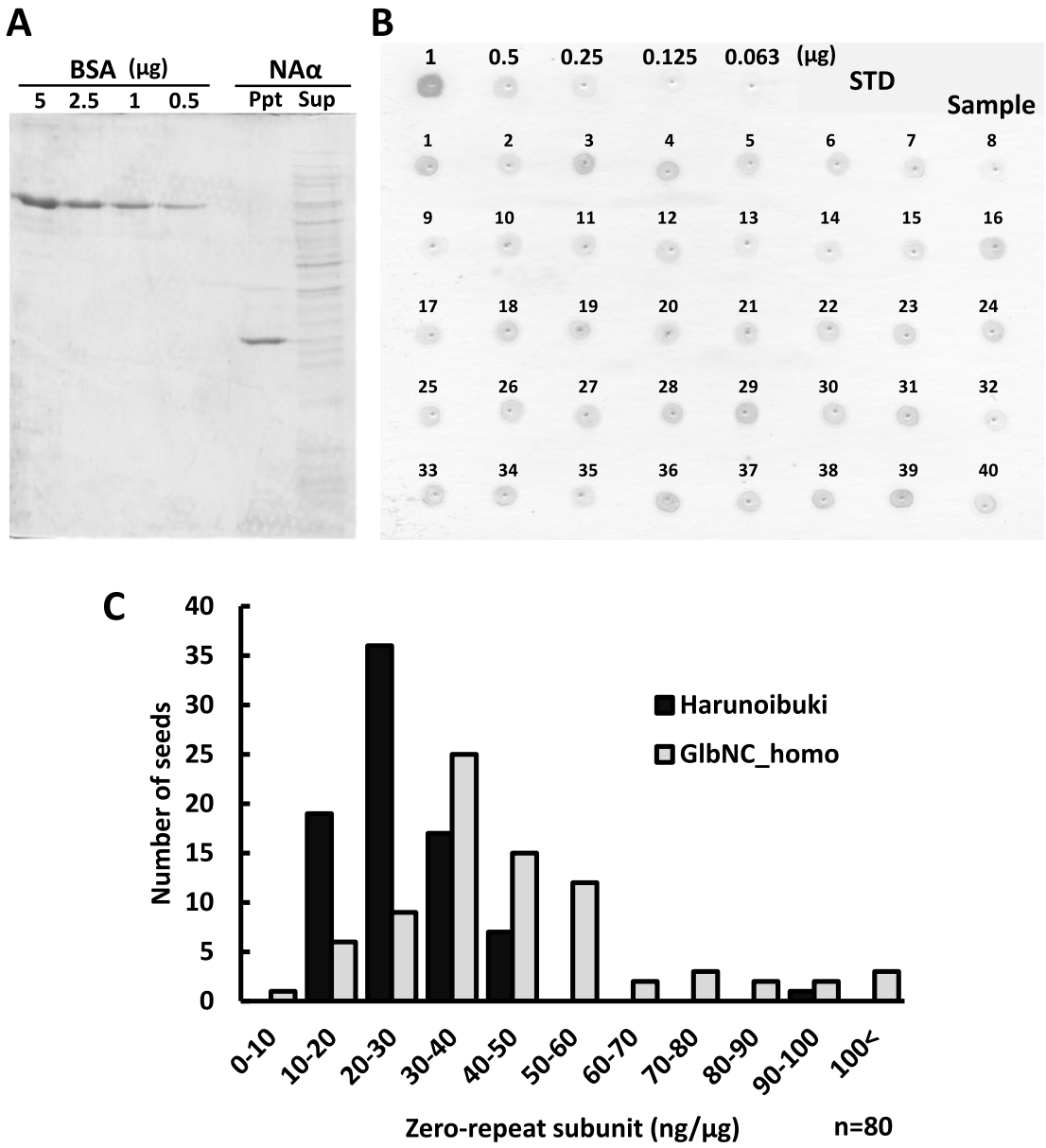


Fig. 2. Accumulation of zero-repeat subunit in wild type and *GlbNC* allele-homozygote in buckwheat seeds. (A) Expression and quantification of recombinant α polypeptide of 13S globulin zero-repeat subunit (NAα) in *E. coli* expression system. Bovine serum albumin (BSA) was used for reference. Insoluble (Ppt) and soluble (Sup) fractions were separated. (B) Exemplified result of zero-repeat subunit quantification by a dot blot. Diluted series of NAα (STD) and 40 seed extracts were dotted and immunologically detected. (C) Distribution of zero-repeat subunit accumulation in wild type (Harunoibuki) and *GlbNC* allele-homozygote (GlbNC_homo) seeds (n = 80 each).

2.9. Statistical analyses

Chi-square test was employed to evaluate the difference in allele frequency of *GlbNB2*. The statistical analysis was conducted using Bell Curve for Excel (ver. 3.21, Social Survey Research Information, Tokyo, Japan).

3. Results

3.1. Accumulation and transcription of zero-repeat subunit in GlbNC homozygotes

The *GlbNC* allele of 13S globulin zero-repeat subunit has a MITE-like sequence ~ 70 bp upstream from the stop codon. The MITE-like sequence was predicted to form a rigid hairpin structure, which may interfere with transcription and/or translation (Fig. 1). Thus, the accumulation of zero-repeat subunit was examined in *GlbNC* homozygote and the wild type seeds, after the establishment of zero-repeat subunit quantification system (Fig. 2). The recombinant α-polypeptide of GlbNA (NAα) was prepared and quantified from the insoluble fraction of *E. coli* cell extract, which was used as a calibration standard for the zero-repeat level quantification with a dot blot method. The accumulation level of zero-repeat subunit varied from less than 20 ng μg^{-1} to over 90 ng μg^{-1} in both the wild type and *GlbNC* homozygote and peaked at 20–30 ng μg^{-1} for the wild type and at 30–40 ng μg^{-1} for *GlbNC* homozygote (Fig. 2). Therefore, the transcription and translation of zero-repeat subunit gene was not suppressed by the MITE-like sequence inserted in *GlbNC*.

Provided that the MITE-like sequence of *GlbNC* is translated as it is, the synthesized polypeptide of *GlbNC* would be shorten by 13 residues compared to that of *GlbNC1*, that is almost identical to *GlbNC* except the MITE-like sequence insertion (Fig. 3). In order to elucidate the

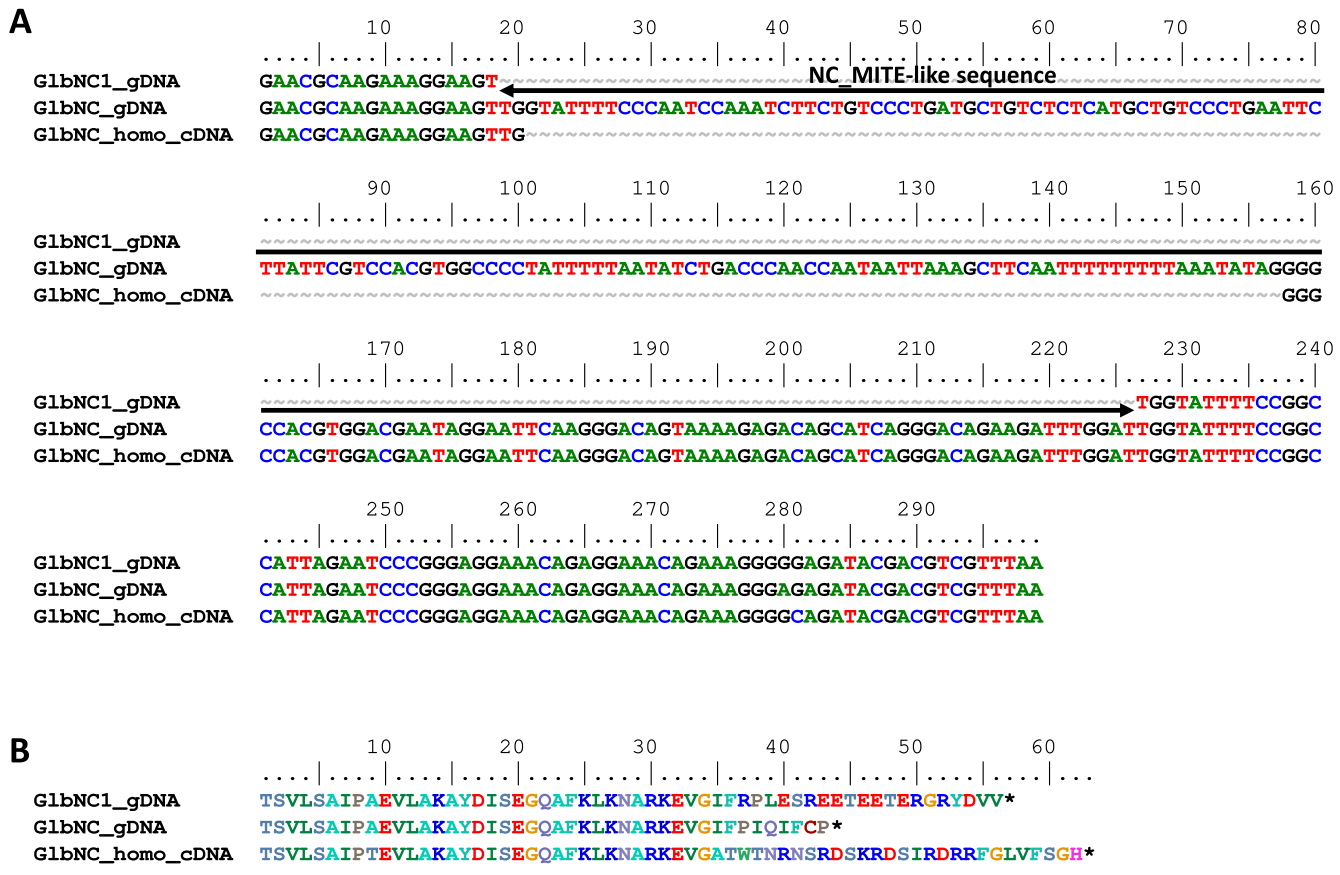


Fig. 3. Nucleotide and amino acid sequences of *GlbNC1* and *GlbNC* alleles. (A) Genomic DNA sequences of *GlbNC1* and *GlbNC* alleles and cDNA sequence of zero-repeat subunit gene detected in *GlbNC* allele-homozygote were shown before and after the MITE-like insertion (NC_MITE-like sequence, horizontal arrows). (B) Difference in amino acid sequences deduced from gDNA and cDNA sequences. * indicates a translation termination.

molecular mechanism underlying the accumulation of GlbNC, RT-PCR analysis with various combinations of primers was conducted. Consequently, by using oligo dT-adapter primer and the zero-repeat subunit gene-specific reverse primer (TAA_3′_0rep_rc), cDNA of *GlbNC* was isolated, in which most of the MITE-like sequence (137 bp out of 208 bp) was removed. Note that the removed sequence started at GT and ended at AG, that is a common rule for splicing of intron in Eukaryote genes. The deduced amino acid sequence of *GlbNC* cDNA was longer by 6 residues compared to that of *GlbNC1*.

3.2. Identification of the novel allele GlbNB2

Because the accumulation level of zero-repeat subunit varied seed-by-seed in the dot blot analysis of 'Harunoibuki' (80 grains) (Fig. 2), we more precisely evaluated the accumulation level using western blot (99 garins), indicating that the α polypeptide of zero-repeat subunit varied in its molecular size and accumulation level (Fig. 4A). Among them, at least two grains showed a very faint band (for example, #23 in Fig. 4A), the genotype of which was *GlbNB*-type allele homozygous. Thus, *GlbNB*-type allele-rich population was developed. During the development, we noticed that some *GlbNB*-type allele homozygotes did not show amplified fragments with PCR primers after 2R or before 4F (Fig. 4B), suggesting that a novel *GlbNB*-type allele, called *GlbNB2* later, exists in 'Harunoibuki'. The *GlbNB2* homozygotes showed no transcription (Fig. 4C) and no translation (Fig. 4D) products of zero-repeat subunit.

3.3. Gene structure of GlbNB2

As described in the section 3.2, normal PCR could not amplify the region between 2R and 4F primers in *GlbNB2*. Therefore, a nested PCR with increased concentration of Mg^{2+} to stabilize a primer-template complex and addition of secondary structure-destabilizing reagents (DMSO and betaine) was conducted. After testing various PCR conditions, we could finally clone the region between 3F and 4R primers. The Sanger sequencing of the clone with a difficult template protocol revealed that *GlbNB2* contains a 164 bp insertion just after the 2R primer site (Fig. 5A and 5B). The insertion sequence was predicted to form a rigid hairpin structure by CentroidFold software (Fig. 5C). Note that the insertion sequence started at GT and ended at AG, that is same with that of the spliced-out region in the MITE-like sequence of *GlbNC*.

3.4. Trypsin digestibility of GlbNB2 homozygotes

Trypsin digestion was conducted using seed globulin fractions of the cultivar 'Shinano 1' and the *GlbNB2* homozygote seeds. The digested samples were analyzed with recombinant α polypeptide of Met-rich subunit [RA(α)] and zero-repeat subunit [NA(α)]. The results showed that α polypeptides of 13S globulin were mostly digested within 2 h in Shinano 1 and *GlbNB2* homozygote except α polypeptide of Met-rich subunit in Shinano 1 and *GlbNB2* homozygote (Fig. 6A) and that of zero-repeat subunit in Shinano 1 (Fig. 6B), which was largely resistant till 2 h incubation. Note that β polypeptides were also fairly resistant till 2 h incubation in both the cultivar and homozygote line (Fig. 6A).

3.5. BLAST search

To compare the frequency of insertion into a buckwheat genome between *GlbNC*_insert and *GlbNB2*_insert, BLAST search was conducted for the cultivar 'Dasha' (GenBank assembly: genomic/3617/

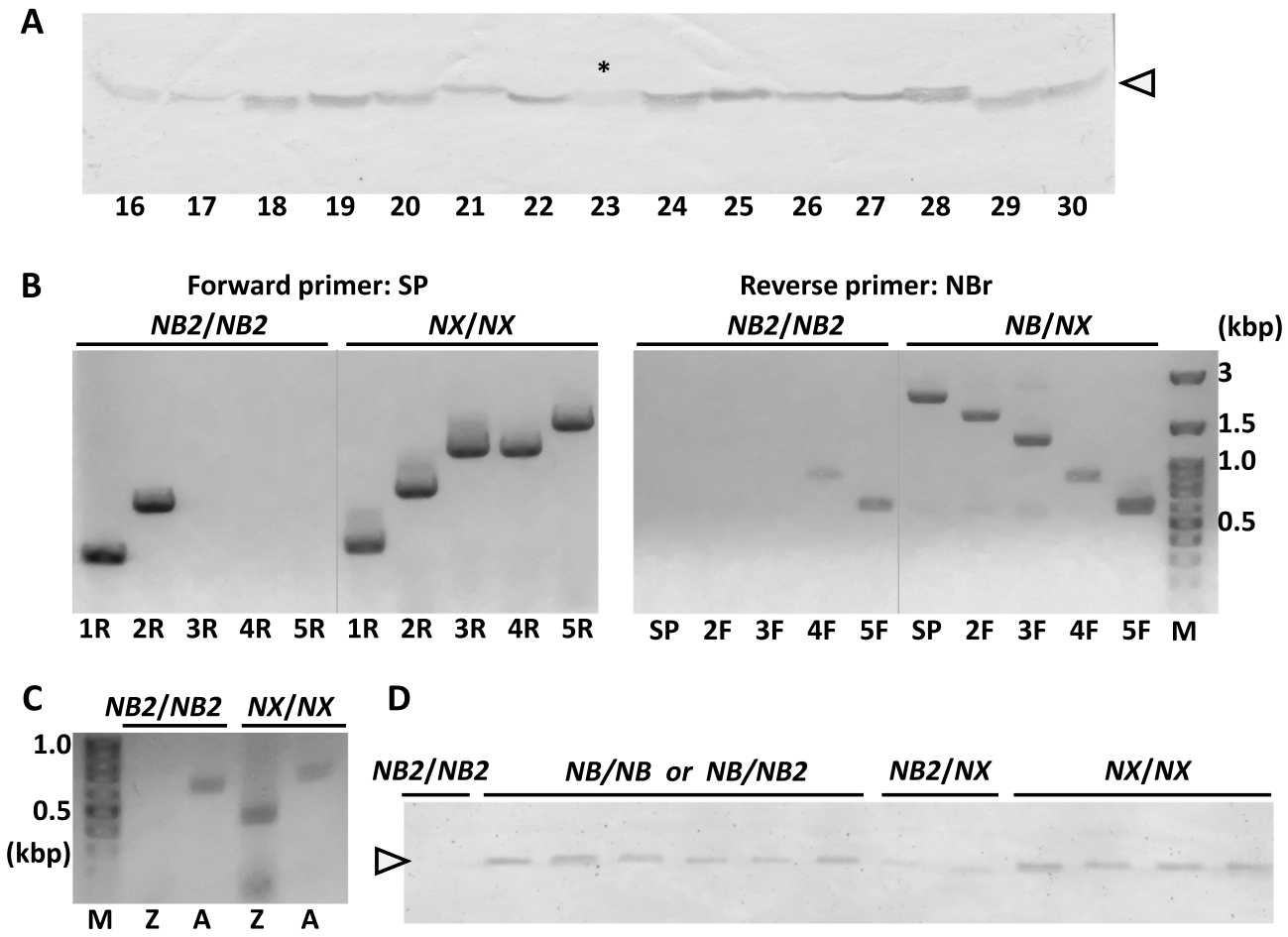


Fig. 4. Western blot, PCR, and RT-PCR analyses for *GlbNB2* allele isolation and characterization. (A) Exemplified results of western blot for zero-repeat subunit accumulation in the cultivar 'Harunoibuki'. Only #16–#30 were shown from 99 grains analyzed. Arrowhead and asterisk indicate the position of zero-repeat subunit α polypeptide and the seed possesses *GlbNB*-type allele, respectively. (B) PCR for the newly identified *GlbNB2* homozygote (*NB2/NB2*), non-*GlbNB*-type (*NX/NX*), and authentic *GlbNB* allele containing (*NB/NX*) seeds. Primer combinations were shown above and underneath the images. M represents a size marker. (C) RT-PCR for the *GlbNB2* homozygote (*NB2/NB2*) and non-*GlbNB*-type (*NX/NX*) immature seeds. Z and A represent zero-repeat subunit and actin gene, respectively. M represents a size marker. (D) Western blot for zero-repeat subunit accumulation in the *GlbNB2* homozygote (*NB2/NB2*), *GlbNB*-type (*NB/NB* or *NB/NB2*), *GlbNB2* heterozygote (*NB2/NX*), and non-*GlbNB*-type (*NX/NX*) seeds. Arrowhead indicates the position of zero-repeat subunit α polypeptide.

GCA_004303065.2) using Megablast algorithm with default parameters optimized for highly similar sequences. The *GlbNC*_insert-like sequence was found to be more in number (141 hits) than the *GlbNB2*_insert-like sequence (10 hits) in the Dasha genome (Supplementary Fig. S1A). Most of hits were aligned with only half of the query in *GlbNB2*_insert (exemplified by Dasha scaffold #7458, Supplementary Fig. S1B) while that was aligned with full in *GlbNC*_insert (exemplified by Dasha scaffold #3465, Supplementary Fig. S1B). It is notable that Dasha scaffold #11849 has the alignments for the first half and the second half of query tandemly with 458 bp of inter-alignment region containing 216 bp of gap (N) (Supplementary Fig. S1B). Meanwhile, Dasha scaffold #4517 has two sets of alignments, which are apart from each other with 5,097 bp of inter-alignment region and each set has two tandem alignments in a same direction in one strand (Supplementary Fig. S1B).

4. Discussion

4.1. Difference between GlbNC_insert and GlbNB2_insert

In this study, we demonstrated that most of *GlbNC*_insert sequence was spliced out like an intron but *GlbNB2*_insert was not, although both the spliced-out sequence of *GlbNC*_insert and the *GlbNB2*_insert have GT and AG nucleotides at their ends (Figs. 3 and 5), that are common for splicing sites of intron in Eukaryote genes. One possible reason for this might be the difference in the robustness of possible hairpin structures caused by strong hydrogen bonds. The 3′ region of the spliced-out sequence of *GlbNC*_insert is predicted to form a non-hairpin (loop) structure (Fig. 1), that is not found in the *GlbNB2*_insert (Fig. 5). It is known that introns during splicing form a lariat structure by connecting 5′ splice site and branch point site within the introns (Peters & Toor, 2015), which might be interfered with the rigid hairpin structure. Another possible reason might be related to the difference in the number of hits found in the homology search. BLAST search demonstrated that homologous sequences with high score were found more for the *GlbNC*_insert than for the *GlbNB2*_insert in the Dasha genome (Supplementary Fig. S1), indicating that the *GlbNC*_insert-like sequence can be retained more acceptably than the *GlbNB2*_insert-like sequence in a buckwheat genome. Although the splice-out of *GlbNC*_insert was not perfect, the insertion of *GlbNC*_insert-like sequence is possibly less harmful to the genome than that of *GlbNB2*_insert-like sequence. Further confirmation of that hypothesis is necessary.

4.2. Trypsin digestibility of GlbNB2 homozygotes

Previously Okada et al. (2023) found a zero-repeat subunit with 10 amino acids insertion (10aa subunit) in natural buckwheat populations and demonstrated that α polypeptide of 10aa subunit is more digestible by trypsin than that of native zero-repeat subunit only when 10aa

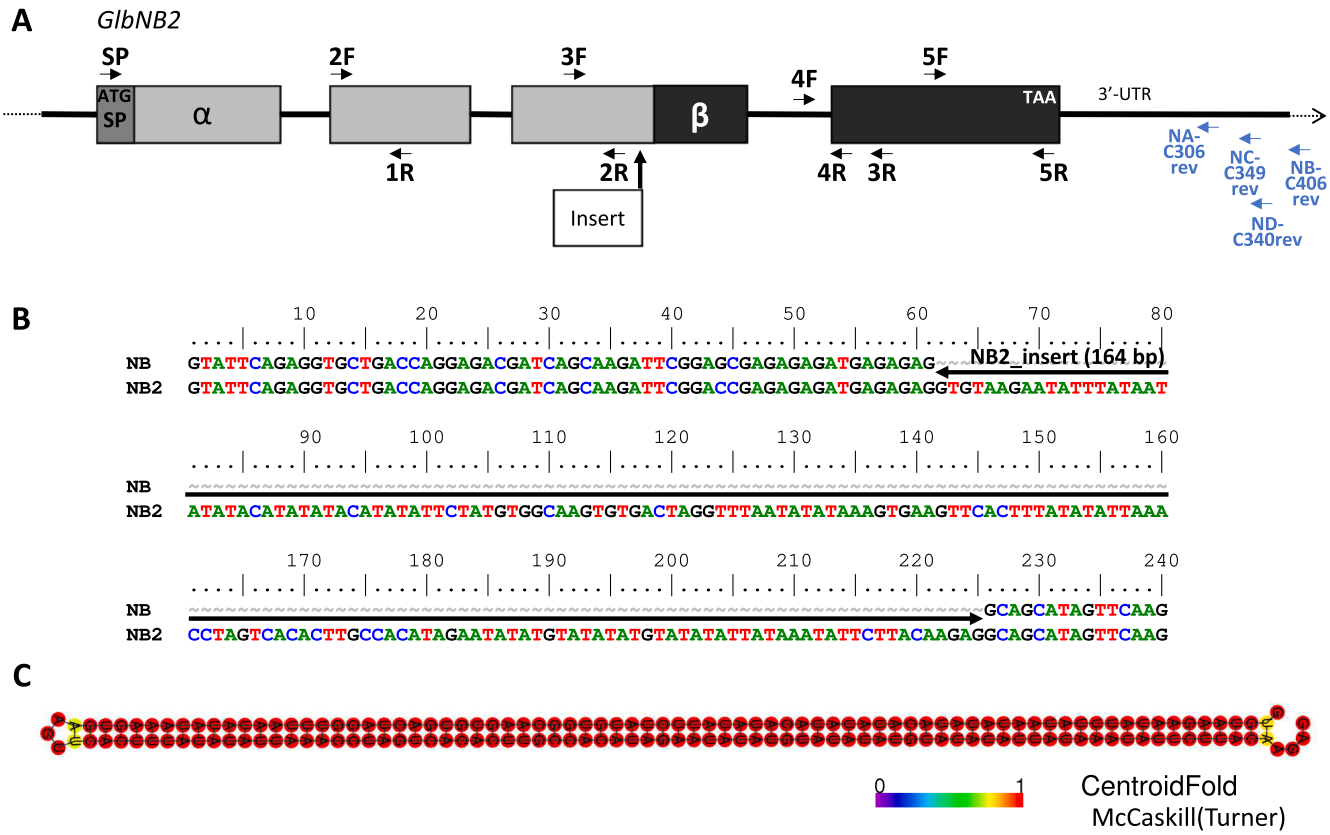


Fig. 5. Schematic representation of *GlbNB2* allele and its insertion sequence. (A) *GlbNB2* allele structure of 13S globulin zero-repeat subunit gene. Shaded boxes represent exons and open box represents a unique insertion sequence. Short black and blue horizontal arrows indicate common and specific PCR primers for each zero-repeat subunit gene, respectively. Regions for signal peptide (SP), α and β polypeptide chains are indicated. (B) Genomic DNA sequences of authentic *GlbNB* and *GlbNB2* alleles were shown before and after the insertion sequence (horizontal arrows). (C) Predicted hairpin structure of the insertion sequence with the CentroidFold software [McCaskill (Turner)]. Warm and cool colors imply strong and weak hydrogen bonds, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

subunit accumulates without native zero-repeat subunit in seeds. 13S globulin molecule accumulates as a hexamer in seeds. We speculated that the subunit composition of hexamer may change the stability of 13S globulin. However, although the interaction and relationship between zero-repeat and 1–6-repeat subunits in the hexamer is still unclear, the presence or absence of native zero-repeat subunit does not affect the apparent trypsin digestibility of α polypeptides in 1–6-repeat subunits (Fig. 6A). Meanwhile, the trypsin digestibility of β polypeptides was also apparently unchanged between wild type (Shinano 1) and *GlbNB2* homozygote (Fig. 6A). However, Okada et al. (2023) pointed out that the pI of α polypeptides, that is an important trait determining physicochemical properties of polypeptides, is different between zero-repeat subunit (pI 5.9) and 1–4-repeat subunits (pI 5.1) in average. The pI of β polypeptides is also different between zero-repeat subunit (pI 6.1) and 1–4-repeat subunits (pI 9.5) in average, suggesting that the individual polypeptides of 13S globulin may differ in allergenicity as well as trypsin digestibility.

4.3. Cultivar difference in structure and allele frequency of GlbNB2

Monshi et al. (2020) demonstrated that the allele frequency of *GlbNB* type (authentic *GlbNB* and *GlbNB2*) was 0.05 for the cultivar 'Harunoibuki' and 0.25 for the cultivar 'Shinano 1', implying that the allele frequency of *GlbNB* type was five times larger in 'Shinano 1' than in 'Harunoibuki'. The similar results were obtained during the development of *GlbNB2* homozygote in this study, indicating 0.06 for 'Harunoibuki' and 0.21 for 'Shinano 1' (Data not shown). Meanwhile, the allele frequency of *GlbNB2* was 0.05 for 'Harunoibuki' and 0.09 for 'Shinano 1', that is not significantly different between them, indicating that 'Harunoibuki' is a good material for the *GlbNB2* homozygote development because most of *GlbNB* type is *GlbNB2* in this cultivar. BLAST search in the cultivar 'Dasha' also found the *GlbNB2*-like sequence in scaffold #11849 (Supplementary Fig. S1), suggesting that the allele frequency of *GlbNB2* is not so low in natural populations. However, the scaffold #11849 contains 458 bp of inter-alignment region including 216 bp of gap (N) between the first half and the second half of *GlbNB2*_insert (Supplementary Fig. S1). Therefore, cultivar difference in the structure and frequency of *GlbNB2* and *GlbNB2*-like sequence may not be small.

5. Conclusion

In this study, the reason why the *GlbNC* allele containing MITE-like insert is functional was elucidated. Instead, the novel non-functional allele of zero-repeat subunit, *GlbNB2*, was identified within natural common buckwheat populations. The *GlbNB2*-homozygous line was efficiently developed from the cultivar 'Harunoibuki', that could be useful for the development of hypoallergenic buckwheat. However, we found that the developed *GlbNB2*-homozygous line still has the trypsin resistant α polypeptide of Met-rich subunit, that would be a next target to eliminate.

Funding sources

This research was partially supported by JSPS KAKENHI Grant Number 20H02967 to TKT.

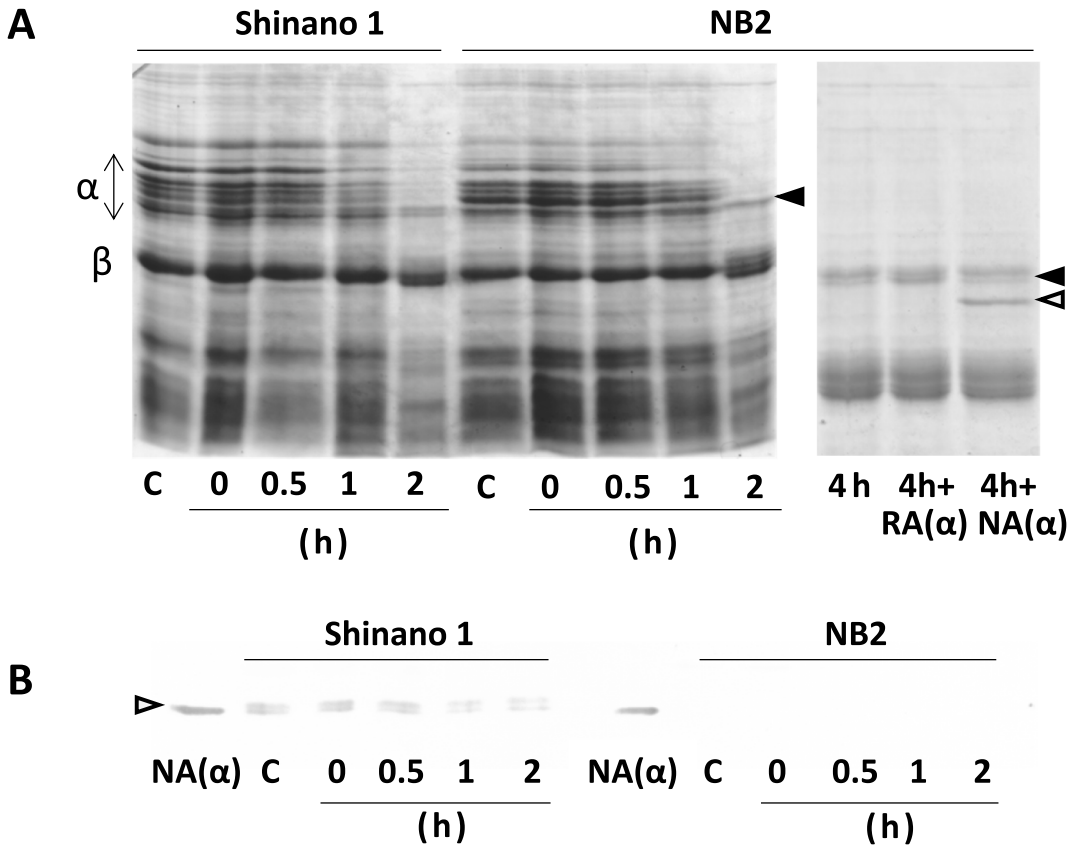


Fig. 6. Trypsin digestion of the cultivar 'Shinano 1' and *GlbNB2* allele homozygote seeds. Seed globulin fractions of the cultivar 'Shinano 1' and the *GlbNB2* homozygote (NB2) were digested (0, 0.5, 1, 2, 4 h) with trypsin followed by SDS-PAGE with CBB staining (A) and immunodetection with anti-NA (B). C indicates undigested control sample. RA(α) and NA(α) represent recombinant α polypeptide of Met-rich and zero-repeat subunit, respectively. Positions of 13S globulin α polypeptide (α) and β polypeptide (β) were indicated. Black and white arrowheads represent the α polypeptide of Met-rich and zero-repeat subunit, respectively.

Author contributions

TKT conceived and designed experiments. TO, KK, and NG performed experiments. TO and TKT analyzed the data and wrote the manuscript. All the authors have read and approved the final manuscript.

Conflict of interest

The authors have no conflicts of interest to declare.

CRediT authorship contribution statement

Takeyuki Okada: Writing – original draft, Investigation, Conceptualization. **Kohtaro Kimura:** Investigation. **Naruha Goto:** Investigation. **Tomoyuki Katsube-Tanaka:** Writing – review & editing, Writing – original draft, Funding acquisition, Data curation, Conceptualization.

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 JSPS. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Technical assistances were acknowledged to Mr. Akifumi Oishi, Mr. Naoto Mukainishi, and Ms. Madoka Sano.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochms.2024.100205.

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. https://doi.org/10.1073/pnas.0832158100

Chen, F., Li, H., Fan, X., Li, Y., Zhang, C., Zhu, L., Hu, J., KOMBE, A. J. K., Xie, J., Yin, D., Zhang, Y., Sun, J.-L., Tang, R. & Jin, T. (2021). Identification of a novel major allergen in buckwheat seeds: Fag t 6, Journal of Agricultural and Food Chemistry, 69 (45), 13315-13322. DOI: 10.1021/acs.jafc.1c01537.

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 and 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.

Katayama, S., Yamaguchi, D., Suzuki, Y., Athamneh, A. M. A., Mitani, T., Satoh, R., Teshima, R., Mine, Y., & 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, T., Kurisaka, N., Ogawa, M., Maruyama, N., Ohtsuka, R., Utsumi, S., & Takaiwa, F. (1999). Accumulation of soybean glycinin and its assembly with glutelins in rice. *Plant Physiology, 120*, 1063–1074.

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., Duldulao, J. B., Kimura, Y., Iida, S., Yamaguchi, T., Nakano, J., & Utsumi, S. (2004). The two subfamilies of rice glutelin differ in both primary and higher-order structures. *Biochimica et Biophysica Acta, 1699*, 95–102. https://doi.org/10.1016/j.bbapap.2004.02.001

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

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

Khan, N., 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.

Mari, A., Rasi, C., Palazzo, P., & Scala, E. (2009). Allergen databases: Current status and perspectives. *Current Allergy and Asthma Reports, 9*, 376–383.

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.

Okada, T., Monshi, F. I., Kudo, S., & Katsube-Tanaka, T. (2023). 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, 6*, Article 100159. https://doi.org/10.1016/j.fochms.2022.100159

Penin, A. A., Kasianov, A. S., Klepikova, A. V., Kirov, I. V., Gerasimov, E. S., Fesenko, A. N., & Logacheva, M. D. (2021). High-resolution transcriptome atlas and improved genome assembly of common buckwheat, Fagopyrum esculentum. *Frontiers in Plant Science, 12*. https://doi.org/10.3389/fpls.2021.612382

Peters, J. K., & Toor, N. (2015). Group II intron lariat: Structural insights into the spliceosome. *RNA Biology, 12*(9), 913–917. https://doi.org/10.1080/15476286.2015.1066956

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.

Sato, K., Hamada, M., Asai, K., & Mituyama, T. (2009). CENTROIDFOLD: A web server for RNA secondary structure prediction. *Nucleic Acids Research, 37*(Web Server issue), W277–W280. https://doi.org/10.1093/nar/gkp367

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, C., Zhang, R., Li, Y. M., Liang, N., Zhao, Y., Zhu, H., He, Z., Liu, J., Hao, W., Jiao, R., Ma, K. Y., & Chen, Z.-Y. (2017). Cholesterol-lowering activity of tartary buckwheat protein. *Journal of Agricultural and Food Chemistry, 65*(9), 1900–1906. https://doi.org/10.1021/acs.jafc.7b00066

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