Breeding Science 70: 118–127 (2020)
doi: 10.1270/jsbbs.19017

Research Paper

Structure and diversity of 13S globulin zero-repeat subunit, the trypsin-resistant storage protein of common buckwheat (*Fagopyrum esculentum* M.) seeds

Fakhrul Islam Monshi†[1,2], Nadar Khan†[1,3], Kohtaro Kimura[1], Seita Suzuki[1], Yuka Yamamoto[1] and Tomoyuki Katsube-Tanaka*[1]

[1] *Graduate School of Agriculture, Kyoto University*, Kitashirakawa, Kyoto 606-8502, Japan
[2] *Faculty of Agriculture, Sylhet Agricultural University*, Sylhet-3100, Bangladesh
[3] *Present address: Bio-Resources Conservation Institute, National Agricultural Research Centre (NARC)*, Park road ChakShehzad, Islamabad, Pakistan

The zero-repeat subunit of 13S globulin, which lacks tandem repeat inserts, is trypsin-resistant and suggested to show higher allergenicity than the other subunits in common buckwheat (*Fagopyrum esculentum* Moench). To evaluate allelic variations and find novel alleles, the diversity of the zero-repeat genes was examined for two Japanese elite cultivars and 15 Pakistani landraces. The results demonstrated that two new alleles *GlbNA1* and *GlbNC1*, plus three additional new alleles *GlbNA2*, *GlbNA3*, and *GlbND*, were identified besides the already-known *GlbNA*, *GlbNB*, and *GlbNC* alleles. In the Pakistani landraces, *GlbNA* was the most dominant allele (0.60–0.88 of allele frequency) in all except one landrace, where *GlbNB* was the most dominant allele (0.50 of allele frequency). Similar to *GlbNC*, the alleles *GlbNA2* and *GlbNA3* had extra ~200 bp MITE-like sequences around the stop codon. Secondary structure predictions of a sense strand demonstrated that the extra ~200 bp sequences of *GlbNC*, *GlbNA2*, and *GlbNA3* can form rigid hairpin structures with free energies of –78.95, –67.06, and –29.90 kcal/mol, respectively. These structures may affect proper transcription and/or translation. In the *GlbNC* homozygous line, no transcript of a zero-repeat gene was detected, suggesting the material would be useful for developing hypoallergenic buckwheat.

Key Words: allergen, common buckwheat, MITEs, 13S globulin, seed storage protein.

Introduction

Common buckwheat (*Fagopyrum esculentum* Moench) is considered a healthy food crop because of its well-balanced amino acid composition, high dietary fiber content, and beneficial physiological functions, such as anti-hypercholesterolemic, anti- hypertensive, anti-carcinogenic, and anti-inflammatory activities (Chen *et al.* 2008, Giménez-Bastida and Zieliński 2015, Liu *et al.* 2001, Tomotake *et al.* 2000, Zhang *et al.* 2012). Because of buckwheat's health-promoting and nutritional benefits, including good palatability, there has been an increase in its consumption and production in developed countries such as France, the United States, and Japan (FAO 2019, Katsube-Tanaka 2016). However, buckwheat seed contains allergenic proteins that cause immunoglobulin E (IgE)-mediated allergenic reactions in humans (Park *et al.* 2000, Satoh *et al.* 2014, Wieslander and Norbäck 2001). Reducing allergic reactions to buckwheat seed is becoming a focus of research with the goal of improving nutritional and physiological quality. This will improve the beneficial impacts of buckwheat food on human health.

To date, several allergenic proteins of buckwheat have been identified and categorized as most prevalent (24 kDa), or less prevalent (30, 43, and 67 kDa) (Park *et al.* 2000). More recently, Cho *et al.* (2015) considered 16, 24, 40, 43, and 48 kDa as major allergenic proteins of common buckwheat. 13S globulin, a salt soluble legumin-like protein, accounts for about 43% of the total seed proteins. 13S globulin is composed of disulphide-bonded heterogeneous acidic (α) and basic (β) polypeptides (Radović *et al.* 1996). The α and β polypeptides are biosynthesized as a larger single precursor with a signal peptide (**Fig. 1**). The signal peptide is processed during translocation to the endoplasmic reticulum. Then the polypeptide is further processed into α and β polypeptides. The β polypeptide of 13S globulin is 24 kDa, which is recognized as one of the major allergens, Fag e 1 (Nagata *et al.* 2000, Nair and Adachi 1999).

The 13S globulin subunits of common buckwheat have been categorized into methionine rich (Met-rich) and

Communicated by Ryo Ohsawa
Received February 1, 2019. Accepted August 25, 2019.
First Published Online in J-STAGE on December 17, 2019.
*Corresponding author (e-mail: tanakato@kais.kyoto-u.ac.jp)
† These authors contributed equally to this work

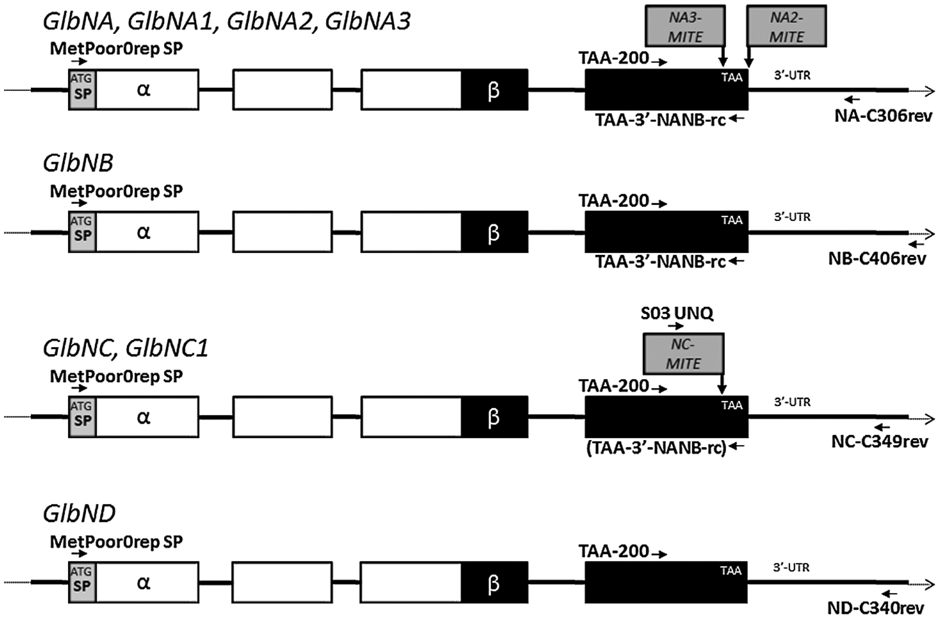

Fig. 1. Schematic representation of zero-repeat subunit gene structure and the position of PCR primers for genotyping and cloning. White, black and gray boxes indicate exon of α polypeptide, β polypeptide, and signal peptide or MITE-like sequence, respectively. Length of the boxes and lines between boxes indicate approximate length of nucleotide sequences. NA2-MITE, NA3-MITE, and NC-MITE indicate a MITE-like sequence specifically inserted in *GlbNA2*, *GlbNA3*, and *GlbNC* allele, respectively. Horizontal arrows indicate the position of PCR primers. Note that the primer 'TAA-3′-NANB-rc' was used for amplification of *GlbNC* but was not perfectly matched to *GlbNC*.

methionine poor (Met-poor) subunits. Both the subunits contain a pair of α and β polypeptides. In our previous report, Met-poor subunits were further divided into two types; 1) with variable number of tandem repeat sequences, and 2) without tandem repeat sequences (Khan *et al.* 2012). The tandem repeat sequences are hydrophilic with many arginine residues which are target of trypsin. Actually α polypeptides with the tandem repeat have been found to be easier to digest by trypsin than those without tandem repeat sequences.

Resistance to digestion by proteinases is one of the major characteristics of food allergens (Kopper *et al.* 2004, Sen *et al.* 2002). Therefore, the allergenicity of 13S globulin might be affected by the presence or absence of tandem repeat sequences. The allergenicity may be reduced by lowering the expression of genes with no tandem repeat sequences, also known as zero-repeat genes (Khan *et al.* 2012).

Recently, Sano *et al.* (2014) detected 17 open reading frames (ORFs) encoding 13S globulin from a genomic DNA library, indicating that the protein composes a multigene family, similarly as other 11S globulins such as glutelin of rice (*Oryza sativa* L.) and glycinin of soybean (*Glycine max* (L.) Merr.) do. Out of the 17 ORFs, there are two zero-repeat genes *GlbNA* (GenBank accession no. AB828117) and *GlbNB* (GenBank accession no. AB828118). Katsube-Tanaka *et al.* (2014) identified *GlbNC* (GenBank accession no. LC484359), which had ~200 bp of MITE (miniature inverted-repeat transposable element)-like sequence inserted ~70 bp upstream from the stop codon. The genes *GlbNA*, *GlbNB*, and *GlbNC* all belong to Met-poor subunits and show a high similarity to each other, possibly suggesting those genes are allelic. However, no other information about the diversity of zero-repeat genes in common buckwheat is available. Additionally, it is unknown whether or not the zero-repeat genes are located at a single locus. We also do not know if varietal and geographical differences in common buckwheat influence allelic frequency, or the structural variation of zero-repeat genes. Understanding the genetic variation of zero-repeat genes is important for the development of hypoallergenic buckwheat. Therefore, assessments of intra- as well as inter-varietal diversity of zero-repeat genes, including improved varieties, are needed.

The Gilgit-Baltistan region in northern Pakistan, which is surrounded by the Karakoram Range, the western Himalayas, the Pamir Mountains, and the Hindu Kush, is the western terminus of buckwheat cultivation in the Himalayan regions (Ohnishi 1994), connecting to the original birthplace of common buckwheat in the northwestern corner of Yunnan province (Ohnishi 1998). Buckwheat, as well as millet, are important crops in mountainous areas where maize cannot grow well due to the short summers (Ohnishi 1994). At least two waves of diffusion of common buckwheat have taken place in northern Pakistan; the first being pink flower genotypes which are currently growing in Hunza, Nagar, and the Hushe valley of Ghanche district, and the second being white flower genotypes in the Indus valley of Baltistan (Ohnishi 1994). Allozyme variability analysis indicated that northern Pakistan populations are closely related with those of Kashmir (Ohnishi 1994), which are distantly related to European populations (Ohnishi 1993). However, no other research on the genotypic diversity of buckwheat from Pakistan has been reported, except for phytochemicals (Abbasi *et al.* 2015), SDS-PAGE pattern of seed proteins (Hussain *et al.* 2016b), and major and trace elements (Hussain *et al.* 2016a). No systematic breeding program exists for the improvement of buckwheat crops. Thus, farmers grow locally adapted genotypes and landraces, which are assumed to conserve valuable genetic resources.

In this study, we briefly confirmed the three published zero-repeat genes were alleles on single locus in common buckwheat and identified new unique zero-repeat alleles, some of which have MITE-like sequences possibly to form a rigid hairpin secondary structure. We also evaluated the inter-variety diversity in allele frequency of zero-repeat genes using Japanese elite cultivars and Pakistani landraces. The findings of this study extend our knowledge about zero-repeat genes in common buckwheat, which will be useful for developing hypoallergenic buckwheat.

Materials and Methods

Plant materials and preparation of genomic DNA samples

The common buckwheat Japanese indigenous cultivars, ‘Harunoibuki’ and ‘Shinano1’, were used as plant materials for preliminary genotyping. Twenty seeds from each cultivar were grown in soil filled plastic pots. Young leaves from three to four week old seedlings were collected from individual plants and stored at –80°C for genomic DNA isolation. Genomic DNA was extracted and purified from leaves using DNeasy Plant Mini Kit (QIAGEN).

Fifteen Pakistani landraces from the Bio-resources Conservation Institute, National Agricultural Research Centre, Pakistan were used for genotyping (**Table 1**). Twenty seeds from each landrace were individually ground with MixerMill (QIAGEN) and used for genomic DNA extraction using DNAs-ici!-S (RIZO Inc., Tsukuba).

Isolation and identification of zero-repeat GlbNC and GlbND genes

The coding region of *GlbNC* was isolated from each seed that did not have the alleles *GlbNA* or *GlbNB* (Katsube-Tanaka *et al.* 2014), using PCR with forward (MetPoor0repSP: 5′-ATGTCTACGAAGCTCAATCTCTTCATCT-3′) and reverse (TAA-3′-NANB-rc: 5′-CGGAGCTCTTAAACGACGTCGTATCTCTC-3′) primers. PCR conditions were as follow: an initial denaturing step at 94°C for 2 min, 45 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 60 s, and extension at 72°C for 60 s, and the final extension step at 72°C for 7 min with ExTaq DNA polymerase (TaKaRa, Japan). The amplified fragment with larger molecular weight by 200 bp than that of *GlbNA* was isolated using an agarose gel and purified using FastGene Gel/PCR Extraction Kit (Nihon Genetics, Japan). The purified fragments were then cloned to pTAC-2 vector using TA cloning Kit (BioDynamics Laboratory Inc. Tokyo, Japan).

The coding and downstream regions of *GlbND* were isolated from seeds, which did not contain alleles *GlbNA*, *GlbNB*, or *GlbNC*, using PCR forward (MetPoor0repSP: 5′-ATGTCTACGAAGCTCAATCTCTTCATCT-3′) and reverse (0rep_TAA+1100rc: 5′-GATGAAGCTAGCCCTACGTACGAAC-3′) primers. PCR conditions were same as the described above except with the use of an extension time of 3 min.

Identification of 3′-Untranslated Region (3′-UTR) and downstream of 13S globulin gene GlbNC

For the isolation of the 3′-UTR of *GlbNC* gene, a genome walking approach was employed using Straight

Table 1. Sample numbers and accessions used for genotyping of Pakistani landraces

Sample #	Accession	Province	District	Town	Altitude (m)	Longitude	Latitude
1	3716	Gilgit-Baltistan	Skardu	Keris	2,200	75°58′	35°13′
2	3717	Gilgit-Baltistan	Skardu	Dognai	2,370	76°11′	35°15′
3	3724	Gilgit-Baltistan	Skardu	Fiazpur	2,405	75°42′	35°27′
4	3726	Gilgit-Baltistan	Ghanche	Lunkha	2,705	76°27′	35°05′
5	3728	n/a	n/a	n/a	n/a	n/a	n/a
6	3732	Gilgit-Baltistan	Gilgit	Murtaza Abad	2,245	74°35′	36°16′
7	21079	Gilgit-Baltistan	Gilgit	Aliabad	2,140	n/a	n/a
8	21081	Gilgit-Baltistan	Gilgit	Aliabad	2,140	n/a	n/a
9	29217	Gilgit-Baltistan	Hunza Nagar	Nasirabad	2,040	74°21′	36°16′
10	29218	Gilgit-Baltistan	Hunza Nagar	Karimabad	n/a	n/a	n/a
11	29221	Gilgit-Baltistan	Hunza Nagar	Nagar	n/a	n/a	n/a
12	29223	Gilgit-Baltistan	Skardu	Skardu	n/a	n/a	n/a
13	29227	Gilgit-Baltistan	Gilgit	n/a	n/a	n/a	n/a
14	29229	Gilgit-Baltistan	n/a	n/a	n/a	n/a	n/a
15	E1-Hol*	n/a	n/a	n/a	n/a	n/a	n/a

n/a, not applicable; * The accession is now unavailable in the genebank collection.

Walk Kit (BEX Co., Ltd, Tokyo, Japan). About 400 ng of genomic DNA extracted from a seed that had the *GlbNC* allele, but not the *GlbNA* and *GlbNB* alleles, was digested with XbaI restriction enzyme. After one nucleotide (dCTP) elongation, the DNA was ligated to SWA-2 adaptor using T4 DNA ligase (Nippon Gene, Tokyo, Japan). The primary PCR amplification was performed with walking primer-1 (5′-CGCAGGCTGGCAGTCTCTTTAG-3′) and sequence-specific primer-1 (Mpoor-TAA-200: 5′-ATTGGAGTGGGT GGAGTTGAAGACC-3′) along with KOD Plus DNA polymerase (Toyobo, Japan). PCR conditions were as follows: denaturation at 94°C for 2 min, followed by 35 cycles at 94°C for 30 s, 65°C for 30 s, and 68°C for 5 min. Following the primary PCR, a 100-fold diluted primary PCR product was used as a template for the secondary PCR. The secondary PCR was performed with walking primer-2 (5′-ATGC GGCCGCTCTCTTTAGGGTTACACGATTGCTT-3′) and sequence-specific primer-2 (Shina-S03UNQ: 5′-CTGACC CAACCAATAATTAAAGC-3′). The PCR conditions were same as for the primary PCR except that only 30 thermal cycles were performed. After the secondary PCR amplification, the PCR product was electrophoresed on agarose gel and the specific band was isolated and purified with FastGene Gel/PCR Extraction kit (Nihon Genetics Co. Ltd). Finally, the purified PCR product was cloned to pTAC-2 vector after the A-attachment reaction to acquire overhanging dA at the 3′-ends (Toyobo, Japan).

Genotyping of 13S globulin zero-repeat genes

Genotyping of 13S globulin zero-repeat genes (*GlbNA*, *GlbNB*, *GlbNC*, and *GlbND*) was conducted for single seed genomic DNA using common forward primer (Mpoor-TAA-200: 5′-ATTGGAGTGGGTGGAGTTGAAGACC-3′) and allele-specific (allele group-specific) reverse primers designed at 3′-UTR and downstream of *GlbNA* (NA-C306-rev: 5′-GAGACATGAATACGACGGGTGTTG-3′), *GlbNB* (NB-C406-rev: 5′-GCTTAACATCATTCCGTTACCGG-3′), *GlbNC* (NC-C349-rev: 5′-TGCTGTTTCGGACTTTTC CTCC-3′), and *GlbND* (ND-C340_rev: 5′-GCTTCCGAAC GATCCCTTAATGCAAG-3′) genes. PCR was carried out under the following conditions: an initial denaturing step at 94°C for 2 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 65°C for 60 s, and extension at 72°C for 60 s, and a final extension step at 72°C for 7 min using ExTaq DNA polymerase (TaKaRa, Japan) or KOD FX neo (Toyobo, Japan) with recommended thermal conditions as described in the instruction manual.

Secondary structure prediction of DNA and molecular evolutionary analysis

The secondary structures of MITE-like sequences of sense and antisense strands were analyzed with the computer programs Centroid Fold (Hamada *et al.* 2009, Sato *et al.* 2009) (http://rtools.cbrc.jp/centroidfold/), and Mfold (DNA folding form) that was developed by Zuker (2003) using the free energies rules by SantaLucia (1998) and the salt correction of Peyret (2000) (http://unafold.rna.albany.edu/?q=mfold/DNA-Folding-Form). Evolutionary divergence analysis was conducted using the Kimura 2-parameter in MEGA6 (Tamura *et al.* 2013).

Development of GlbNC-homozygous line

GlbNC-containing plants were identified using the allele-specific and *GlbNC*-MITE specific primer (Shina_S03_UNQ: 5′-CTGACCCAACCAATAATTAAAGC-3′) from the cultivar ‘Harunoibuki’. The *GlbNC* containing plants were naturally crossed with each other in the isolated environment of a phytotron to avoid cross pollination with other genotypes. The genotype of the *GlbNC*-homozygous line was confirmed with the allele group-specific primers and the size of PCR amplified products of the coding region.

For RNA experiments, immature seeds of the *GlbNC*-homozygous line and the ‘Harunoibuki (non *GlbNC*-homozygous)’ were sampled periodically, frozen in liquid nitrogen, and stored in a freezer at –80°C. Total RNA was extracted with RNAs-ici!-P (Rizo, Japan) and treated with DNaseI that was followed by column purification (Total RNA Extraction Column, Favorgen). RT-PCR was employed with TaKaRa RNA PCR Kit (AMV) Ver. 3.0 (TaKaRa, Japan) using a random 9 mer primer for cDNA synthesis and primers of Fw: MetPoor 0 repeat SP, ATGTC TACGAAGCTCAATCTCTTCATCT and Rv: TAA_3′_0rep_rc, TTAAACGACGTCGTATCTsyCCC for zero-repeat genes; and Fw: BW MetPoor SP, ATGTCAACTAA ACTCATACTCTCCTTCT and Rv: TAA 3′ 1A2A3D4A rc, CGGAGCTCTTAAACTATGGAGAAACGCTC for repeat-containing genes of 13S globulin of succeeding PCRs.

Results

3′-UTR and downstream region of 13S globulin gene GlbNC and development of specific primers

Because the coding region homology is high between *GlbNA*, *GlbNB*, and *GlbNC*, with the exception of the MITE-like sequence of *GlbNC*, the 3′-UTR and downstream region of *GlbNC* was isolated to develop specific primers to distinguish the three genes. The alignment of the 3′-UTR and downstream region of *GlbNC*, plus the regions of *GlbNA* and *GlbNB*, which were determined by primer walking of BAC clones 269I19 and 336B7 that were isolated by Sano *et al.* (2014), respectively, demonstrated large downstream sequence variations more than 250 bp away from stop codon (**Supplemental Fig. 1**). Evolutionary divergence between *GlbNA* and *GlbNB*, *GlbNB* and *GlbNC*, *GlbNA* and *GlbNC* were 0.026, 0.022, 0.022, respectively. This resulted in the development of specific primers for amplifying *GlbNA*, *GlbNB*, and *GlbNC*.

Preliminary genotyping of zero-repeat genes in Japanese elite cultivars

Genotyping of zero-repeat genes was performed using Japanese indigenous cultivars ‘Harunoibuki’ and

‘Shinano1’, with the specific primers designed at 3′-UTR and downstream of *GlbNA*, *GlbNB*, and *GlbNC*. Because of the cross-pollination and diploid natures of common buckwheat, individual seed is expected to have different gene composition, thus 20 individual plants from each variety were analyzed separately. In ‘Harunoibuki’, the fragment amplification with *GlbNA*-specific primer resulted in two bands with the expected size of 500 bp, and a smaller one by 24 bp (**Fig. 2**). Thus, the corresponding genes were designated as *GlbNA* and *GlbNA1*, respectively. The amplification with *GlbNB*-specific primer showed a single band of 600 bp corresponding to the authentic *GlbNB* gene. The PCR amplification with *GlbNC*-specific primer resulted in two fragments that were approximately 760 bp and smaller 550 bp. Thus, the corresponding genes for the smaller band was named *GlbNC1*. The difference in fragment size was due to the insertion of the MITE-like sequence of *GlbNC*. The genotyping results demonstrated that all the seeds of two different cultivars showed only one or two fragments, suggesting the allelism of the five genes. Thus, hereafter, the five genes are treated as allele. Because *GlbNA* plus *GlbNA1*, and *GlbNC* plus *GlbNC1*, were amplified with the same specific primer, respectively, the alleles within each combination were treated as belonging to the same allele group.

When the varietal differences and allele frequencies in the two Japanese indigenous cultivars were compared, the allele frequencies of *GlbNA* were highest in both cultivars (0.33 and 0.53), followed by the *GlbNC1* and *GlbNC* alleles in ‘Harunoibuki’, and the *GlbNB* and *GlbNC1* alleles in ‘Shinano1’ (**Table 2**). The *GlbNA1* allele was observed only in ‘Harunoibuki’. Thus, the allele frequencies of *GlbNB* and *GlbNC*, as well as *GlbNA1*, differed between the two cultivars, suggesting that this genotyping methodology seems to be useful for a varietal characterization.

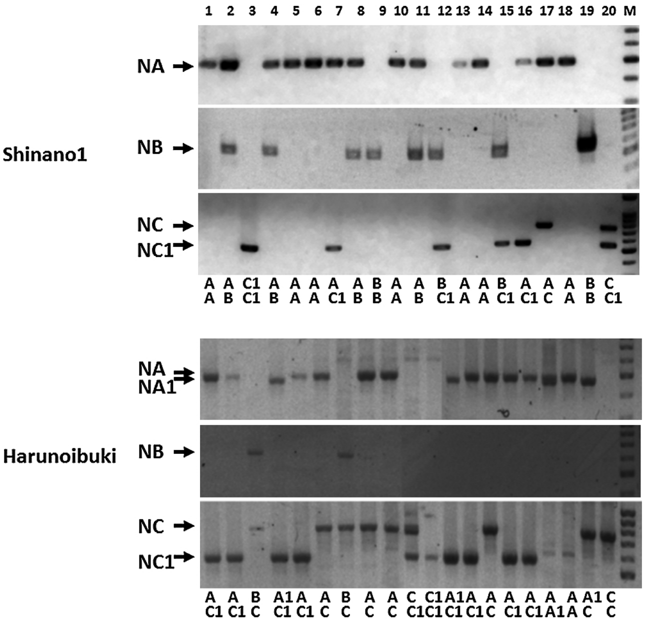
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 M
Shinano1
NA
NB
NC
NC1
A A C1 A A A A A B A A B A A B A A A B C
A B C1 B A A C1 B B A B C1 A A C1 C1 C A B C1
Harunoibuki
NA
NA1
NB
NC
NC1
A A B A1 A A B A A C C1 A1 A A A A A A A1 C
C1 C1 C C1 C1 C C C C C1 C1 C1 C1 C C1 C1 A1 A C C

Fig. 2. Genotyping of Japanese elite cultivars ‘Shinano1’ and ‘Harunoibuki’ for zero-repeat subunit gene of 13S globulin. Twenty seeds each cultivar were analyzed by three rounds of PCR with allele group-specific primers for *GlbNA (GlbNA1)*, *GlbNB*, and *GlbNC (GlbNC1)*. The position of amplified fragments were shown by horizontal arrows with abbreviated names (NA, NA1, NB, NC, NC1) of the alleles (*GlbNA*, *GlbNA1*, *GlbNB*, *GlbNC*, *GlbNC1*). The lane M indicates molecular size marker. The letters (A, A1, B, C, C1) under the figures indicate the combination of alleles (*GlbNA*, *GlbNA1*, *GlbNB*, *GlbNC*, *GlbNC1*) identified for each seed.

Table 2. Allele frequency for zero-repeat subunit gene of 13S globulin (n = 20)

Allele	Harunoibuki	Shinano1
GlbNA	0.33	0.53
GlbNA1	0.1	0
GlbNB	0.05	0.25
GlbNC	0.25	0.05
GlbNC1	0.28	0.18
Total	1	1

Genotyping of zero-repeat genes in Pakistani landraces

A similar genotyping was conducted for 20 seeds from each of the 15 Pakistani landraces, most of which were collected in Gilgit-Baltistan region in northern Pakistan (**Fig. 3**). Because two seeds of #4_6 and #6_9 could not be genotyped with the three allele group-specific primers, the zero-repeat genes of the two seeds were isolated and sequenced. Subsequently, the new allele designated as *GlbND* was identified and a new allele group-specific primer for *GlbND* was developed and added to the genotyping routine.

The genotyping results showed the allele frequencies of *GlbNA* were highest in all landrace accessions (0.600–0.875), except for the E1-Hol landrace accession where the *GlbNB* was the highest (0.500), and the *GlbNA* allele frequency was the second highest (0.48) (**Table 3**). With the exception of sample #15 (acc. E1-Hol), the maximum allele frequencies of *GlbNA1*, *GlbNB*, *GlbNC*, *GlbNC1*, and *GlbND* were 0.075 (#1), 0.150 (#3, 10, 14), 0.050 (#1, 2), 0.150 (#3), and 0.225 (#4), respectively. The combined allele frequencies of non-*GlbNA* were high in most accessions from Skardu and Ghanche, which are located at the east side of the Gilgit-Baltistan (0.325: #1, 0.400: #2, 0.375: #3, 0.400: #4). The lowest allele frequencies of non-*GlbNA* (0.125–0.250) were found in the accessions from Hunza-Nagar and Gilgit, which are located at the middle and the north side of the Gilgit-Baltistan (#6, 7, 8, 9, 10, 11, 13) (**Fig. 4**).

Secondary structure prediction of MITE-like sequences

During the genotyping and preliminary genotyping of *GlbNA(NA1)* in sample #11, two larger fragments with MITE-like sequences were amplified and the alleles were named *GlbNA2* and *GlbNA3* (**Fig. 1**). The MITE-like

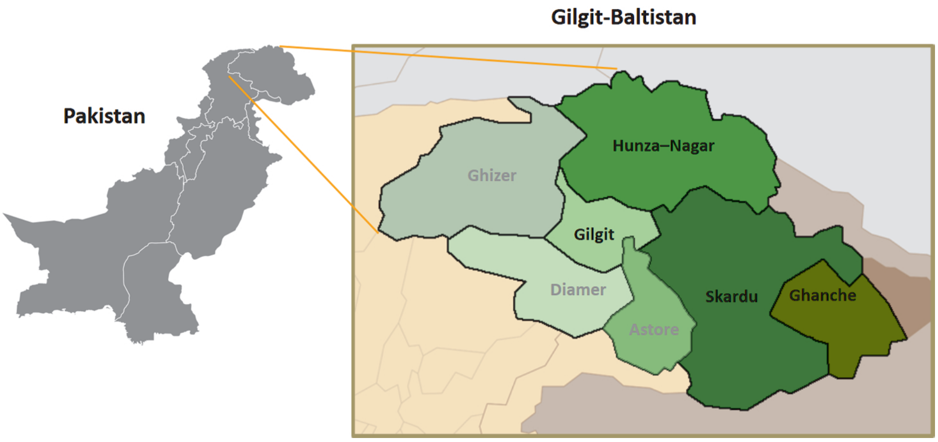

Fig. 3. The maps of Pakistan and its Gilgit-Baltistan region where most of Pakistani buckwheat landrace germplasms were collected.

Table 3. Allele frequency and expected heterozygosity of zero repeat gene

Sample #	Allele								Total
	NA	NA1	NA2	NA3	NB	NC	NC1	ND	
1	0.675	0.075	0	0	0.075	0.050	0.075	0.050	1
2	0.600	0.050	0	0	0.100	0.050	0.125	0.075	1
3	0.625	0	0	0	0.150	0.025	0.150	0.050	1
4	0.600	0	0	0	0.050	0.025	0.100	0.225	1
5	0.800	0	0	0	0.075	0	0.100	0.025	1
6	0.825	0	0	0	0.025	0	0.050	0.100	1
7	0.825	0	0	0	0.125	0	0.050	0	1
8	0.875	0.050	0	0	0.075	0	0	0	1
9	0.825	0	0	0	0.125	0	0.050	0	1
10	0.750	0.025	0	0	0.150	0	0.025	0.050	1
11	0.800	0	0.025	0.025	0.100	0	0	0.050	1
12	0.850	0.050	0	0	0.100	0	0	0	1
13	0.775	0	0	0	0.100	0.025	0.100	0	1
14	0.700	0	0	0	0.150	0.025	0.100	0.025	1
15	0.475	0.025	0	0	0.500	0	0	0	1
p	0.733	0.018	0.002	0.002	0.127	0.013	0.062	0.043	1
p^2	0.538	0.000	0.000	0.000	0.016	0.000	0.004	0.002	0.560

p; mean allele frequency: Expected heterozygosity is 0.44.

sequence of *GlbNA*3 (229 bp) was inserted 70 bp upstream of the stop codon, at exact position of the MITE-like sequence of *GlbNC* (208 bp), with the same direct repeat sequence of 5′-TGGTATTTTCC-3′. Meanwhile the MITE-like sequence of *GlbNA2* (198 bp) was inserted downstream of stop codon, with direct repeat sequence of 5′-GTTTAAA GG-3′.

The predicted secondary structures of the MITE-like sequences of *GlbNC*, *GlbNA3*, and *GlbNA2* had significant numbers of hydrogen bonds with high base-pairing probability, resulting in a rigid hairpin structure with free energies at 37°C of –78.95, –67.06, and –29.90 kcal/mol for the sense strand, and –76.58, –71.65, and –24.24 kcal/mol for the antisense strand (**Fig. 5**). The MITE-like sequences forming rigid hairpin secondary structures might affect the expression or function of *GlbNC*, *GlbNA3*, and *GlbNA2* genes and may have evolutionary implications.

Development of the GlbNC homozygous line and expression of the zero-repeat gene

Because the allele frequency of *GlbNC* was five-fold higher in ‘Harunoibuki’ than that in ‘Shinano1’, we isolated *GlbNC* containing plants from ‘Harunoibuki’ and cross pollinated the plants to enrich the *GlbNC* allele. In the next generation, *GlbNC* homozygous plants with long-style and short-style flowers were isolated to propagate the *GlbNC* homozygous seeds. Within the maturing seeds of the *GlbNC* homozygous line, no transcripts of a zero-repeat gene were detected, whereas transcripts of repeat containing genes were detected (data not shown). In the maturing seeds of ‘Harunoibuki (non *GlbNC* homozygous)’, both transcripts of the zero-repeat gene and the repeat-containing gene were detected.

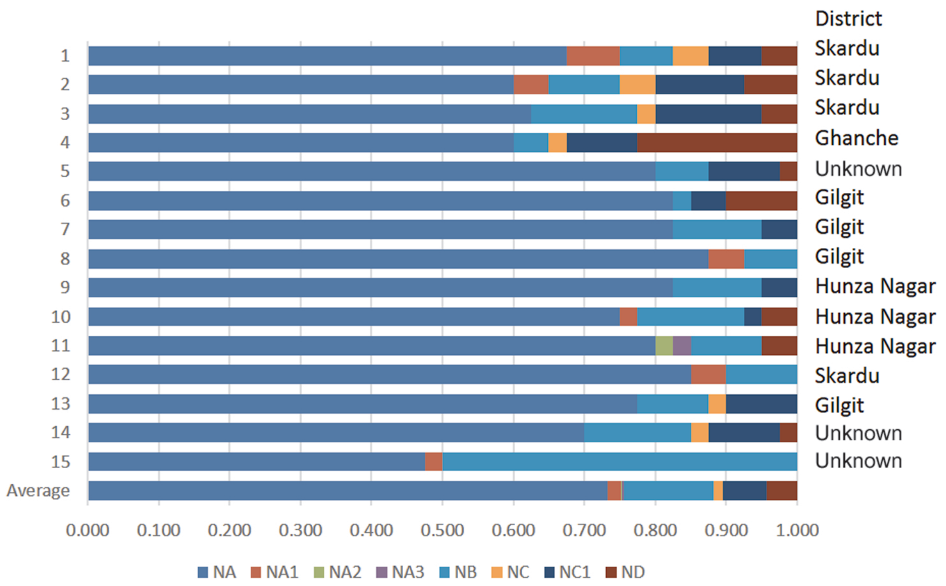

Fig. 4. Comparison of allele frequencies for zero-repeat subunit gene and the collected location of germplasms. Allele frequency for *GlbNA*, *GlbNA1*, *GlbNA2*, *GlbNA3*, *GlbNB*, *GlbNC*, *GlbNC1*, and *GlbND* and the origin of the accessions were compared for fifteen accessions. The sample number of the fifteen accessions were shown left. Mean allele frequency was also calculated.

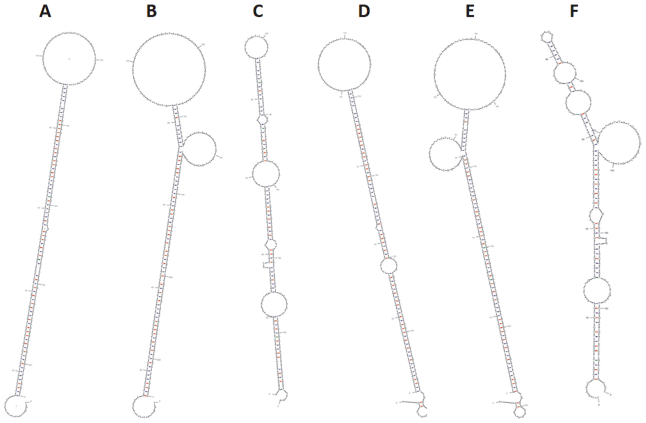

Fig. 5. Predicted secondary structures of the MITE-like sequences found in *GlbNC*, *GlbNA3*, and *GlbNA2*. The secondary structures of sense (panels A, B, and C) and antisense (panels D, E, and F) strands of MITE-like sequences were analyzed with the computer program Mfold (DNA folding form) for *GlbNC* (panels A and D), *GlbNA3* (panels B and E), and *GlbNA2* (panels C and F).

Discussion

Importance of characterization of zero-repeat genes and identification of new alleles

The zero-repeat genes (without tandem repeat sequences) of 13S globulin of common buckwheat encode a trypsin resistant protein that is, therefore, considered to show higher allergenicity than the other repeat-containing, trypsin digestible 13S globulins (Khan *et al.* 2012). Because the genes have not been fully described, a characterization of the zero-repeat genes is crucially important in understanding allergenic allelic variation. Thus, the primary focus of this study was to explore a novel allele of the zero-repeat gene, which could be useful for developing a hypoallergenic buckwheat using the diversified sequence information at 3′-UTR and downstream. We identified two new alleles, *GlbNA1* and *GlbNC1*, from the Japanese cultivars and three new alleles *GlbNA2*, *GlbNA3*, and *GlbND*, from the Pakistani germplasms. Yasui *et al.* (2016) developed the Buckwheat Genome DataBase (BGDB) that is composed of 387,594 scaffolds, 286,768 predicted coding sequences (CDSs), and 36,763 annotated CDSs, which were generated from a draft assembly of the buckwheat genome using short reads of 264.5 Gb obtained by Next-Generation Sequencing (NGS) of Illumina HiSeq 2000. A BLAST search of the BGDB with the *GlbNA* sequence as a query resulted in finding a complete, but intron mis-predicted, sequence of *GlbNC1* (Fes_sc0022676.1.g000001.aua.1), plus an incomplete and divided sequence of *GlbNA* (Fes_sc0203109.1.g000001.aua.1, Fes_sc0028260.1.g000002.aua.1), an indication of the difficulty in the analysis of complicated 13S globulin sequence structures with short reads of NGS. Meanwhile, an amplicon deep sequence for zero-repeat gene coding regions divided into quarters, which were amplified from pooled genomic DNA in several cultivars, showed large numbers of diversified sequences but did not detect a novel allele containing a large insert similar to the *GlbNC* sequence (data not shown). Thus, the traditional genotyping and cloning methodology, such conducted in this study, are still effective for the analysis of diversified 13S globulin genes in common buckwheat.

MITE-like sequences of GlbNA2, GlbNA3, and GlbNC

Three types of MITE-like sequences of ~200 bp were found around the stop codon of the alleles *GlbNA2*, *GlbNA3*, and *GlbNC*. A NCBI-BLAST search for the three MITE-like sequences resulted in the identification of only 25–33 bp similar sequences in other organisms, whereas a BLAST search (BLASTN 2.2.26) in the BGDB showed many hits in other scaffolds of the BGDB. For example, *GlbNC*-MITE yielded 152 sequences with E-values better than 1.0e-01, with many having high scores with extra 8–9 bases of thymine in the middle of sequence. *GlbNA3*-MITE gave 30 sequences with E-values better than 1.0e-01, with many having high scores that lack approximately 80 bases in the middle. The sequence of these 80 bases was not found in the BGDB. *GlbNA2*-MITE gave 50 sequences with E-values better than 1.0e-01. These results suggest that the three MITE-like sequences are unique in common buckwheat and are widely dispersed in the genome.

The secondary structure predictions demonstrated that the extra 200 bp sequence of *GlbNC* might form a rigid hairpin structure. MITEs play important roles in gene regulation and genome evolution (El Amrani *et al.* 2002, Kuang *et al.* 2009, Naito *et al.* 2006, Oki *et al.* 2008, Yang *et al.* 2005). MITEs may down regulate gene expression through MITE-derived small RNAs by either disrupting or altering gene structure (Kuang *et al.* 2009). The down regulation of genes by MITEs insertion has been reported by several researchers. For example, genome-wide analysis in rice showed that genes associated with MITEs have significantly lower expression than genes away from MITEs, and all genes with MITEs insertions have a larger proportion of weakly expressed genes than the genes with no MITEs insertions (Chen *et al.* 2012, Lu *et al.* 2012).

Formation of hairpin or stem-and-loop structures affects gene expressions, DNA recombination, and DNA transposition (Lah *et al.* 2011), plus inhibition of molecular biology techniques such as PCR and sequencing (Nelms and Labosky 2011). mRNA pseudoknots mediate ribosomal frameshifts or acts as roadblocks, whereby synthesis of multiple proteins is controlled from a single polycistronic mRNA (Tholstrup *et al.* 2012). Information carried by RNA in their primary, secondary, and tertiary structures influence the transcription, splicing, cellular localization, translation and turnover of RNA (Wan *et al.* 2011). Hairpins, stem-loops or triplexes formed during the lagging strand synthesis would disrupt DNA replication causing slippage or blockage (Aguilera and Gómez-González 2008). Thus, the possible secondary structure formation in the MITE-like sequences should be noted from this study.

The reason the MITE-like sequences were found only around the stop codons is noteworthy. A possible explanation is that the primers used for genotyping amplified the region between 200 bp upstream to 300–400 bp downstream from a stop codon. If we focused on other regions with other primers, we may have detected other MITE-like sequences being inserted. Another possible explanation may be that the C-terminus of β polypeptide or the 3′-UTR may be structurally tolerant to short polypeptide insertions and truncations or changes in mRNA secondary structure. Before verifying such hypotheses, expression and accumulation of mRNA and polypeptides that coincide with MITE-like sequences should be carefully studied.

Evaluation of the allele frequency of zero-repeat gene

The second focus of this study was to identify allele frequencies of *GlbNC* and other zero-repeat alleles. Although the number of zero-repeat gene loci have not been determined and is necessary to be examined with the progeny produced by the cross between each allele, the maximum number of alleles detected in a single seed did not exceed two, suggesting the locus is only one in a diploid common buckwheat genome. The number of zero-repeat alleles detected in the BGDB (Yasui *et al.* 2016) and the BAC clone library (Sano *et al.* 2014, Yasui *et al.* 2008) was also two. Meanwhile, we categorized eight zero-repeat alleles, with four allele group-specific primers. The four primers amplified at least one fragment in all of the single seeds examined, but the possibility of an unknown allele exist has not been eliminated. The current genotyping methodology will identify an unknown allele only when the unknown allele is homozygous. We examined more than 300 seeds, so at least one unknown allele-homozygous seed would be detected if the allele frequency of the unknown allele exceeded 0.058, which is estimated by the following calculation: $0.058^2 \times 300 = 1.009$, which is greater than one.

The observed heterozygosity at the zero-repeat locus averaged 0.36 in the 15 Pakistani germplasms (108 heterozygous seeds out of 300 examined seeds), while the expected heterozygosity was 0.44 (**Table 3**), suggesting a reduced diversity due to interbreeding.

Thus, when considering the small number of genotyped seeds (20 seeds from each line), the accuracy in the evaluation of allele frequency may have been limited in this study. Nonetheless, the detection of the *GlbNC* allele was without ambiguity and the allele frequency of *GlbNC* seemed reliable. Consequently ‘Harunoibuki’ showed the highest allele frequency of *GlbNC*, with a value of 0.25 and this cultivar was used for development of *GlbNC* homozygous lines.

Development of the GlbNC homozygous lines

Seven out of 15 accessions of Pakistani landraces showed amplified fragments derived from the allele *GlbNC*, with the allele frequency of *GlbNC* at a maximum of 0.05 in the accessions. Meanwhile the allele frequency of *GlbNC* in ‘Harunoibuki’ was 0.25, indicating that the cultivar would be good for developing a *GlbNC* homozygous line. After the enrichment of *GlbNC*-alleles in isolated populations, we successfully developed a *GlbNC* homozygous line.

Because the inserted MITE-like sequence of *GlbNC* was predicted to form a rigid hairpin structure, the transcription and/or translation of the *GlbNC* was expected to be affected. At a minimum, the translational product would be truncated because a new stop codon appeared in the MITE-like sequence. At the moment, RT-PCR experiments demonstrated no zero-repeat gene transcripts were detected in the *GlbNC* homozygous line, although repeat-containing gene transcripts were detected in both the *GlbNC* homozygous line and 'Harunoibuki (non *GlbNC* homozygous)'. Future work should be conducted, including western blot experiments with zero-repeat subunit-specific antibodies to determine the accumulation of zero-repeat subunits in *GlbNC* homozygous lines.

The accumulation of zero-repeat subunits may vary across other types of alleles and genotypes. The MITE-like sequences containing alleles *GlbNA2* and *GlbNA3* will be interesting for analysis. Even though the alleles *GlbNA2* and *GlbNA3* were found in only one of the 300 seeds, both alleles were identified in the accession '3728', suggesting this accession is a promising germplasm to further analyze for finding a new allele.

Further genotyping using a wide range of buckwheat genetic resources, including land races as well as improved varieties, might offer an opportunity for discovering more diversity in zero-repeat genes. In our study, we only assessed the zero-repeat gene diversity in a limited number of Pakistani landraces and Japanese elite cultivars; nonetheless, our results will be helpful for better understanding zero-repeat genes. The findings of our current study will contribute to the efforts of developing hypoallergenic buckwheat.

Author Contribution Statement

T.K.-T. designed the project, and wrote the initial draft of the manuscript. F.I.M. and N.K. contributed to analysis and interpretation of data, and assisted in the preparation of the manuscript. The other authors have contributed to data collection and interpretation.

Acknowledgments

This work was supported by a fellowship to N.K. by the Japan Society for the Promotion of Science (JSPS) and a grant by the Ministry of Education, Culture, Sports, Science, and Technology, Japan [Scientific Research (C) 23580020, 2011-2013 to T.K.-T.]. Technical assistance provided by Ms. Madoka Sano, Mr. Naoto Mukainishi, and Ms. Mariko Nakagawa is acknowledged.

Literature Cited

Abbasi, R., S. Janjua, A. Rehman, K. William and S.W. Khan (2015) Some preliminary studies on phytochemicals and antioxidant potential of *Fagopyrum esculentum* cultivated in Chitral, Pakistan. J. Anim. Plant Sci. 25: 576–579.

Aguilera, A. and B. Gómez-González (2008) Genome instability: a mechanistic view of its causes and consequences. Nat. Rev. Genet. 9: 204–217.

Chen, J., C. Lu, Y. Zhang and H. Kuang (2012) Miniature inverted-repeat transposable elements (MITEs) in rice were originated and amplified predominantly after the divergence of Oryza and Brachypodium and contributed considerable diversity to the species. Mob. Genet. Elements 2: 127–132.

Chen, Z.-Y., R. Jiao and K.Y. Ma (2008) Cholesterol-lowering nutraceuticals and functional foods. J. Agric. Food Chem. 56: 8761–8773.

Cho, J., J.-O. Lee, J. Choi, M.-R. Park, D.-H. Shon, J. Kim, K. Ahn and Y. Han (2015) Significance of 40-, 45-, and 48-kDa proteins in the moderate-to-severe clinical symptoms of buckwheat allergy. Allergy Asthma Immunol. Res. 7: 37–43.

El Amrani, A., L. Marie, A. A ïnouche, J. Nicolas and I. Couée (2002) Genome-wide distribution and potential regulatory functions of *AtATE*, a novel family of miniature inverted-repeat transposable elements in *Arabidopsis thaliana*. Mol. Genet. Genomics 267: 459–471.

FAO (2019) FAOSTAT, Statistics Division, Food and Agriculture Organization of the United Nations. Date accessed on 19th Jan 2019 at http://www.fao.org/faostat/en/#home

Giménez-Bastida, J.A. and H. Zieliński (2015) Buckwheat as a functional food and its effects on health. J. Agric. Food Chem. 63: 7896–7913.

Hamada, M., H. Kiryu, K. Sato, T. Mituyama and K. Asai (2009) Prediction of RNA secondary structure using generalized centroid estimators. Bioinformatics 25: 465–473.

Hussain, I., A. Bano, Faizanullah and A. Nosheen (2016a) Multivariate analysis for elemental composition among indigenous common buckwheat genotypes of Baltistan. J. Anim. Plant Sci. 26: 1725–1731.

Hussain, I., A. Bano, F. Ullah, M. Ali and S.K. Sherwani (2016b) Genotypic variation among indigenous Common Buckwheat of Baltistan based on SDS-PAGE. Int. J. Biol. Biotech. 13: 171–176.

Katsube-Tanaka, T., M. Nakagawa, M. Sano and Y. Yasui (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—. J. Crop Res. 59: 31–35.

Katsube-Tanaka, T. (2016) Buckwheat production, consumption, and genetic resources in Japan. *In*: Zhou, M., I. Kreft, S.-H. Woo, N. Chrungoo and G. Wieslander (eds.) Molecular Breeding and Nutritional Aspects of Buckwheat. Elsevier, Amsterdam, pp. 61–80.

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

Kopper, R.A., N.J. Odum, M. Sen, R.M. Helm, J.S. Stanley and A.W. Burks (2004) Peanut protein allergens: Gastric digestion is carried out exclusively by pepsin. J. Allergy Clin. Immunol. 114: 614–618.

Kuang, H., C. Padmanabhan, F. Li, A. Kamei, P.B. Bhaskar, S. Ouyang, J. Jiang, C.R. Buell and B. Baker (2009) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the *Solanaceae*: new functional implications for MITEs. Genome Res. 19: 42–56.

Lah, J., M. Seručnik and G. Vesnaver (2011) Influence of a hairpin loop on the thermodynamic stability of a DNA oligomer. J. Nucleic Acids 2011: 513910.

Liu, Z., W. Ishikawa, X. Huang, H. Tomotake, J. Kayashita, H. Watanabe and N. Kato (2001) A buckwheat protein product suppresses 1, 2-dimethylhydrazine-induced colon carcinogenesis in rats by reducing cell proliferation. J. Nutr. 131: 1850–1853.

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

Nagata, Y., K. Fujino, S. Hashiguchi, N. Abe, Y. Zaima, Y. Ito, Y. Takahashi, K. Maeda and K. Sugimura (2000) Molecular characterization of buckwheat major immunoglobulin E-reactive proteins in allergic patients. Allergol. Int. 49: 117–124.

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

Naito, K., E. Cho, G. Yang, M.A. Campbell, K. Yano, Y. Okumoto, T. Tanisaka and S.R. Wessler (2006) Dramatic amplification of a rice transposable element during recent domestication. Proc. Natl. Acad. Sci. USA 103: 17620–17625.

Nelms, B.L. and P.A. Labosky (2011) A predicted hairpin cluster correlates with barriers to PCR, sequencing and possibly BAC recombineering. Sci. Rep. 1: 106.

Ohnishi, O. (1993) Population genetics of cultivated common buckwheat, *Fagopyrum esculentum* Moench. VIII. Local differentiation of land races in Europe and the silk road. Jpn. J. Genet. 68: 303–316.

Ohnishi, O. (1994) Buckwheat in Karakoram and the Hindukush. Fagopyrum 14: 17–25.

Ohnishi, O. (1998) Search for the wild ancestor of buckwheat III. The wild ancestor of cultivated common buckwheat, and of tatary buckwheat. Econ. Bot. 52: 123–133.

Oki, N., K. Yano, Y. Okumoto, T. Tsukiyama, M. Teraishi and T. Tanisaka (2008) A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. *japonica*. Genes Genet. Syst. 83: 321–329.

Park, J.W., D.B. Kang, C.W. Kim, S.H. Koh, H.Y. Yum, K.E. Kim, C.S. Hong and K.Y. Lee (2000) Identification and characterization of the major allergens of buckwheat. Allergy 55: 1035–1041.

Peyret, N. (2000) Prediction of Nucleic Acid Hybridization: Parameters and Algorithms PhD dissertation, Wayne State University, Department of Chemistry, Detroit, MI.

Radović, S.R., V.R. Maksimović and E.I. Varkonji-Gašić (1996) Characterization of buckwheat seed storage proteins. J. Agric. Food Chem. 44: 972–974.

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

SantaLucia, J.J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc. Natl. Acad. Sci. USA 95: 1460–1465.

Sato, K., M. Hamada, K. Asai and T. Mituyama (2009) CentroidFold: a web server for RNA secondary structure prediction. Nucleic Acids Res. 37: W277–W280.

Satoh, R., R. Nakamura, M. Ohnishi-Kameyama and R. Teshima (2014) Identification of IgE-binding proteins in buckwheat. Clin. Transl. Allergy 4: P13.

Sen, M., R. Kopper, L. Pons, E.C. Abraham, A.W. Burks and G.A. Bannon (2002) Protein structure plays a critical role in peanut allergen stability and may determine immunodominant IgE-binding epitopes. J. Immunol. 169: 882–887.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30: 2725–2729.

Tholstrup, J., L.B. Oddershede and M.A. Sørensen (2012) mRNA pseudoknot structures can act as ribosomal roadblocks. Nucleic Acids Res. 40: 303–313.

Tomotake, H., I. Shimaoka, J. Kayashita, F. Yokoyama, M. Nakajoh and N. Kato (2000) A buckwheat protein product suppresses gallstone formation and plasma cholesterol more strongly than soy protein isolate in hamsters. J. Nutr. 130: 1670–1674.

Wan, Y., M. Kertesz, R.C. Spitale, E. Segal and H.Y. Chang (2011) Understanding the transcriptome through RNA structure. Nat. Rev. Genet. 12: 641–655.

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

Yang, G., Y.H. Lee, Y. Jiang, X. Shi, S. Kertbundit and T.C. Hall (2005) A two edged role for the transposable element Kiddo in the rice ubiquitin2 promoter. Plant Cell 17: 1559–1568.

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

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

Zhang, Z., M.-L. Zhou, Y. Tang, F.-L. Li, Y.-X. Tang, J.-R. Shao, W.-T. Xue and Y.-M. Wu (2012) Bioactive compounds in functional buckwheat food. Food Res. Int. 49: 389–395.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31: 3406–3415.