Rapid genotyping with DNA micro-arrays for high-density linkage mapping and QTL mapping in common buckwheat (Fagopyrum esculentum Moench)

Rapid genotyping with DNA micro-arrays for high-density linkage mapping and QTL mapping in common buckwheat (*Fagopyrum esculentum* Moench)

Shiori Yabe[†1)], Takashi Hara[†2)], Mariko Ueno[3)], Hiroyuki Enoki[4)], Tatsuro Kimura[4)], Satoru Nishimura[4)], Yasuo Yasui[3)], Ryo Ohsawa[2)] and Hiroyoshi Iwata*[1)]

[1)] *Graduate School of Agricultural and Life Sciences, The University of Tokyo*, 1-1-1 Yayoi, Bunkyo, Tokyo 113-8657, Japan
[2)] *Graduate School of Life and Environmental Sciences, University of Tsukuba*, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan
[3)] *Graduate School of Agriculture, Kyoto University*, Kitashirakawa, Kyoto 606-8502, Japan
[4)] *Future Project Division, TOYOTA MOTOR CORPORATION*, 1 Toyota-cho, Toyota, Aichi 471-8572, Japan

For genetic studies and genomics-assisted breeding, particularly of minor crops, a genotyping system that does not require *a priori* genomic information is preferable. Here, we demonstrated the potential of a novel array-based genotyping system for the rapid construction of high-density linkage map and quantitative trait loci (QTL) mapping. By using the system, we successfully constructed an accurate, high-density linkage map for common buckwheat (*Fagopyrum esculentum* Moench); the map was composed of 756 loci and included 8,884 markers. The number of linkage groups converged to eight, which is the basic number of chromosomes in common buckwheat. The sizes of the linkage groups of the P1 and P2 maps were 773.8 and 800.4 cM, respectively. The average interval between adjacent loci was 2.13 cM. The linkage map constructed here will be useful for the analysis of other common buckwheat populations. We also performed QTL mapping for main stem length and detected four QTL. It took 37 days to process 178 samples from DNA extraction to genotyping, indicating the system enables genotyping of genome-wide markers for a few hundred buckwheat plants before the plants mature. The novel system will be useful for genomics-assisted breeding in minor crops without *a priori* genomic information.

Key Words: common buckwheat, *Fagopyrum esculentum* Moench, array-based genotyping system, genome-wide linkage analysis, QTL analysis.

Introduction

Advances in next-generation sequencing and genotyping technologies have paved the way for the application of genomics-assisted breeding (Varshney *et al.* 2005) to various major crops, such as maize, wheat, and rice. In the last several years, minor (or orphan) crops, which are those crops that have received less attention from researchers compared with major crops, have also started to garner attention as targets for genomics-assisted breeding (Armstead *et al.* 2009, Varshney *et al.* 2012). Minor crops are of great importance in the marginal environments of Africa, Asia and South America because they are often more nutritious and more adapted to the harsh environments of these areas than many of the major crop varieties. Because they play an important role in regional food security, the development of genomic resources has rapidly progressed for various minor crops (e.g., Ly *et al.* 2013, Varshney *et al.* 2013).

Common buckwheat, *Fagopyrum esculentum* Moench

(2*n* = 2*x* = 16), is an example of a minor yet important crop. Common buckwheat is easy to grow and can be adapted to almost all types of soil, except for sandy, heavy crusted, or wet soils. Genus *Fagopyrum* has great potential as a crop for forage, human consumption, and medicine because of its high-quality protein content (Campbell 1997). For common buckwheat in particular, its rapid rate of growth and ability to grow at high altitudes make it an especially important crop in the mountainous regions of South Asia (e.g., Baniya 1990, Joshi 1999). Despite its importance, the genetic improvement of common buckwheat has achieved only limited success, mainly due to buckwheat's heteromorphic self-incompatibility (Campbell 1997). Common buckwheat is a heteromorphic, self-incompatible species that has heterostylous flowers controlled by the *S*-locus (reviewed in Lewis and Jones 1992). Genomics-assisted breeding is expected to accelerate the genetic improvement of common buckwheat; however, the genomic resources currently available remain incomplete. Furthermore, the development of genomic resources for use in common buckwheat breeding has been hampered because buckwheat is the only crop species within the Polygonaceae family, and therefore is phylogenetically distant to other major crops and model plants. While some linkage map has been published (e.g., Konishi

and Ohnishi 2006, Pan and Chen 2010, Yasui *et al.* 2004), the utility of the map has been limited due to buckwheat's large genetic diversity.

The potential of genomics-assisted breeding has been demonstrated in various plant species (e.g. Varshney *et al.* 2013). Yabe *et al.* (2014) expect genomic selection (Meuwissen *et al.* 2001), which is a promising method in genomics-assisted breeding, to accelerate the genetic improvement of allogamous crops such as common buckwheat. However, to fully utilize genomic selection, a rapid, high-throughput genotyping technology is necessary because a large number of genome-wide markers need to be genotyped for all of the plants in a breeding population before the plants mature. To widen the scope for the application of genomics-assisted breeding technologies, genotyping technologies should be able to conduct genotyping or constructing linkage map even without any *a priori* genomic information of target species.

In the present study, we evaluated the potential of a novel marker system (Enoki *et al.* 2012, Iehisa *et al.* 2014) for use as a rapid, high-throughput genotyping technology for genomics-assisted breeding. We used common buckwheat as the test plant, and conducted a genome-wide linkage analysis without using any *a priori* genomic information (e.g., DNA sequences) or any existing genomic tools (e.g., DNA markers). Moreover, to verify the utility of the linkage map we constructed, we applied it to a quantitative trait loci (QTL) analysis of the same population. We selected two plants from a breeding population and crossed them to generate a segregating population. This approach can be generalized to linkage and QTL mapping in other crop species, thus the potential of the novel marker system demonstrated in this study may be feasible in various 'orphan' or 'less-studied' crops. Since common buckwheat is a completely allogamous species, breeding populations have large effective population sizes; common buckwheat is therefore expected to have a low level of linkage disequilibrium (Sved 1971). Linkage map containing a high number of markers (i.e., a high-density linkage map) could be highly beneficial for the success of genomics-assisted breeding.

Materials and Methods

Development of mapping population and agronomic characterization

To construct a linkage map and detect QTL related to agronomic characteristics, we employed 92FE1-F4, a population produced by bulk crossing among 'Tempest', 'Kitawasesoba', 'Natsusoba', and 'Shinanonatsusoba'. These cultivars are classified into a single agroecotype: summer type. We developed a mapping population consisting of 178 F_1 progeny derived from a cross between P1 and P2, which were selected from 92FE1-F4. In the present study, P1 had thrum-type flowers and P2 had pin-type flowers. The genotype of the *S*-locus was heterozygous (i.e., *S/s*) in P1, and recessive homozygous (*s/s*) in P2.

The F_1 progeny were sowed at a density of one seed per plastic pot (diameter, 24 cm; height, 24 cm) on 6 August 2012 and cultivated under natural conditions in an isolation chamber (L × W × H, 630 × 540 × 230 cm) at the University of Tsukuba (36°06′N, 140°05′E). We examined main stem length as the trait for QTL mapping after harvesting on 9 October 2012.

DNA extraction

Total genomic DNA from both parental individuals and the F_1 progeny was extracted as follows: About 100 mg milled leaf tissue was mixed with 500 μl lysis buffer (0.3% sodium dodecyl sulfate, 20 mM Tris-HCl [pH 8.0 at 25°C], 5 mM EDTA, 400 mM NaCl) and 4 μl RNase A (100 mg/ml; Qiagen Inc., CA, USA), and incubated at 65°C for 10 min. The mixture was then centrifuged (14,000 rpm, 20°C) for 2 min. Buffer AP2 (130 μl; Qiagen Inc.) was added to the cleared lysate and the samples were mixed, incubated on ice for 5 min, and centrifuged for 2 min (14,000 rpm, 20°C). Then, 500 μl isopropanol (Wako Pure Chemical Industries, Ltd., Osaka, Japan) was added to the supernatant, and the samples were mixed by inversion, incubated for 5 min at room temperature, and centrifuged for 10 min. The DNA pellet was washed twice with 500 μl of 70% ethanol (Wako Pure Chemical Industries, Ltd.). After each wash, the sample was centrifuged for 5 min, and 500 μl of 99.5% ethanol (Wako Pure Chemical Industries, Ltd.) was added to the pellet. After incubation for 5 min at room temperature, the sample was centrifuged for 5 min and resuspended in 100 μl Tris-EDTA buffer.

Sequence library construction and next-generation sequencing analysis for marker preparation

Total genomic DNA was prepared from the mapping parents (P1 and P2) and from 40 plants of the 92FE1-F4 population (hereafter referred to as the 40-mix population). Total genomic DNA was digested with *Pst*I (New England Biolabs, Inc., Ipswich, MA) at 37°C for 1 h and then with *Bst*NI (New England Biolabs) at 60°C for 1 h. Genomic DNA fragments obtained were ligated with *Pst*I-sequence adapters (5′-CACGATGGATCCAGTGCA-3′, 5′-CTGGAT CCATCGTGCA-3′) by using T4 DNA Ligase (New England Biolabs) at 16°C for 16 hours, after which, polymerase chain reaction (PCR) amplifications were performed with Primestar DNA polymerase (Takara Bio, Inc., Otsu, Japan). The PCR primer was designed based on the adapter sequences (5′-GATGGATCCAGTGCAG-3′). The PCR conditions were as follows: 30 cycles at 98°C for 10 s, 55°C for 15 s, and 72°C for 1 min, followed by 1 cycle at 72°C for 3 min. These steps amplify only the *Pst*I–*Pst*I DNA fragment but not *Pst*I–*Bst*NI fragments. After sonication, DNA fragments from 200 to 500 bp were recovered by electrophoresis on a 1.5% agarose gel. DNA fragments from P1 and P2 were recovered separately, and two sequence libraries (LIB-P1 and LIB-P2) were constructed by using a Paired-End DNA Sample Preparation Kit (Illumina, Inc.,

San Diego, CA). PCR products from each plant in the 40-mix population were equally mixed before sonication and sizing, and a sequence library (LIB-40mix) was constructed by using a Paired-End DNA Sample Preparation Kit (Illumina). LIB-P1 and LIB-P2 were analyzed by using an Illumina GAII with a paired-read of 300 nt, and LIB-40mix was analyzed by using an Illumina Hiseq2000 with a paired-read of 200 nt. The sequences were submitted to the DNA Data Bank of Japan (submission no. DRA001140).

Processing of raw reads and design of probes for microarray analysis

The procedure for the processing of row reads and the design of probes was similar to the one provided in Iehisa *et al.* (2014), in which the same genotyping system was applied to the linkage mapping of common wheat. First, we performed trimming of adaptor and primer sequences and low-quality $3'$ ends (quality score below 2) from raw reads using in-house developed software. Discarding pairs of reads when either of pairwise reads had length of <70 nt, we obtained three trimmed sequence libraries, TRIM-P1, TRIM-P2 and TRIM-40mix. CONT-P1, CONT-P2 and CONT-40mix were then assembled, respectively, from TRIM-P1, TRIM-P2 and TRIM-40mix using Velvet ver. 1.1.05 (Zerbino and Birney 2008). For the assemblies, velveth was run with a k-mer size of 65, and velvetg was run with expected coverage: 'auto', coverage cut-off: 'auto', insert length of 220 bp, minimum contig length of 100 bp, and allowing scaffolding. From assembled contigs and reads unused in the assemblies, tiling probes were designed with a Tm of around 76°C and the length of 50–75 bp. The probe design fundamentals were described in the NimbleGen technical note (http://www.roche-biochem.jp/products/2010/08/array/pdf/probe_design_2008_06_04.pdf). To construct a mapping array, designed probes were first aligned to two trimmed libraries, TRIM-P1 and TRIM-P2, by using Bowtie ver. 0.12.7 (Langmead *et al.* 2009), where no more than 3 mismatches were allowed and up to 30 alignments were reported (-v 3, -k 30). P1-derived probes that met all the following five conditions were selected: (1) the number of aligned P1 reads with no mismatch was larger than 3, (2) the ratio of aligned P1 reads with no mismatch over aligned P1 reads with one or two mismatches was larger than 4, (3) the ratio of aligned P1 reads with no mismatch over aligned P2 reads with less than three mismatch was larger than 10. Similar selection was performed for P2 probes. All candidate probes from CONT-40mix were mapped to TRIM-40mix by using the Bowtie software, and the probes that the number of aligned reads with no mismatch was larger than 30 were selected. Then, the selected probes were aligned to the trimmed sequence library TRIM-166, which were obtained from *Pst*I–*Pst*I DNA fragments that did not harbor *Bst*NI sites, was obtained from 166 F_1 plants derived from the 40-mix population. The number of aligned reads with no mismatch to the sequences of each plant was counted. Probes that showed identical patterns of absence/presence

of probe hits were merged. All probes from P1, P2 and 40mix and control probes were synthesized in triplicate on a NimbleGen HD-2 135K × 12plex microarray (Roche Diagnostics, Madison, WI). Although the company has terminated services for the microarrays, microarrays provided by other companies, such as Agilent Technologies, Inc., will also be usable for the genotyping assay of the marker system.

Hybridization of microarray

Genotypes of the parental plants, P1 and P2, and their 178 F_1 progeny were determined by using HD-2 135K × 12plex microarrays as described. After digestion of total DNA with *Pst*I and *Bst*NI, adapter ligation, PCR amplification and purification of PCR products, as described in library preparation section, they were labelled with Cy3 using NimbleGen One-Color DNA Labeling Kit (Roche Diagnostics) according to the manufacturer's instructions. Hybridization was performed at 42°C during 24 hours on NimbleGen Hybridization System (Roche Diagnostics). The microarrays were scanned with a NimbleGen MS200 Microarray Scanner (Roche Diagnostics) and genotype calling was performed based on the signal intensity. Present/absent calls of the probe-sets reflected the present/absent DNA fragments hybridized to the probe-sets. Thus, DNA polymorphisms genotyped with this system were used as dominant markers.

Linkage map construction

Linkage maps of P1 and P2 were constructed by using the pseudo-testcross strategy (Grattapaglia and Sederoff 1994). First, the deviation from the Mendelian segregation ratio was tested for each marker by using the Chi-square test ($p < 0.01$; statistically significance). Markers segregating in a 1 : 1 ratio were used to construct the linkage map of P1 and P2. Markers were assigned to linkage groups by setting the recombination rate threshold at 0.3 and the threshold for the minimum number of markers at 3. Locus ordering was performed by using AntMap software (Iwata and Ninomiya 2006) with a 50-run of an optimization process (i.e., maximization of the log-likelihood). The Kosambi mapping function (Kosambi 1943) was used to calculate map distances. To connect linkage groups constructed in the P1 and P2 linkage map, recombination rates between markers segregating in a 3 : 1 ratio and markers represented in the P1 and P2 map were calculated. Among the markers segregating in a 3 : 1 ratio, those that had low recombination rates (<0.02) with both markers on the P1 and P2 map were used as "bridges" between the P1 and P2 map. When multiple markers on the P1 or P2 map had recombination rates <0.02 and were segregated in a 3 : 1 ratio, the centers of gravity of these multiple markers were connected by a bridge.

QTL analysis

For the QTL analysis, 171 plants for which both marker data and phenotypic data (i.e., scores of main stem length

[cm]) were available were used. Composite interval mapping (Zeng 1993, 1994) was performed by using the QTL Cartographer software ver. 1.17 (Basten *et al.* 2003). To analyze our pseudo-testcross data, we employed a model for inbred backcross design. In the analysis, we adopted the linkage phase estimated at the step of linkage map construction. A permutation test with 100 replicates was performed for each trait to estimate the empirical threshold value corresponding to the 5% significance level. The proportion of total (i.e., phenotypic) variance explained by each of detected QTL was calculated as ([residual variance under the null hypothesis]–[residual variance under the alternative hypothesis])/[phenotypic variance].

Results

Construction of microarray

To design probes for developing a microarray, we sequenced complexity-reduced genomic DNA samples from P1, P2 and 40mix-population. After trimming and filtering of raw reads, libraries of 13,055,602, 11,576,694 and 267,438,466 reads were generated (TRIM-P1, TRIM-P2

and TRIM-40mix; Fig. 1). All libraries were assembled independently by using the Velvet, yielding 9,054 contigs and 5,535,673 unused reads for TRIM-P1, 8,187 contigs and 4,684,155 unused reads for TRIM-P2, 63,558 contigs and 128,957,022 unused reads for TRIM-40mix. From assembled contigs and reads unused in the assemblies, tiling probes were designed. The probes developed from P1 and P2 were aligned to the TRIM-P1 and TRIM-P2, and the probes that showed presence/absence polymorphisms were selected. The probes developed from 40mix were aligned to the TRIM-40mix, and the probes that showed more than 30 hits were selected. The selected probes were then mapped to TRIM-166, and probes that showed identical patterns of absence/presence of probe hits were merged. Finally, we obtained 16,583 probes from P1, 16,584 from P2, and 11,480 from 40mix population.

Preparing the data for linkage map construction

DNA microarray genotyping was performed for 44,836 markers (i.e., 44,647 probes mentioned above and 189 control probes). For marker genotypes, we treated three signal types: present call, absent call and missing. The present/
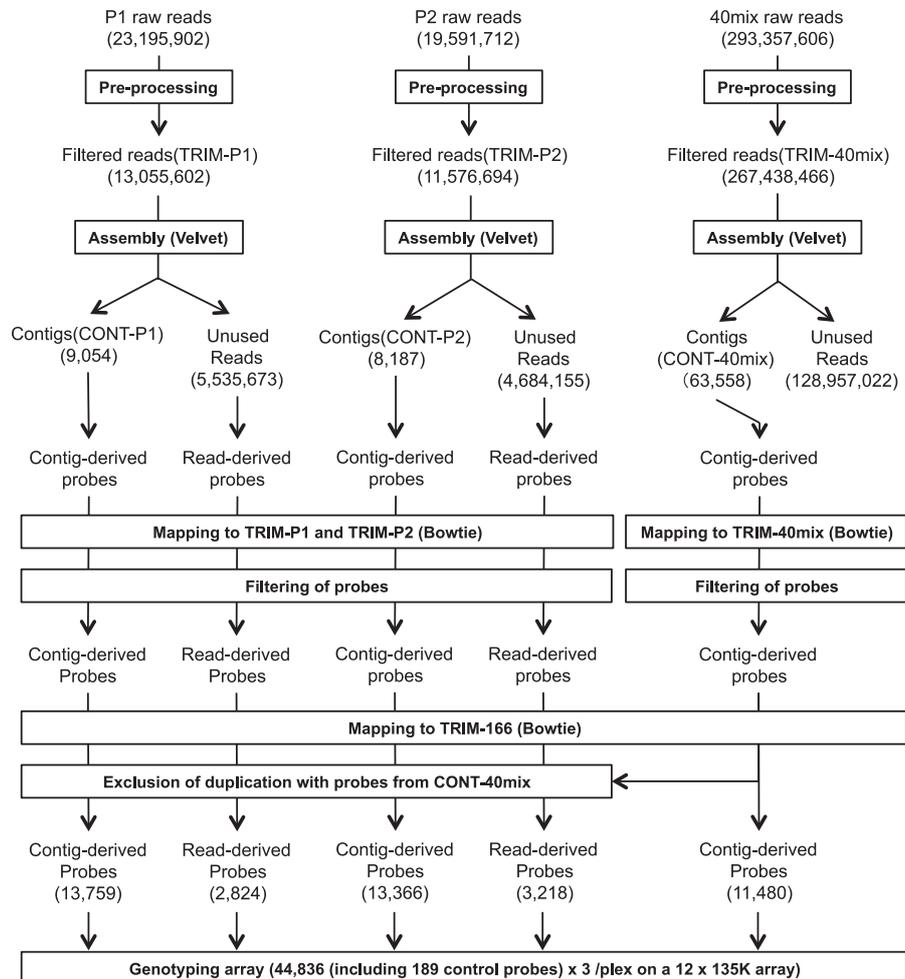

Fig. 1. Data flow diagram of the processing of raw sequence reads and the design of probes for microarray analysis. See text for details.

absent calls were determined according to the bimodal microarray signals. The ambiguous signals were classified as missing values. From those markers, we used 16,841 markers that were heterozygous in one or both parents for further analysis. In the F_1 population of 178 plants, the numbers of markers segregating in 1 : 1 and 3 : 1 were 14,442 (P1: 6,875 markers; P2: 7,567 markers) and 2,399, respectively. A segregation ratio of 1 : 1 is expected when one parent has a heterozygous genotype and the other has a recessive homozygous genotype. A segregation ratio of 3 : 1 is expected when both parents have a heterozygous genotype. Of the markers with segregation ratio of 1 : 1 and 3 : 1, 9,112 (P1: 4,325 markers; P2: 4,787 markers) and 1,701 markers, respectively, showed a clear distinction between the two genotypes. Of all 10,813 markers, 1,339 markers contained one or more missing values. Of these, 824 contained missing values for less than four plants, and 106 contained missing values for more than 17 (10%) plants. Markers derived from an identical contig with the same segregation type (i.e., P1 or P2 markers with a 1 : 1 segregation ratio or bridging markers with a 3 : 1 segregation ratio) were expected to reflect genotypes at the location of the contig because of the close linkage between the markers. We therefore collected markers from a single contig to make consensus genotypes. The missing genotypes of the collected markers were imputed on the basis of the consensus genotypes. In a few cases, two or three patterns of marker genotypes were observed in one contig. In those cases, we treated the groups of markers having different genotypes as separate contigs (sep-contigs). When sep-contigs derived from a single contig belonged to different segregation types, we excluded them from the subsequent analysis. Sep-contigs from a sin-

gle contig showing a low level of consensus (i.e., genotypes were discordant in more than 30% of plants) were also excluded. We used flower morphology (i.e., pin or thrum) controlled by the S-locus as the phenotypic marker. The marker was located on the P1 map because P1 was heterozygous (i.e., *S/s*) at the locus. In total, we had 1,455, 1,631, and 869 contigs for P1, P2, and bridging markers, respectively. Next, to analyze the pseudo-testcross data, we inverted the genotype data by duplicating and converting all of the contigs (i.e., homozygous genotypes were converted into heterozygous genotypes and vice versa). We grouped contigs that had an identical genotype into a single marker group and then separated those marker groups into three marker types—P1 markers, P2 markers, and bridging markers. Marker groups were grouped into two and four clusters for P1 and P2 type, respectively, via a single-linkage cluster analysis based on Manhattan distances among marker groups. The distances were calculated based on the genotypes of marker groups, which were scored with 0 and 1. We used single-linkage clustering because the method is similar to the one used in the linkage grouping. In total, we attained 346, 410, and 360 marker groups, respectively, for P1, P2, and bridging markers. The genotypes of these marker groups were used to construct the linkage map.

Linkage map construction

By using a pseudo-testcross strategy, we constructed linkage map for P1 and P2, and connected them with bridging markers (Fig. 2; Supplemental Tables 1, 2). We mapped 346 loci on the P1 map, and 410 loci on the P2 map (Table 1; Supplemental Table 1). We used 283 groups segregating 3 : 1 (bridging markers), which represented small
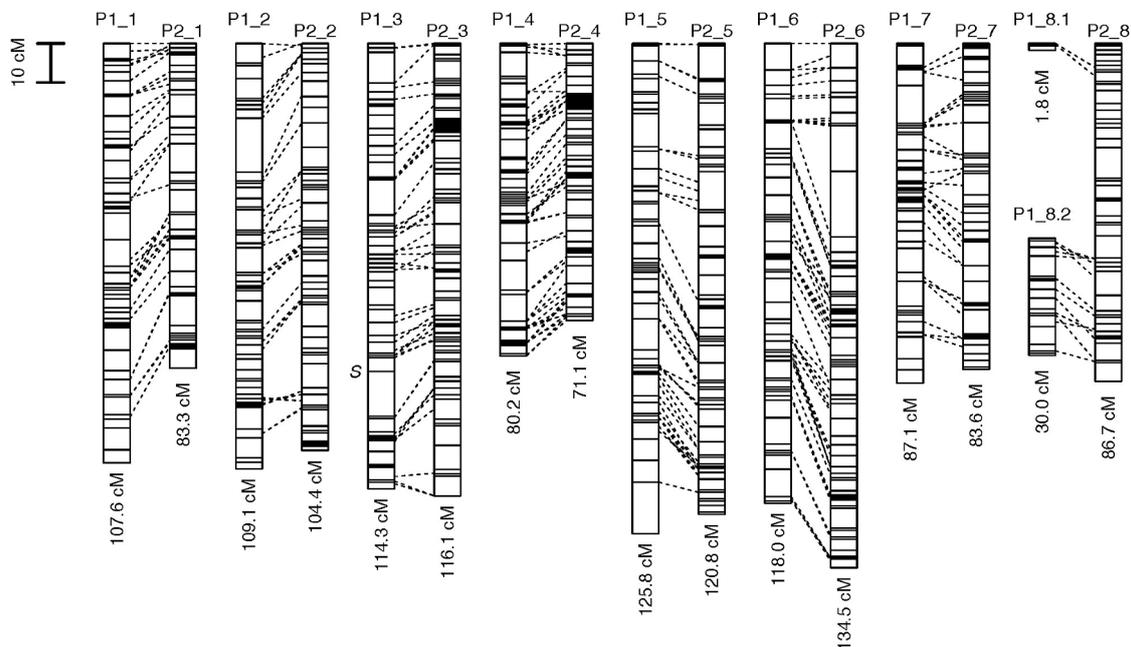
Fig. 2. Linkage map for common buckwheat. Dashed lines represent "bridge" markers that connect marker groups between the linkage maps of P1 and P2. Phenotypic marker position (i.e., *S*-locus) is indicated by "*S*".

Table 1. Summary information for the linkage map of common buckwheat

| Linkage group	No. of loci	Genetic length (cM)	Average interval of adjacent loci (cM)	No. of contigs	No. of markers
P1_1	45	107.6	2.45	198	674
P1_2	51	109.1	2.18	211	549
P1_3	51	114.3	2.29	223	691
P1_4	42	80.2	1.96	193	511
P1_5	47	125.8	2.73	203	626
P1_6	49	118.0	2.46	196	560
P1_7	41	87.1	2.18	161	453
P1_8.1	3	1.8	0.88	10	18
P1_8.2	17	30.0	1.88	60	145
P2_1	37	83.3	2.32	119	235
P2_2	54	104.4	1.97	206	621
P2_3	70	116.1	1.68	273	886
P2_4	49	71.1	1.48	184	498
P2_5	57	120.8	2.16	217	598
P2_6	57	134.5	2.40	224	695
P2_7	43	83.6	1.99	189	544
P2_8	43	86.7	2.06	219	580
P1	346	773.8	2.30	1,455	4,227
P2	410	800.4	1.99	1,631	4,657
All	756	1,574.3	2.13	3,086	8,884

recombination rates (i.e., <0.02) with loci in both the P1 and P2 map, to combine the loci in the P1 and P2 linkage map (Supplemental Table 2). We used markers segregating 3 : 1 only for the P1 and P2 map bridging, not for mapping on the linkage map. Because the precision of the estimation of recombination rates between markers segregating 3 : 1 and between markers segregating 1 : 1 and 3 : 1 was low (Ritter *et al.* 1990), the inclusion of 3 : 1 markers in the linkage map was thought to affect the estimation of the linkage map positions of 1 : 1 markers. Thus, as shown in Fig. 2, we used bridging markers only to connect the linkage groups between the P1 and P2 linkage map. The phenotypic marker,

(flower morphology) was located on linkage group P1_3 ("*S*" in Fig. 2). After connecting the P1 and P2 linkage map, the number of linkage groups converged to eight, which is the basic chromosome number of common buckwheat. The eighth linkage group was divided into two groups of short length in the P1 map. The P1 and P2 linkage map covered 773.8 and 800.4 cM, and contained 1,455 and 1,631 contigs, consisting of 4,227 and 4,657 markers, respectively (Table 1; Supplemental Table 1). The means of the intervals between adjacent positions were 2.30 and 1.99 cM (Table 1) and the medians were 1.68 and 1.15 cM in the P1 and P2 linkage map, respectively. Most (90%) adjacent positions had intervals shorter than 5.07 cM.

On the linkage map, one position (i.e., a single marker group) harbored a number of contigs, and one contig consisted of a number of markers. Fig. 3 shows the number of contigs per loci (A) and the number of markers per contig (B). Among the 756 loci on the map, 555 loci consisted of more than one contig, and 492 loci consisted of less than 10 contigs. Among the 3,086 contigs, 1,140 contigs consisted of more than one marker, and among those, 1,036 contigs consisted of less than 15 markers and 13 contigs consisted of more than 35 markers.

QTL analysis

Phenotypic values for main stem length observed in the mapping population had a unimodal, continuous distribution (Fig. 4), suggesting that main stem length is controlled by multiple QTL and is influenced by environmental effects. For main stem length, significant QTL were detected at map positions 9.3 cM on the P1-1 group, 49.0 cM on the P1-2 group, 9.0 cM on the P1-5 group, and 16.9 cM on the P2-4 group (Fig. 5; Table 2). The four QTL accounted for 5.64% to 8.51% of the phenotypic variance observed in main stem length (Table 2).
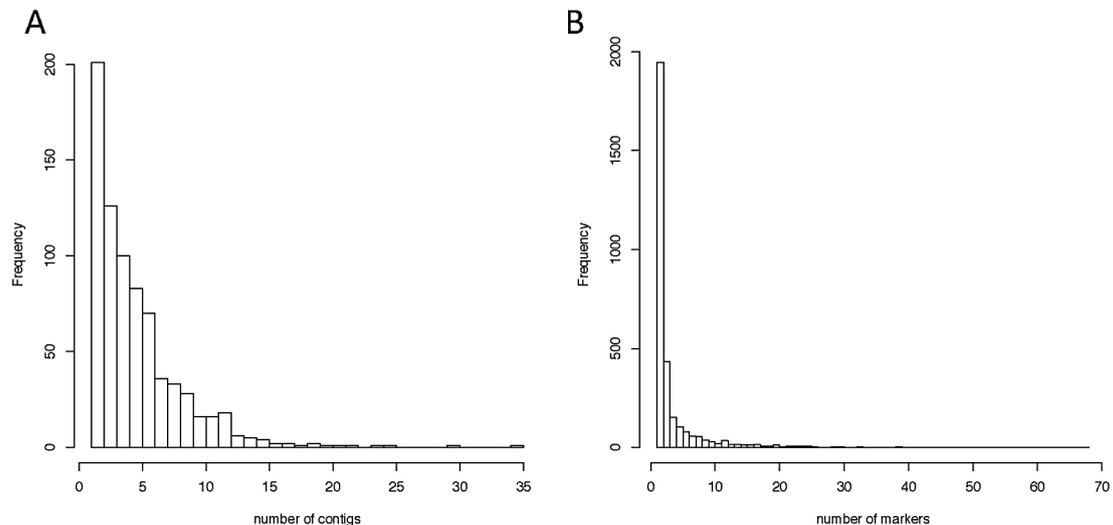
Fig. 3. Distribution of the number of contigs per map position (A) and the number of markers per contig (B).
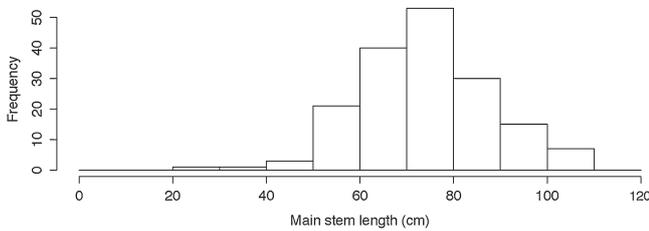
Fig. 4. Distribution of phenotypic value for main stem length.

Discussion

By using a novel marker system, we successfully constructed a high-density linkage map for common buckwheat. The number of linkage groups converged to eight, which is the basic number of chromosomes in common buckwheat. The size of the linkage groups in the P1 and P2 linkage map were 773.8 and 800.4 cM, respectively. Some linkage maps have previously been developed for common buckwheat. Yasui *et al.* (2004) constructed a linkage map with eight linkage groups covering a total of 508.3 cM that contained 223 amplified fragment-length polymorphism (AFLP) markers. However, some of these markers were not included in the eight linkage groups. Konishi and Ohnishi (2006) conducted linkage mapping with the pseudo-testcross strategy. Their female map had 12 linkage groups covering 911.3 cM that contained 54 microsatellite and 77 AFLP markers; their male map had 12 linkage groups covering 909.0 cM that contained 37 microsatellite and 34 AFLP markers. The female and male map converged to eight paired linkage groups, although three of the linkage groups in the female map did not converge to the eight groups. Pan and Chen (2010) constructed a linkage map with 10 linkage groups covering a total of 692.4 cM that included 87 randomly amplified polymorphic DNA markers, 12 sequence-tagged sites, four seed protein subunit markers, and three morphological character alleles. Hara *et al.* (2011) constructed a linkage map with nine linkage groups covering a

total of 311.6 cM that contained two photoperiod-sensitivity candidate genes and 63 expressed sequence tag markers. Compared with the previously constructed map, the map constructed in the present study had the largest number of markers on one side (i.e., P1 map or P2 map), and the average interval of adjacent markers was 2.13 cM, which is the smallest interval reported to date. Moreover, the map constructed in the present study converged to eight linkage groups including all markers used in the mapping analysis, whereas the others did not. Together, these results show that the map constructed in the present study will be suitable for use as a basic linkage map in the future studies of common buckwheat.

The linkage map constructed in the present study had 756 independent loci and 8,884 markers (Table 1; Fig. 3). Thus, the map harbored multiple markers at a single position. This characteristic may contribute to the versatility of the linkage map. Common buckwheat has a high level of genetic variation; therefore, genetic composition differs by population. This high level of genetic variation makes it difficult to apply markers detected in one population to another population. However, if a single position has multiple markers, at least one of the markers may be polymorphic in a different target population. This allows the linkage-mapping step to be skipped for new target populations because the positions of the co-located polymorphic markers are already known.

QTL were detected for main stem length (Fig. 5; Table 2). The linkage map consisted of 756 loci, and marker density was high. QTL detected on the P1 map were not detected on the P2 map, suggesting that the QTL detected in one population (i.e., the loci heterozygous in one or both parents) may not be detected in another population. That is, QTL efficient in one population may not totally explain the genetic variation of the target trait in another population. This suggests that marker-assisted selection may not be a suitable method for use in common buckwheat populations. Because it is necessary to map QTL on a population-by-
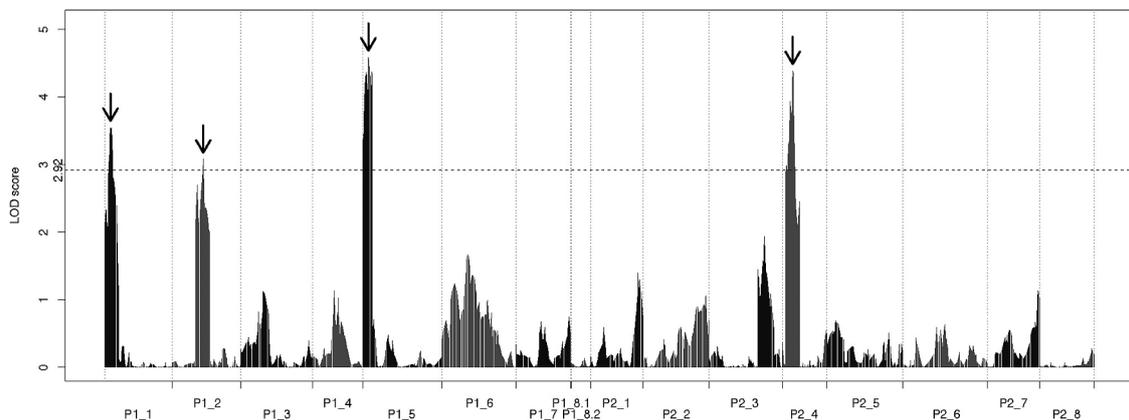

Fig. 5. Logarithm of the odds (LOD) score profile of the quantitative trait loci analysis of main stem length. Map positions are shown on the *x*-axis. The horizontal dashed line shows the 5% threshold of LOD score (2.92) obtained from a permutation test. Arrows show the detected peaks of LOD scores.

Table 2. Positions, logarithm of the odds (LOD) scores, additive effects and explained phenotypic variations of the quantitative trait loci detected for main stem length

Linkage group	Position (cM)	LOD	Additive effect	Explained phenotypic variance (%)
P1-1	9.3	3.55	−7.48	6.62
P1-2	49.0	3.08	−6.85	5.64
P1-5	9.0	4.59	−8.45	8.51
P2-4	16.9	4.39	8.29	8.18

population basis, genome-wide association studies (GWAS) may instead be better to detect QTL that contribute to the genetic variation of a target trait in multiple populations. When undertaking GWAS or multi-family (i.e., population-by-population) QTL mapping, a high-throughput genotyping system is essential. Therefore, the genotyping system proposed in the present study may be useful for GWAS and multi-family QTL mapping.

The results of the present study suggest that genomics-assisted breeding can be used in crop plants for which genomic resources are poor. We conducted linkage mapping and QTL mapping in a breeding population of common buckwheat without using any *a priori* genomic information or any existing genomic tools. It took eight months to construct the microarray markers in this study. Because we did not use any *a priori* genomic information in common buckwheat, our results show the timescale in which the genotyping system can be used for the analysis of other crops for which no genomic information is available. When a genotyping system is used for genomic selection (Meuwissen *et al.* 2001), the genotyping of plants needs to be conducted at every selection cycle. Yabe *et al.* (2014) conducted genomic selection simulations with open-pollinated plants, and the results suggest that it is important to control pollination among parents in allogamous species. To select parent plants before pollination, it is necessary to obtain their marker genotypes prior to flowering. In this study, it took 37 days (5 days for DNA extraction, 7 days for the Cy3 labeling, 17 days for microarray analysis, 6 days for data analysis) to process 178 samples from DNA extraction to genotyping, suggesting the genotyping system allows us to obtain genotypes of genome-wide markers of a few hundreds of common buckwheat plants before the plants mature.

The marker system used in this study was a microarray-based method. Diversity arrays technology (DArT) is a well-known microarray-based method for genotyping DNA polymorphisms, which can detect a single base pair change in a restriction site. Although DArT has potential to detect insertion, deletion, and rearrangement type DNA polymorphisms (Jaccoud *et al.* 2001), DArT mainly detected restriction site polymorphisms (Wittenberg *et al.* 2005). The method used in this study can detect insertion, deletion, and rearrangement type DNA polymorphisms as well as restriction site polymorphisms, because a mutation at a single to several nucleotides can be detected accurately by a probe

whose length is shorter than a genomic DNA fragment (Enoki *et al.* 2012). In the present study, we obtained a large number of markers (10,813) showing clear segregation patterns, suggesting that the system has high capability to detect DNA polymorphisms in a common buckwheat population.

Currently, genotyping systems using next generation sequencing, such as restriction site-associated DNA sequencing (RAD-Seq; Baird *et al.* 2008) and genotyping-by-sequencing (GBS; Elshire *et al.* 2011), are used in various crops to score a massive number of SNPs for a large number of genotypes. Genotyping using next generation sequencing has several advantages over genotyping using microarrays: (i) it does not require SNP discovery and array design (Davey *et al.* 2011), (ii) it is free from ascertainment bias attributable to SNP discovery (Heslot *et al.* 2013), and (iii) it is lower in cost (Elshire *et al.* 2011). Genotyping using next generation sequencing, however, shows higher rates of missing data than genotyping using microarrays in most cases, and requires a special procedure to estimate markers with missing data (e.g., Poland *et al.* 2012). The genotyping system used in this study showed a low rate (0.4%) of missing data. Low rate of missing data is necessary to estimate recombination fraction accurately in the linkage mapping. Iehisa *et al.* (2014) used the same genotyping system to construct a high-density linkage map of common wheat and reported low levels of missing data. The result also suggests a high yield of data from the genotyping system.

Here, we demonstrated the potential of a novel marker system for the linkage and QTL mapping of common buckwheat. Our results suggest that this system can be used as a rapid, high-throughput genotyping technology for genomics-assisted breeding, and that it will be particularly useful for use in crop plants for which no prior genomic information is available. Because of their high marker density, the linkage map constructed in this study will be useful for genomic studies based on other common buckwheat populations. By using the novel marker system, marker genotypes of a few hundreds of plants can be acquired in ca. 40 days, suggesting that the system can be used for genomic selection that requires marker genotypes to be obtained prior to flowering.

Acknowledgments

Literature Cited

Armstead, I., L. Huang, A. Ravagnani, P. Robson and H. Ougham (2009) Bioinformatics in the orphan crops. Brief. Bioinform. 10: 645–653.

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W.A. Cresko and E.A. Johnson (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3: e3376.

Baniya, B.K. (1990) Buckwheat in Nepal. Fagopyrum 10: 86–94.

Basten, C.J., B.S. Weir and Z.-B. Zeng (2003) QTL Cartographer, version 1.17. A reference manual and tutorial for QTL mapping. Department of Statistics, North Carolina State University, Raleigh, NC 187.

Campbell, C.G. (1997) Buckwheat, *Fagopyrum esculentum* Moench. Vol. 19. Bioversity International.

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen and M.L. Blaxter (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12: 499–510.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler and S.E. Mitchell (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6: e19379.

Enoki, H., S. Nishimura and A. Murakami (2012) Method for designing probe in DNA microarray, and DNA microarray provided with probe designed thereby. United States Patent Application Publication; US2012190582 (2012.7.26).

Grattapaglia, D. and R. Sederoff (1994) Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137: 1121–1137.

Hara, T., H. Iwata, K. Okuno, K. Matsui and R. Ohsawa (2011) QTL analysis of photoperiod sensitivity in common buckwheat by using markers for expressed sequence tags and photoperiod-sensitivity candidate genes. Breed. Sci. 61: 394–404.

Heslot, N., J. Rutkoski, J. Poland, J.L. Jannink and M.E. Sorrells (2013) Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. PLoS ONE 8: e74612.

Iehisa, J.C.M., R. Ohno, T. Kimura, H. Enoki, S. Nishimura, Y. Okamoto, S. Nasuda and S. Takumi (2014) A high-density genetic map with array-based markers facilitates structural and quantitative trait locus analyses of common wheat genome. DNA Res. 21: 555–567.

Iwata, H. and S. Ninomiya (2006) AntMap: constructing genetic linkage maps using an ant colony optimization algorithm. Breed. Sci. 56: 371–377.

Jaccoud, D., K. Peng, D. Feinstein and A. Kilian (2001) Diversity Arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res. 29: e25.

Joshi, B.D. (1999) Status of buckwheat in India. Fagopyrum 16: 7–11.

Konishi, T. and O. Ohnishi (2006) A linkage map for common buckwheat based on microsatellite and AFLP markers. Fagopyrum 23: 1–6.

Kosambi, D.D. (1943) The estimation of map distances from recombination values. Ann. Eugenics 12: 172–175.

Langmead, B., C. Trapnell, M. Pop and S.L. Salzberg (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Lewis, D. and D.A. Jones (1992) The genetics of heterostyly. *In*:

Barrett, S.C.H. (ed.) Evolution and function of heterostyly. Springer, Berlin. pp. 129–150.

Ly, D., M. Hamblin, I. Rabbi, G. Melaku, M. Bakare, H.G. Gauch Jr., R. Okechukwu, A.G.O. Dixon, P. Kulakow and J.L. Jannink (2013) Relatedness and genotype × environment interaction affect prediction accuracies in genomic selection: a study in cassava. Crop Sci. 53: 1312–1325.

Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829.

Pan, S.J. and Q.F. Chen (2010) Genetic mapping of common buckwheat using DNA, protein and morphological markers. Hereditas 147: 27–33.

Poland, J.A., P.J. Brown, M.E. Sorrells and J.L. Jannink (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7: e32253.

Ritter, E., C. Gebhardt and F. Salamini (1990) Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. Genetics 125: 645–654.

Sved, J.A. (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Pop. Biol. 2: 125–141.

Varshney, R.K., A. Graner and M.E. Sorrells (2005) Genomics-assisted breeding for crop improvement. Trends Plant Sci. 10: 621–630.

Varshney, R.K., W. Chen, Y. Li, A.K. Bharti, R.K. Saxena, J.A. Schlueter, M.T.A. Donoghue, S. Azam, G. Fan, A.M. Whaley *et al.* (2012) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat. Biotech. 30: 83–89.

Varshney, R.K., C. Song, R.K. Saxena, S. Azam, S. Yu, A.G. Sharpe, S. Cannon, J. Baek, B.D. Rosen, B. Tar'an *et al.* (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat. Biotech. 31: 240–246.

Wittenberg, A.H.J., T. van der Lee, C. Cayla, A. Kilian, R.G.F. Visser and H.J. Schouten (2005) Validation of the high-throughput marker technology DArT using the model plant Arabidopsis thaliana. Mol. Genet. Genomics 274: 30–39.

Yabe, S., R. Ohsawa and H. Iwata (2014) Genomic selection for the traits expressed after pollination in allogamous plants. Crop Sci. 54: 1448–1457.

Yasui, Y., Y. Wang, O. Ohnishi and C.G. Campbell (2004) Amplified fragment length polymorphism linkage analysis of common buckwheat (*Fagopyrum esculentum*) and its wild self-pollinated relative Fagopyrum homotropicum. Genome 47: 345–351.

Zeng, Z.-B. (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA 90: 10972–10976.

Zeng, Z.-B. (1994) Precision mapping of quantitative trait loci. Genetics 136: 1457–1468.

Zerbino, D.R. and E. Birney (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18: 821–829.