

De novo sequencing of seed transcriptome and development of genic-SSR markers in common buckwheat (*Fagopyrum esculentum*)

Taoxiong Shi · Ruiyuan Li · Qijiao Chen · Yue Li · Fan Pan · Qingfu Chen

Received: 3 January 2017 / Accepted: 18 October 2017
© Springer Science+Business Media B.V. 2017

Abstract Large-scale identification of simple sequence repeat (SSR) markers has not been reported in the genus *Fagopyrum* due to a lack of genomic resources. High-throughput transcriptomic sequencing can provide a rich resource for development of genic-SSR markers. In our study, the immature seeds transcriptome of common buckwheat (*Fagopyrum esculentum*) was sequenced using the Illumina HiSeq™ 2000 platform. We obtained a total of 21 million paired-end reads after filtering. The reads were assembled into 54,947 transcripts with an average length of 1235 bp. In total, 46,126 (83.95%) transcripts showed significant hits to known proteins or domains. Among these, 43,647 (79.43%) transcripts was assigned at least one gene ontology (GO) term. A total of 12,677 (23.07%) transcripts were mapped to the 141 Kyoto Encyclopedia of Genes and Genome (KEGG) pathways. Overall, 2326 transcripts contained at least one SSR with a frequency of one SSR per 1.17 kb. Mono-nucleotide repeats were the most abundant type (45.48%) followed by tri-nucleotide repeats

(35.17%). Of the 150 SSRs chosen randomly, 36 markers showed polymorphism among 24 common buckwheat accessions, and 141 genic-SSRs (94.0%) were transferable to tartary buckwheat. A low level of genetic diversity was found among 35 registered tartary buckwheat varieties in China assessed using 31 polymorphic SSR markers. Genic-SSR markers developed herein will be valuable for construction of linkage maps, genetic diversity studies, and marker-assisted selection in *Fagopyrum* species.

Keywords Buckwheat · Seed · Transcriptome · Genic-SSR · Genetic diversity

Introduction

Buckwheat (*Fagopyrum* Mill) belongs to the eudicot family Polygonaceae comprising 23 known species (Chen 2012). Among these, common buckwheat (*F. esculentum*, an outbreeding species) and tartary buckwheat (*F. tataricum*, a self-pollinating species) are two major species cultivated for human consumption (Ikeda et al. 2003). Common buckwheat is a traditional pseudo-cereal mainly grown in temperate regions of Asia, Europe, and North America, while tartary buckwheat is an important functional food widely cultivated in the cold mountainous regions of Asia and Europe (Chen 1999). As an important part of the human diet, common buckwheat seeds are a rich source of high biological-value proteins due to its balanced amino acids composition. Furthermore, common buckwheat

Taoxiong Shi and Ruiyuan Li contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11032-017-0743-4>) contains supplementary material, which is available to authorized users.

T. Shi · Q. Chen · Y. Li · F. Pan · Q. Chen (✉)
Research Center of Buckwheat Industry Technology, Guizhou Normal University, Guiyang 550001, China
e-mail: chenqf1966@126.com

R. Li
Key Laboratory of Information and Computing Science Guizhou Province, Guizhou Normal University, Guiyang 550001, China

seeds contain higher amounts of flavonoids, dietary fiber, vitamins (B₃, B₆, and K) and minerals (Zn, Cu, and Mn) than cereals (Przybylski and Gruczynska 2009).

Despite the high nutritional and nutraceutical value of common buckwheat, the seed yield is low due to its self-incompatibility. Molecular breeding of common buckwheat has been impeded due to the lack of genomic resources and tightly linked markers for agronomically important genes. Molecular markers play an important role in genetic studies and marker-assisted selection (MAS) in crop breeding. A variety of molecular markers such as amplified fragment length polymorphism (AFLP) (Yasui et al. 2004; Konishi et al. 2005; Konishi and Ohnishi 2006), random amplified polymorphic DNA (RAPD) (Sharma and Jana 2002; Pan and Chen 2010), micro-satellite or simple sequence repeat (SSR) (Iwata et al. 2005; Ma et al. 2009), expressed sequence tag-based single nucleotide polymorphism (EST-SNP) (Hara et al. 2011), and array-based single nucleotide polymorphism (SNP) markers (Yabe et al. 2014) have been developed and used for genetic diversity and relationship studies, linkage maps construction and quantitative trait locus (QTL) mapping in common buckwheat.

SSRs are the preferred markers for analysis of genetic diversity, population genetics, and MAS because of their high abundance, multi-allelic nature, high polymorphism, and rapid and readily detection. Until now, only a few SSR markers have been developed and reported for genetic studies in common buckwheat (Iwata et al. 2005; Konishi et al. 2006; Ma et al. 2009). Furthermore, the majority of the reported markers were developed from the anonymous genomic DNA sequences and have been difficult to associate with target agronomic traits. Recently, a large number of SSRs were identified and characterized from the genomic survey sequence of tartary buckwheat (Hou et al. 2016). However, previous studies of SSR transferability suggested that genomic SSRs showed a lower rate of transferability to related species or genera than SSR markers from expressed sequence tags (genic-SSRs or EST-SSRs) (Varshney et al. 2005; Somta et al. 2009; Souframanien and Reddy 2015), which is probably because genic-SSR markers, being part of the transcriptome, are highly conserved. Due to their association with coding sequences, genic-SSRs are more useful as genetic markers in comparative mapping and direct gene/QTL mapping of agronomically important traits and MAS.

RNA sequencing (RNA-seq) using next-generation sequencing (NGS) technologies, such as 454 pyrosequencing and Illumina, not only accelerates our understanding of the complexity of gene expression and regulation but also provides a wealthy resource for development of genic-SSR markers. In recent years, genic-SSRs have already been developed using NGS approaches for various minor crops such as legumes (Souframanien and Reddy 2015; Kaur et al. 2012; Zhang et al. 2012; Chen et al. 2015a, b), rubber tree (Li et al. 2012), yellow lupin (Parra-González et al. 2012), and primrose species (Zhang et al. 2013). Although high-throughput mRNA sequencing technologies had been used in detection of the candidate gene (S-ELF3) controlling the short-styled phenotype of common buckwheat (Yasui et al. 2012), differentially expressed genes in inflorescence transcriptomes between common and tartary buckwheat (Logacheva et al. 2011), and genes responsible for Al-tolerance in tartary buckwheat (Zhu et al. 2015), genic-SSR markers have rarely been reported in *Fagopyrum* species so far.

In this study, the immature seeds transcriptome of common buckwheat was deep sequenced using the Illumina HiSeq™ 2000 platform, and a great number of genic-SSRs was identified and characterized from the assembled transcripts. Genic-SSR markers were developed to test their utilization for the study of genetic diversity in both common and tartary buckwheat.

Materials and methods

Plant material

Immature seeds of common buckwheat cultivated variety (Datian-1) for transcriptome sequencing were procured from Research Center of Buckwheat Industry Technology, Guizhou Normal University (Guiyang, China). Twenty-four common buckwheat accessions collected from China and Japan (Electronic Supplementary Material [ESM] Table S1) and 35 registered tartary buckwheat varieties in China (ESM Table S2) were used for SSR polymorphism detection and genetic diversity analysis.

RNA isolation and sequencing

Approximately 100 immature seeds were collected and ground into fine power with liquid nitrogen. Total RNA

was extracted from immature seeds using the Quick RNA isolation Kit (Huayueyang Biotech Co., Ltd., Beijing, China) following the manufacturer's instructions and cleaned with RNase-free DNaseI (TaKaRa Biotech Co., Ltd., Dalian, China). The integrity and concentration of the RNA was determined using agarose gel electrophoresis (1.2%) as well as an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). High-quality RNA sampling (concentration ≥ 200 ng/ μ L; RIN ≥ 6.5 ; OD_{260/280} ≥ 1.8 and OD_{260/230} ≥ 1.8) was used to construct the complementary DNA (cDNA) libraries.

The cDNA library construction and sequencing reactions were conducted at Biomarker Technologies Corporation (Beijing, China). RNA was processed for the preparation of the paired-end cDNA sequencing library using the TruSeq RNA sample Prep Kit (Illumina, San Diego, CA, USA) as per the manufacturer's instruction. The cDNA was purified using a QiaQuick PCR extraction kit (Qiagen, Inc., Hilden, Germany). Short fragments were subjected to end repair and poly (A) addition, sequencing adapter connection, and agarose gel electrophoresis isolation. Next, 200 bp cDNA fragments were selected as templates for PCR amplification. The cDNA libraries were sequenced using the Illumina HiSeq™ 2000 platform using the paired-end model to obtain read lengths of 101 bp.

Sequence cleaning, assembly, and annotation

NGSQCToolkit v2.5 (Patel and Jain 2012) was used to screen adaptors, trim polymers on the end of reads and remove reads with more than 5 Ns or more than 30% of poor bases (quality score less than 20) (Garg et al. 2011). The screened reads were assembled with Trinity (Grabherr et al. 2011) using default parameter settings and transcripts shorter than 200 bp were discarded. The unigene was the longest transcript of the cluster extracted from the transcript names in the Trinity.fasta file (e.g., the transcript 'comp86566_c0_seq1' belonged to the cluster 'comp86566_c0').

To deduce the functions, the assembled transcripts were firstly searched using BLASTx against the National Center for Biotechnology Information's (NCBI) non-redundant (Nr) protein database with an E value of $< 10^{-5}$. Interproscan V5.0 was used to detect the domains of transcripts with default setting. The BLAST2GO v2.6.0 program was then used to obtain gene ontology (GO) (Ashburner et al. 2000) and Kyoto

Encyclopedia of Genes and Genomes (KEGGs) (Kanehisa et al. 2008) pathway annotation of the transcripts based on the NCBI Nr hits and Interproscan results. The classification and macro-distribution of the GO annotation was exported to an Excel® (Microsoft Corp., Redmond, WA, USA) spreadsheet to draw the chart.

Simple sequence repeat (SSR) identification and primer design

SSR motifs were detected from transcripts using MISA 1.0 (MICroSAtellite: <http://pgrc.ipk-gatersleben.de/misa/>) (Thiel et al. 2003). The minimum number of repeat units was set as follows: ten for mono-nucleotide, six for di-nucleotides, five for tri-nucleotides, and three for tetra-, penta-, and hexa-nucleotides. Adjacent microsatellites within 100-nt were counted as compound repeats. If a microsatellite was detected in multiple transcripts belonging to the same cluster, only the longest transcript was used in the subsequence analysis and the remains were dropped. One hundred and fifty SSRs were chosen randomly to synthesize (Sangon Biotech, Shanghai, China). Primer pairs were designed using Primer 3 (Untergasser et al. 2012) with the following parameters: primer lengths of 18–22 bp, GC content of 40–60%, annealing temperature of 55–60 °C, and target PCR product size of 100–300 bp.

Validation of SSR markers

The 150 pairs of SSR primers were tested in 24 common buckwheat accessions (ESM Table S1) and 35 registered tartary buckwheat varieties in China (ESM Table S2) to confirm their suitability and polymorphism. Total genomic DNA of each accession was extracted from young leaves of multiple individual plants using a CTAB methodology (Doyle and Doyle 1990). PCR reactions were carried out in 10 μ L volumes containing 20 ng of template genomic DNA, 1 $\times$ Taq Buffer, 1.5 mM MgCl₂, 200 mM deoxynucleotide triphosphates (dNTPs), 1.6 μ M of each primer, and 0.5 U Taq DNA polymerase (TAKARA Biotech, Dalian, China). The program of PCR amplification was set as follows: 94 ° for 3 min, followed by 35 cycles of 94 °C for 30 s, 55–62 °C (depending on optimized annealing temperature of primers) for 45 s, 72 °C for 45 s, and, finally, extension at 72 °C for 10 min. All PCR products were separated by 6% polyacrylamide gel electrophoresis

and subsequently stained with 0.15% AgNO₃ for 10 min. The PCR products were sized roughly based on the 100 bp DNA ladder (TIANGEN, Beijing, China).

Genetic diversity analysis

Descriptive genetic diversity was measured using PopGen software version 32 (<http://popgen-popgen.software.informer.com>) based on the genotyping data. The number of observed alleles (Na) per locus was calculated, and the effective number of alleles per locus (Ne) was estimated using the formula $Ne = 1/\sum Pi^2$, where Pi is the frequency of the i^{th} allele (Kimura and Crow 1964). The gene diversity/expected heterozygosity (H_e) was calculated as $H_e = 1 - \sum Pi^2$ (Nei 1973). Shannon's information index (I) was computed as $I = -\sum Pi^2 \ln Pi^2$ (Lewontin 1974). The polymorphism information content (PIC) of each SSR locus was calculated as $PIC = \sum_{i=1}^m p_i^2 - \sum_{i=1}^{m-1} \sum_{j=i+1}^m 2p_i p_j$, where p_i and p_j are the frequency of the i^{th} and j^{th} allele for the locus (Botstein et al. 1980). NTSYS-pc version 2.1 software (<http://www.exetersoftware.com/>) was used to calculate genetic distances and UPGMAM (Unweighted Pair Group Method with Arithmetic Mean) values in order to construct the dendrogram among accessions.

Availability of sequencing data

All raw sequencing data from this study have been deposited in the NCBI Sequence Read Archive (SRA; <http://www.ncbi.nlm.nih.gov/sra>). The BioProject and SRA accession numbers are PRJNA315634 and SRP072050, respectively.

Results

Illumina sequencing, assembly, and assessment

Using the Illumina HiSeq™ 2000 platform, we obtained a total of approximately 21 million high-quality reads. A total of 54,947 transcripts were assembled, with an average length of 840 bp and an N50 value of 1235 bp. The mean reads per transcript was 31.6 bp (ESM Table S3). The lengths of the transcripts ranged from 201 to 7231 bp. Of these transcripts, 23,732 (43.2%) were 201–500 bp; 14,399 (26.2%) were 501–1000 bp; 8302 (15.1%) were 1001–1500 bp; 4719

(8.6%) were 1501–2000 bp; 2190 (4.0%) were 2001–2500 bp, and 1901 (3.5%) were longer than 2500 bp (Fig. 1). The combined sequences of these transcripts were assembled into 36,133 unigenes (ESM Table S3).

The assembly quality of the transcripts was assessed using a BLASTn search against 28 full-length cDNA sequences of common buckwheat genes available in GenBank (ESM Table S4). Significant hits were found for all 28 sequences in our assembly, with average coverage of 87.18%. Of the 28 gene sequences, 18 were covered more than 95%, and 11 were covered completely. Logacheva et al. (2011) performed floral transcriptome sequencing in common and tartary buckwheat by 454 sequencing. The contigs from their results were also mapped onto these 28 gene sequences (Table S4). Compared to BLAST results from the floral contigs, a higher coverage was produced in our study, indicating the good quality and integrity of the assembled transcripts.

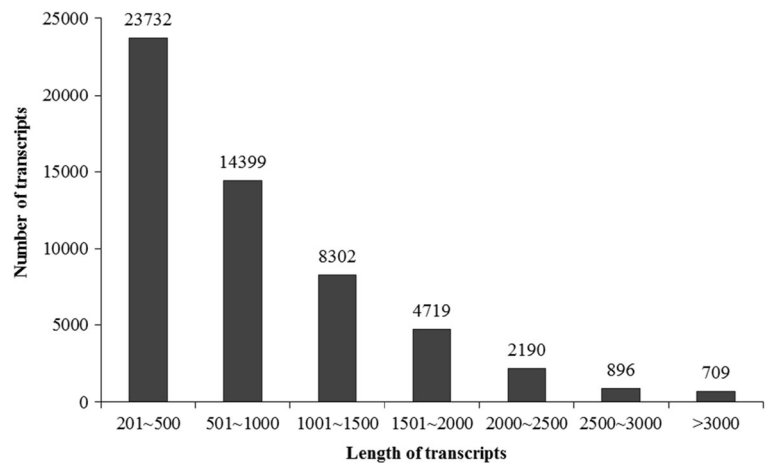
Annotation

For functional annotation, the assembled transcripts were first searched against the Nr protein database using BLASTx. The results indicated that out of 54,947 transcripts, 41,412 (75.37%) showed significant similarity to known proteins (ESM Table S3). In the domain analysis implemented in Interproscan, 16,246 (29.57%) transcripts have significant hits. These together resulted in a total of 46,126 (83.95%) annotated transcripts (ESM Table S3). The species that gave the most top-hits and strongest similarity were *Vitis vinifera* (12,715; 50%) followed by *Prunus persica* (4284; 7.80%), *Ricinus communis* (4114; 7.49%), and *Populus trichocarpa* (4055; 7.38%) (Fig. 2).

In total, 43,647 (79.43%) transcripts matching known proteins or domains were classified into the three main GO categories: biological process (28,038 [51.03%]), cellular component (29,863 [54.35%]), and molecular function (29,294 [53.31%]). There were 13, seven, and nine level-2 categories for the three main categories, respectively. The most highly represented terms in the level-2 categories for the three main categories were 'metabolic and cellular process', 'cell and organelle', and 'binding and catalytic activity', respectively (ESM Fig. S1).

To identify the biological pathways, the annotated transcripts were searched for in KEGG based on the

Fig. 1 Length distributions of the assembled transcripts for common buckwheat

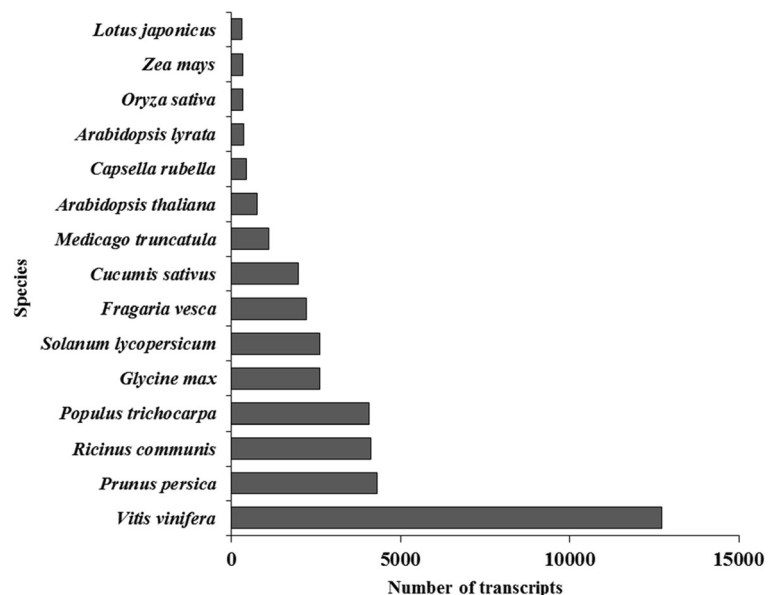


Nr and Interproscan hits using BLAST2GO. A total of 12,677 (23.07%) transcripts were mapped to the 141 pathways (ESM Table S5). The most highly represented category was ‘carbohydrate metabolism’ with 4043 transcripts involved. ‘Amino acid metabolism’ and ‘lipid metabolism’ were the two next best represented pathways, with 2252 and 1809 transcripts involved, respectively. Many transcripts were grouped into the biosynthesis and metabolism of flavonoids (162), alkaloids (201), and glucosinolate (13). In addition, a few transcripts were distributed to the pathways of ‘signal transduction’, ‘immune system’, ‘translation’, and ‘plant hormone biosynthesis’.

Frequency and distribution of SSRs in the common buckwheat transcriptome

Of the 54,947 transcripts assembled in the study, 2454 SSRs were detected in 2326 transcripts. Of these, 2208 transcripts (94.93%) contained an SSR, 117 (5.03%) contained two or three SSRs, one (comp57058_c1_seq4) had six SSRs and 74 (3.18%) had compound SSRs (ESM Table S6). The frequency of SSRs was 4.47%, and the mean distance was 1.17 kb in the transcripts that contained SSRs. Mono-nucleotide repeats were the most abundant type followed by tri-nucleotide repeats, making up about 45.48% (1116) and 35.17% (863) of the total,

Fig. 2 Top-hit species distribution for transcripts from common buckwheat by searching against the National Center for Biotechnology Information's non-redundant (NCBI-Nr) database



respectively; di-nucleotide repeats accounted for 15.16% (372) of the total SSRs. Tetra- and penta-nucleotide and compound repeats in total accounted for only 4.20% of the SSRs, and no hexa-nucleotide repeats were detected (Table 1).

Ten repeats were the most abundant for mono-motifs. For di-, tri-, tetra-, and penta-motifs, five repeats were the most abundant, followed by six and seven repeats. Among mono-nucleotide repeats, A/T (98.7%) was the most abundant, including 61 A/T located at the 3' end

Table 1 Summary of the number of simple sequence repeat (SSR) motifs in common buckwheat

Repeat motif	No. of repeats								Total	%
	5	6	7	8	9	10	11	≥ 12		
Mono-nucleotide (1116)										
A	0	0	0	0	0	499	232	370	1101	44.87
C	0	0	0	0	0	8	3	4	15	0.61
Di-nucleotide (372)										
AC	0	15	5	1	0	0	1	0	22	0.90
AG	0	125	44	28	6	3	4	1	211	8.60
AT	0	80	28	16	8	4	2	0	138	5.62
CG	0	0	1	0	0	0	0	0	1	0.04
Tri-nucleotide (863)										
AAC	57	23	7	1	0	0	0	0	88	3.59
AAG	178	55	35	2	0	0	0	0	270	11.00
AAT	24	8	3	2	0	0	0	0	37	1.51
ACC	76	25	14	2	0	0	0	0	117	4.77
ACG	8	7	0	0	0	0	0	0	15	0.61
ACT	13	4	3	0	0	0	0	0	20	0.81
AGC	46	7	4	1	0	0	0	0	58	2.36
AGG	57	21	3	0	0	0	0	0	81	3.30
ATC	97	36	12	2	0	0	0	0	147	5.99
CCG	26	3	1	0	0	0	0	0	30	1.22
Tetra-nucleotide	22	3	0	0	0	0	0	0	25	1.02
Penta-nucleotide	3	0	0	0	0	0	0	0	3	0.12
Compound SSRs									75	3.06
Total	607	412	160	55	14	514	242	375	2454	100.00
%	24.7	16.8	6.5	2.2	0.6	20.9	9.9	15.3	100.0	

and 68 located at the 5' end. AAG/CTT was the most frequent tri-motifs repeat, accounting for 31.28% (270) of total tri-nucleotide motifs. The AG/CT repeat was dominant among the di-motifs, accounting for 56.73% (211) of total di-nucleotide repeats. There were no prominent motifs observed for the tetra- and penta-nucleotide motifs (Table 1).

Evaluation of genic-SSR markers in common buckwheat

To test the amplification and polymorphism of the identified genic-SSRs, 150 SSRs (ESM Table S7) chosen randomly were synthesized and amplified in 24 common buckwheat accessions. Of the 150 genic-SSR primer pairs, 60 yielded clear products with an expected size (Fig. 3a), 85 showed clear non-specific products, and five did not amplify any fragment. Among the successful markers, 36 (24.83%) showed polymorphisms and 143 alleles were detected in total (Table 2). The N_a per SSR locus varied from two to eight, with a mean value of 3.97. Moreover, the N_e was equal to 2.582, ranging from 1.18 to 4.80. The Shannon index (I) ranged from 0.287 to 0.1.712, with an average of 1.028. The mean observed heterozygosity (H_o) and expected heterozygosity (H_e) was equal to 0.043 and 0.556, varying from 0.000 to 0.333 and 0.153 to 0.792, respectively. The PIC per SSR locus ranged from 0.141 to 0.761, with a mean value of 0.504 (Table 2).

Further genetic relationships among the 24 common buckwheat accessions were determined using UPMGA (Unweighted Pair Group Method with Arithmetic Mean) methods (Fig. 4). Most of the accessions were distinguished by the SSR markers and the genetic similarity coefficient between accessions varied from 0.61 to 0.90. In addition to Yuqiao2, other accessions were classified into two subgroups. One subgroup contained five accessions collected from southwestern China and one accession (Bendidalqiao) from northwestern China. The other subgroup included 12 accessions from

northwestern China, three accessions from southwestern China (Qujingtianqiao, Datian1, and Fengtian1), and two Japanese accessions. These results showed an association between the genetic relationship of the common buckwheat and its geographical origin.

Transferability of common buckwheat genic-SSR markers to tartary buckwheat

To assess the transferability of common buckwheat genic-SSR markers, amplification of the 150 genic-SSRs was also carried out in 35 registered tartary buckwheat varieties in China. The PCR results showed that 141 (94.0%) genic-SSR primer pairs produced amplified products, including 59 with expected size and 82 with non-specific products. Among the successful primer pairs, 31 (20.7%) were polymorphic (Fig. 3b), including three common primer pairs (Marker 2310, 2312 and 6805) observed in common buckwheat (Table 2). The mean values of N_a and N_e per SSR locus was 2.650 and 1.496, respectively. The I , H_o , H_e , and PIC values ranged from 0.130 to 1.364, 0.000 to 0.238, 0.056 to 0.626 and 0.054 to 0.594, with an average of 0.501, 0.033, 0.284 and 0.248 respectively (Table 2). For each common marker, the N_a , N_e , I , H_e , and PIC values in tartary buckwheat were lower than those in common buckwheat.

Discussion

Prior to our study, high-throughput mRNA sequencing technologies such as the 454 pyrosequencing and Illumina had been used in genetic studies in buckwheat. Comparative analysis of flowers and inflorescence transcriptomes between common and tartary buckwheat revealed differentially expressed genes involved in sugar biosynthesis and metabolism (Logacheva et al. 2011). In the present study, the immature seed transcriptome of common buckwheat was firstly analyzed using the

Fig. 3 Polymorphic bands revealed using genic-marker 9011 in 24 common buckwheat genotypes (a) and 2216 in 35 Chinese-approved tartary buckwheat cultivars (b)

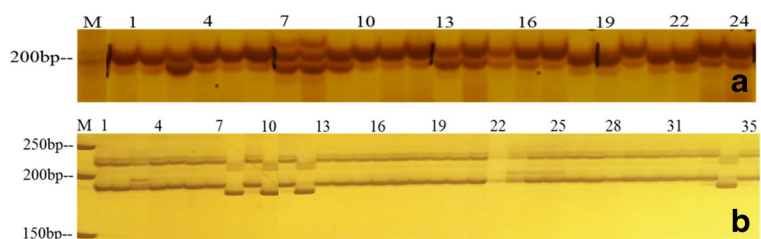


Table 2 Polymorphic primer pairs in common and tartary buckwheat

Marker	Repeat motifs	Primer sequence (F: 5'-3')	Reverse sequence (R: 5'-3')	PS	Na	Ne	<i>I</i>	<i>Ho</i>	<i>He</i>	PIC
Common buckwheat										
1722	(GAT)5	AAGCAAAGGCAAGTGGGC	TTGGGTTGGTCTTCAGCATC TA	295	5	4.364	1.540	0.000	0.771	0.735
2204	(AG)6	TGGCGCCTTTGCCT AGAATC	CCTTAACTCACCAG ACCCCA	144	3	1.646	0.675	0.000	0.392	0.341
2210	(TC)6	TCTAACATGTGCCC CTGGTG	CTCCTCAAGACGCG TGAAC	143	3	2.072	0.815	0.000	0.517	0.420
2212	(GA)6	GGGTCTGAGAACGA GATGCG	GACCCCCACCCAGACCAAC	172	4	2.420	1.072	0.000	0.587	0.535
2228	(TA)6	ATGGCGTAGGGCAT ACAGAG	TGCAGTTCAAGTCC ATAAACAAAGTC	185	3	2.341	0.968	0.000	0.573	0.510
2270	(AT)6	CGTGCAACTACCAC ACTTGC	ACGTTACTCGTGCA AGAAGC	178	4	3.024	1.211	0.042	0.669	0.611
2276	(TC)8	CCGGCACTACTCCATTTCCC	TTGTGACTTCTGATCCGCCA	165	3	2.462	0.990	0.000	0.594	0.523
2310*	(TTC)5	GGGAATCCTCCACC GTTGAT	TCCTGAAGCCGAGT CTCCTA	129	4	2.304	1.013	0.333	0.566	0.501
2312*	(CCA)5	ATGAGTGCACCACA TCCACC	ATTGCAGCATTAGG AGGCCG	131	4	3.556	1.321	0.000	0.719	0.667
5184	(AAG)5	AAGCATTTCTGAGTTTATCAG CTAC	ACTAGTGACCATTCATTTGA AACCT	194	4	2.246	1.041	0.042	0.555	0.510
5206	(CT)6	TGGTTAGCTTGTATGATGGT GG	AGGGGAAAAGAAGG AATGTCTC	272	6	4.448	1.602	0.292	0.775	0.740
6783	(A)10	ACTCCCGATACCT GTACCC	CTGAAGTGGCCGTC ACAGAT	147	6	4.664	1.659	0.125	0.786	0.756
6785	(T)11	GGTTCGGAGAACTG AGCCAA	AGTTCGGACAATCC AAGTGC	190	5	3.273	1.316	0.000	0.694	0.640
6805*	(CT)6	AAGACAGTGGTGGT AGTGGC	TAGAGGAAAGTAGT GCGGCG	136	4	2.286	0.965	0.000	0.563	0.477
6813	(AG)8	CGGCTGATCGGTTA GGGTTT	CCGTCGACGCCTTTGTTTTT	172	3	1.946	0.824	0.000	0.486	0.424
6835	(CA)6	TGTTGAAGCCCCAA TGGTCA	TGCGTTCTTGTGGCTACA	200	4	2.648	1.094	0.042	0.622	0.559
6841	(TC)7	CACCGCTCTCCAGA TCCATAC	CGTCATCTGCTCTGAATCCC A	195	8	2.790	1.459	0.045	0.642	0.618
6855	(AT)7	TGTACGTACGCTGG AGAAACT	CAAGAGCTGGCAAG GTTAGC	118	4	2.233	1.035	0.000	0.552	0.506
6861	(GGT)6	CCAGCGTTCTCAAG TTTACGC	CTGTGTGCGACTCTTCCCT	198	7	4.800	1.712	0.000	0.792	0.761

Table 2 (continued)

Marker	Repeat motifs	Primer sequence (F: 5'-3')	Reverse sequence (R: 5'-3')	PS	Na	Ne	<i>I</i>	<i>Ho</i>	<i>He</i>	PIC
6871	(ATC)5	GAGAGATGAGCACA GCGTCA	TCGATGGCGAAACC CAAGTT	277	3	1.534	0.639	0.042	0.348	0.316
7570	(T)11	ACAAGCAACTCCAC AAGTGC	GGCTTATGCACCCT ACCCAA	250	3	2.642	1.033	0.000	0.622	0.550
7586	(GGA)5	GCATCAACCTTGGT AGCCGA	CTCCGCCACTTTCTCTCGAT	201	5	3.105	1.261	0.167	0.678	0.620
7596	(TTC)5	GCTTGAAATGGCCT TGAGCC	TTCTCGTGGCTGAG AAGTCC	222	4	2.909	1.191	0.000	0.656	0.595
7614	(GTT)6	CTGAACTCGCCGTG CTATGA	CTAAGGTGGCTGCA GCAGAA	222	4	3.282	1.274	0.042	0.695	0.641
7618	(TCT)5	CGGGAGGCCAGTGT AAAAGT	TTCCTCCATTGAAG CACCCT	222	4	2.909	1.164	0.000	0.656	0.589
7628	(ACA)5	TCCAGATCGAGCAA CTGAGC	CCTGGCACTGCAAA ATGCTT	216	3	2.072	0.815	0.000	0.517	0.420
7648	(CGG)5	CCTTTAAGCCTCCG ATGTCGT	GGAATCGTGGTGGGA TTCGGA	121	3	1.185	0.345	0.000	0.156	0.150
7672	(CAT)6	TGCCTGGAACCTC TTCGTC	TGGAAGATGAGCGC AGGAAG	250	4	3.200	1.271	0.000	0.688	0.638
7682	(TGA)6	CCACGTTCCCACCA TACGAA	CGGCATTACCAAGA AACCCA	185	2	1.180	0.287	0.000	0.153	0.141
7684	(CAA)5	CAGATGGTACAAGG ACGGCA	ACTGCAACTGAGCT GATGACA	214	4	1.419	0.624	0.000	0.295	0.281
8995	(AT)7	CCAAACGGCTAGTC TCGGTC	ATGGTCCTTGGCTG AAGCTC	187	5	2.924	1.299	0.083	0.658	0.596
9011	(GA)6	GGCTCTTCTTTAGC ACCAGC	ACTGGACATGGTGA AATGGG	176	3	1.524	0.616	0.083	0.344	0.307
9033	(TA)6	AGGGTGGTGGTAGC CTTAGT	TGACATGTGCTCCA AACGCT	164	3	1.882	0.762	0.000	0.469	0.391
9069	(TCT)5	TTCGATACCTCATCGTCGGC	GCCGGCATCTAAAA TGGACG	129	2	1.704	0.604	0.167	0.413	0.328
9077	(GGA)6	GAAGGCCGATTACG CTTCCT	GTCTCCGTAGCAGC AAGAGG	205	4	2.713	1.122	0.000	0.631	0.563
9079	(TGA)5	TTCCATTTCGCCGCTTCCTG	GACGCAAGAAAATG GCCCAC	128	3	1.237	0.404	0.042	0.192	0.183
Tartary buckwheat										
2158	(A)11	GCTAGTGAGGCAGC TGAGAA	TCTGCACGGACTAG TGAAGC	209	2	1.192	0.298	0.000	0.161	0.148
2214	(AT)6			102	2	1.319	0.406	0.156	0.242	0.212

Table 2 (continued)

Marker	Repeat motifs	Primer sequence (F: 5'-3')	Reverse sequence (R: 5'-3')	PS	Na	Ne	<i>I</i>	<i>Ho</i>	<i>He</i>	PIC
		GTGAGTCGGAACCC GTAGCA	ACCCTATCGCAGGA AAGCAA							
2216	(TA) ₆	TTTTCCACCTCAGA GGGCAG	TTGCACATTGTCGTTTCCCA	180	3	1.533	0.649	0.000	0.348	0.320
2218	(CA) ₉	TATCCGGCTAACCCTACCA	ACACCTTCACCCTTGAATCT CG	175	2	1.486	0.508	0.000	0.327	0.274
2234	(AT) ₇	CTTGCCCAAGAGCCA AGGTAT	AGCAAAAACCTATGCTTTTAC TGC	168	5	2.676	1.216	0.121	0.626	0.579
2252	(GT) ₆	CACGCCCTCCACTTTTTCTT	TGCAAGCGATAGTA GCGATCC	117	2	1.346	0.425	0.000	0.257	0.224
2298	(TGT) ₅	CCCAGTTAAGGCCG GAATGA	TAAGGCTGGTGGTG GGGATA	175	3	1.270	0.428	0.000	0.213	0.199
2310*	(TTC) ₅	GGGAATCCTCCACC GTTGAT	TCCTGAAGCCGAGT CTCCTA	129	2	1.280	0.377	0.000	0.219	0.195
2312*	(CCA) ₅	ATGAGTGCACCACA TCCACC	ATTGCAGCATTAGG AGGCGG	131	2	1.147	0.251	0.000	0.128	0.120
5176	(AAG) ₅	GCTTTGGAGAAGGG ACTTTTGC	GATCGAAGGACCAA GCCCTTA	200	2	1.186	0.293	0.000	0.157	0.144
6805*	(CT) ₆	AAGACAGTGGTGGT AGTGGC	TAGAGGAAAGTAGT GCGGCG	136	2	1.841	0.649	0.000	0.457	0.352
6819	(AT) ₆	CAGTTTGAAGGGCA GCAGGA	TCTCCACCAGCAAC AACGTA	190	5	1.615	0.795	0.000	0.381	0.357
6821	(GA) ₆	ACCTGCAAAGCAAC CTAGTGA	GCCCCAATCGAAGT TTCACC	200	2	1.061	0.133	0.000	0.057	0.055
6827	(CT) ₆	AAACCACGAGCTTC CTCTCC	AGAATTTGCGGGAA GGGTGA	138	3	1.882	0.741	0.000	0.469	0.382
6843	(GA) ₆	TGAGTAATTGAAAA TGGAGATCGAC	TACTCCCAACCAGC AATCCG	194	3	1.331	0.481	0.156	0.249	0.229
6853	(A) ₁₀	GGCCTCCTTGAACA GCTAGA	GCAGTGACCGAAGT GCAGAT	109	7	2.624	1.364	0.088	0.619	0.594
6873	(AGA) ₅	TTGATGGACACTAG GCAAGGT	GTGAGGAGCAATTG GCGTTT	170	2	1.480	0.505	0.111	0.324	0.272
6875	(CAT) ₅	GAATCCACCATCAC CAGGCA	GTTTTCGATGCCAA CTGCGA	141	2	1.147	0.251	0.000	0.128	0.120
6889	(ATG) ₅	TGAGGGGCCAATGC ATGAGA	GGAGGCTTCTTAGTTCTTTT CCC	142	2	1.095	0.185	0.030	0.087	0.083
7582	(TCA) ₅	ATCCGAAACCGCCT CCTTAC	GGGGTTTTGGTGCA GGTACT	234	2	1.220	0.325	0.000	0.180	0.164
				232	4	2.150	0.947	0.000	0.535	0.467

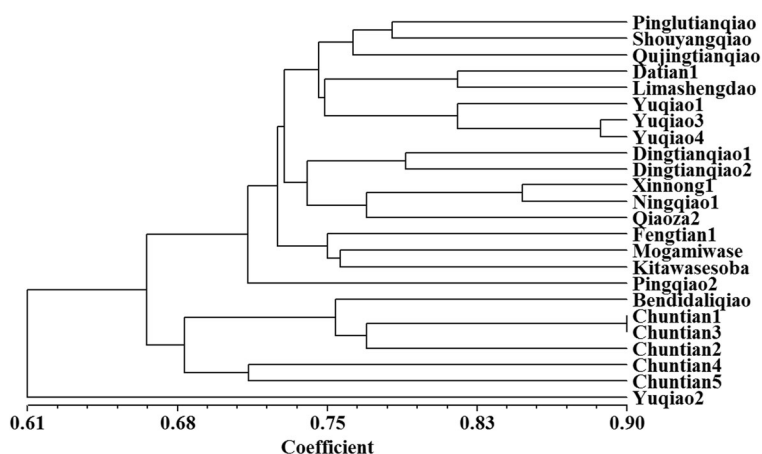
Table 2 (continued)

Marker	Repeat motifs	Primer sequence (F: 5'-3')	Reverse sequence (R: 5'-3')	PS	Na	Ne	<i>I</i>	<i>Ho</i>	<i>He</i>	PIC
7606	(GCA) ⁵	AGTTCGGCGGGAAG ATGAAT	AGTAGTCCCGGAGC TCACAT							
7622	(GTT) ⁵	GGACTGGCCAGCTG ATTCT	GTGAGGGCTGGTGT TGTAAGA	218	2	1.113	0.209	0.036	0.101	0.096
7662	(TCT) ⁵	AATGGCGGCAAATC TCGCTA	TGAGGATTTCAGGG GCTTGG	191	3	1.934	0.760	0.000	0.483	0.392
7678	(TGG) ⁵	GCCAACATGCAAAT GACCCA	CGGTGGATACGGGT AGGTTG	249	2	1.061	0.133	0.000	0.057	0.055
8963	(T) ¹⁰	GAGGATTGGCTGTC TGCTCT	TCAGCTCAATGCAA ACCTCC	122	2	1.059	0.130	0.000	0.056	0.054
8983	(A) ¹¹	TAGAAGATTGTGG CCGCCT	GGGAAATCACAATC ATAGTAAAGCA	150	2	1.849	0.652	0.238	0.459	0.354
9007	(AT) ⁶	AGTTCACGTGAGAT CCTAGTGT	CAACAGGGGAGAAT GACGCT	103	2	1.637	0.578	0.000	0.389	0.314
9013	(A) ¹⁴	TGTAGCTGCAGTTG AAGACG	CTTCCCGGTTTCTCTGGTGA	192	4	1.796	0.801	0.065	0.443	0.390
9023	(CT) ⁸	TCGGGGCCACAAGT CAATAC	TCTGGATAAGGGTT GGGGGT	136	2	1.062	0.136	0.000	0.059	0.057
9045	(AG) ⁶	ACGGTTTGCCTTTG GATTGC	CACCGACAACAACC GTTTCC	274	2	1.153	0.257	0.029	0.133	0.124
9095	(TGC) ⁵	AGCTGCAGCCATGT CCATAA	CATCATGAGACTCG GGACGG	175	2	1.841	0.649	0.000	0.457	0.352

Asterisks indicate the common primers between common and tartary buckwheat

He gene diversity/expected heterozygosity, *Ho* observed heterozygosity, *I* Shannon's information index, Na number of observed alleles, Ne effective number of alleles per locus, PIC polymorphism information content, PS expected product size (bp)

Fig. 4 Clustering analysis of 24 common buckwheat accessions based on a dataset of 36 genic-simple sequence repeat (SSR) markers



Illumina deep-sequencing platform. In comparison with the results of Logacheva et al. (2011), we produced a larger number of reads (20,508,824 vs. 266,782), more assembled transcripts (54,947 vs. 25,435), and a longer average transcript length (840 vs. 532) (ESM Table S3). The high N50 value of 1235 bp along with a large percentage of transcripts greater than 500 bp (Fig. 1) indicated that we generated a high-quality assembly of the immature seed transcriptome for common buckwheat. Moreover, the integrity of the assembled transcripts was better than the contigs from the floral transcriptome (ESM Table S4). The high quality and integrity of the assembled transcripts in our study can be ascribed to the fact that Illumina yielded longer and more accurate contigs despite the substantially shorter read length of 454 (Li et al. 2014). The assembled transcripts in this study could be a particularly useful source for future genetic research on buckwheat.

In our study, a total of 46,126 (83.95%) transcripts were functionally annotated after comparison against the NCBI's Nr and Interproscan databases. The proportion of annotated transcripts in our study was two-fold higher than the floral transcriptome (Logacheva et al. 2011) (ESM Table S3). GO and KEGG enrichment analysis revealed that 22 level-2 categories (75.9%) and 64 pathways (42.7%) showed significant difference in the transcript frequency between the two transcript sets (ESM Table S5). Furthermore, transcripts categorized into 'extracellular matrix', 'enzyme regulator activity', 'receptor activity', 'retinol metabolism', 'tetracycline biosynthesis', 'Fluorobenzoate degradation', 'penicillin and cephalosporin biosynthesis', and 'xylene degradation' were only presented in the seed transcriptome (ESM Table S5 and Fig. S1). The newly annotated transcript, pathway, and GO enables the search for candidate genes responsible for seed development.

The Buckwheat Genome DataBase was recently constructed and the draft genome of common buckwheat was assembled, which has been used as a reference sequence for the rapid identification of homolog genes identified in other plants, i.e., genes controlling flavonoid biosynthesis and genes encoding 2S albumin-type allergens and granule-bound starch synthases (Yasui et al. 2016). In our study, the top-hits taxonomic distribution of BLAST hits were *Vitis vinifera*, followed by *Prunus persica*, *Ricinus communis*, and *Populus trichocarpa* (Fig. 2). A similar taxonomic distribution of BLAST hits has also been found in transcriptome of

buckwheat floral (Logacheva et al. 2011), primrose species (Zhang et al. 2013), and an asterid, *Crataegostigma plantagineum* (Rodriguez et al. 2010), which indicated that buckwheat is close to most of the described Rosidae species, such as *Vitis vinifera*, *Prunus persica*, *Ricinus communis*, and *Glycine max*. The genome of *V. vinifera* (Jaillon et al. 2007), *P. trichocarpa* (Tuskan et al. 2006), *P. persica* (Ahmad et al. 2011), and *R. communis* (Chan et al. 2010) has been completely sequenced and released. The annotated and available whole-genome sequences of the four species will greatly facilitate the genomic alignment and allow candidate gene prediction in the genus *Fagopyrum*.

In our study, 2454 genic-SSRs were identified from 2326 transcripts, which will provide rich information for development of the genic-SSR marker in buckwheat. The mono-nucleotide (45.48%) was the most abundant SSR motifs followed by tri-nucleotide (35.17%) and di-nucleotide (15.16%) repeats (Table 1). Excluding mono-nucleotide repeats, our results were consistent with genic-SSR identification in black gram (Souframanien and Reddy 2015) and most cereal crops (Varshney et al. 2002), but differed from sesame (Zhang et al. 2012) and rubber tree (Li et al. 2012) and genomic-SSR in tartary buckwheat (Hou et al. 2016), in which di-nucleotides were the most abundant repeats followed by tri-nucleotide repeats. To our knowledge, only a few reports on genic-SSR or genomic-SSR marker identification in plants drew attention to mono-nucleotide repeats as it was difficult to distinguish genuine mononucleotide repeats from those generated by polyadenylation products, base mismatches, or sequencing errors. However, several studies have established that mono-nucleotide SSRs had high amplification and polymorphism rates followed by those of tetra-nucleotide SSRs (Varshney et al. 2002; Chen et al. 2015a, b). In our study, eight mono-nucleotide SSRs showed good amplification and high polymorphism, and among them four were very informative, with PIC values higher than 0.5. Therefore, more attention should be given to this type of repeat in future genic-SSR development and evaluations.

In tartary buckwheat genome, Hou et al. (2016) found a higher frequency of AT/AT (78.6%) for di-nucleotide and AAT/ATT (31.83%) for tri-nucleotide repeats. However, in our study AG/CT (56.72%) were the most abundant di-nucleotide repeats followed by the AT/AT motif (37.1%), and AAG/CTT (35.17%) were the most frequent tri-nucleotide repeats followed by ATC/GAT (17.03%) (Table 1), which was consistent

with the finding of genic-SSRs in most dicot plants, but differed from cereals, in which GGC/CCG was the most abundant tri-nucleotide repeat motif due to their increased GC content in the coding regions (Varshney et al. 2002). The differences in distribution frequency of SSR motif types mainly resulted from differences in the SSR search criteria, dataset, searching algorithm used, as well as different selection pressures on each plant species.

In the study of the 150 genic-SSRs randomly chosen, 36 showed polymorphism among 24 common buckwheat accessions. The polymorphic ratio (24.6%) of primer pairs was lower than the 26.7% reported by Konishi et al. (2006) but higher than the 7.4% reported by Ma et al. (2009) for genomic-SSR development in common buckwheat. The PIC of individual genic-SSR loci ranged from 0.141 to 0.761, with an average of 0.504. Of the 36 polymorphic markers, 23 (63.9%) were very informative, having PIC values higher than 0.5; these can be used as core primers for genetic diversity analysis and variety identification for buckwheat. The SSR genetic diversity observed in our study can be compared with the diversities detected in common buckwheat using genomic-SSRs. The average PIC value was lower than the 0.79 reported by Konishi et al. (2006) using a worldwide collection of common buckwheat accessions. The genetic diversity ($H_e = 0.556$) in this study was lower than that in 174 landraces common buckwheat from the Alps (0.762) used by Barcaccia et al. (2016), but proved to be comparable to that in 179 Korean common buckwheat accessions (0.53) used by Song et al. (2011) and 41 common buckwheat accessions (0.412), mostly collected from Korea, reported by Ma et al. (2009), in which the PICs were also comparable to our results. The low detection of heterozygosity and polymorphism in our study might not reveal the full genetic variability of common buckwheat germplasm as genetic diversity is highly dependent on the number of samples and on the geographic regions the samples came from. As a general rule, varieties sampled from a local geographic region would be inherently less diverse than germplasm from a wider range of geographic and ecologically distinct regions.

The UPGMA analysis clearly distributed the common buckwheat accessions into separate clusters, and most of the accessions were grouped according to

geographic distribution (Fig. 4). An association between genetic relationship and geographical origin was also reported in buckwheat using genomic-SSR (Barcaccia et al. 2016; Hou et al. 2016).

Cross-species transferability of genic-SSRs has been reported in many crops. Ma et al. (2009) reported 136 genomic-SSR markers in common buckwheat, and 61.0% were found to be transferable to tartary buckwheat. In our study, about 94.0% of genic-SSR markers were found to be transferable to tartary buckwheat, and 31 (20.7%) of these were polymorphic. The results are in accord with previous studies indicating that genic-SSR markers have a higher transfer rate than genomic-SSR markers. The polymorphic ratio (20.7%) was higher than the 10.4% reported on genomic-SSRs development in tartary buckwheat by Hou et al. (2016). The N_a , N_e , Shannon's information index (I), H_e , and PIC values of all polymorphic markers were lower in tartary buckwheat than in common buckwheat (Table 2). These results were consistent with the expectation that self-pollinated species possessed lower genetic diversity than outcrossing species, mainly due to their inherent lower levels of polymorphism and higher conservation in the coding regions. The H_e values ranged from 0.056 to 0.626 and the I score ranged from 0.130 to 1.364 in the 35 registered tartary buckwheat varieties in China, which was comparable to that in 64 tartary buckwheat accessions collected worldwide as reported by Hou et al. (2016), using genomic-SSRs (i.e., H_e values of 0.030713 to 0.747487, I values of 0.0795 to 1.3828). This indicating that the new SSR sequences and genic-SSR markers developed in this study will be useful resources for genetic diversity and genetic mapping studies in *Fagopyrum* species.

Conclusions

In this study, the immature seed transcriptome of common buckwheat was firstly analyzed using the Illumina deep sequencing platform. A total of 54,947 transcripts were de novo assembled and 83.95% transcripts were functionally annotated. The high quality and integrity of the assembled transcripts can be a particularly useful source for future genetic research on *Fagopyrum* species. In total, 2454 genic-SSRs were detected from 2326 transcripts. Randomly selected primers showed high polymorphism in common buckwheat and high transferability in tartary buckwheat. As the first report on

large-scale exploitation of genic-SSRs in buckwheat, the developed genic-SSRs will be valuable for construction of linkage maps, genetic diversity studies, and MAS in *Fagopyrum* species.

References

- Ahmad R, Parfitt DE, Fass J, Ogundiwin E, Dhingra A, Gradziel TM et al (2011) Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. *BMC Genomics* 12:569
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet* 25:25–29
- Barcaccia G, Volpato M, Gentili R, Abeli T, Galla G, Orsenigo S et al (2016) Genetic identity of common buckwheat (*Fagopyrum esculentum* Moench) landraces locally cultivated in the alps. *Rossi Genet Resour Crop Evol* 63:639–651
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331
- Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D et al (2010) Draft genome sequence of the oilseed species *Ricinus communis*. *Nat Biotech* 28:951–956
- Chen QF (1999) A study of resources of *Fagopyrum* (Polygonaceae) native to China. *Bot J Linn Soc* 130:54–65
- Chen QF (2012) Plant sciences on genus *Fagopyrum*, 1st edn. Science Press, Beijing (in Chinese)
- Chen HL, Liu LP, Wang LX, Wang SH, Somta P, Cheng XZ (2015a) Development and validation of EST-SSR markers from the transcriptome of adzuki bean (*Vigna angularis*). *PLoS One* 10(7):e0131939
- Chen HL, Wang LX, Wang SH, Liu CJ, Blair MW, Cheng XZ (2015b) Transcriptome sequencing of mung bean (*Vigna radiata* L.) genes and the identification of EST-SSR markers. *PLoS One* 10(4):e0120273
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Garg R, Patel RK, Tyagi AK, Jain M (2011) De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. *DNA Res* 18:53–63
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652
- Hara T, Iwata H, Okuno K, Matsui K, Ohsawa R (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
- Hou S, Sun Z, Linghu B, Xu D, Wu B, Zhang B et al (2016) Genetic diversity of buckwheat cultivars (*Fagopyrum tartaricum*, Gaertn.) assessed with SSR markers developed from genome survey sequences. *Plant Mol Biol Rep* 34(1):233–241
- Ikeda K, Asami Y, Lin R, Arai R, Hongda Y, Suzuki T et al (2003) Comparison of mechanical and chemical characteristics between common and tartary buckwheat. *Fagopyrum* 20:53–58
- Iwata H, Imon K, Tsumura Y, Ohsawa R (2005) Genetic diversity among Japanese indigenous common buckwheat (*Fagopyrum esculentum*) cultivars as determined from amplified fragment length polymorphism and simple sequence repeat markers and quantitative agronomic traits. *Genome* 48(3):367–377
- Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463–467
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M et al (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36:D480–D484
- Kaur S, Pembleton LW, Cogan NO, Savin KW, Leonforte T, Paull J et al (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics* 13:104
- Kimura M, Crow J (1964) The number of alleles that can be maintained in a finite population. *Genetics* 49:725–738
- Konishi T, Ohnishi O (2006) A linkage map of common buckwheat based on microsatellite and AFLP markers. *Fagopyrum* 23:1–6
- Konishi T, Yasui Y, Ohnishi O (2005) Original birthplace of cultivated common buckwheat inferred from genetic relationships among cultivated populations and natural populations of wild common buckwheat revealed by AFLP analysis. *Genes Genet Syst* 80:113–119
- Konishi T, Iwata H, Yashiro K, Tsumura Y, Ohsawa R, Yasui Y, Ohnishi O (2006) Development and characterization of microsatellite markers for common buckwheat. *Breed Sci* 56:277–285
- Lewontin RC (1974) The genetic basis of evolutionary change. Columbia University Press, New York
- Li DJ, Deng Z, Qin B, Liu XH, Men ZH (2012) De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg). *BMC Genomics* 13:192
- Li JJ, Chapman B, Charlebois P, Hofmann O, Weiner B, Porter AJ et al (2014) Comparison of 454 and Illumina deep sequencing in participants failing raltegravir-based antiretroviral therapy. *PLoS One* 9(3):e90485
- Logacheva MD, Kasianov AS, Vinogradov DV, Samigullin TH, Gelfand MS, Makeev VJ et al (2011) De novo sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). *BMC Genomics* 12:30–46
- Ma KH, Kim NS, Lee GA, Lee SY, Lee JK, Yi JY et al (2009) Development of SSR markers for studies of diversity in the genus *Fagopyrum*. *Theor Appl Genet* 119(7):1247–1254
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci U S A* 70:3321–3323
- Pan SJ, Chen QF (2010) Genetic mapping of common buckwheat using DNA, protein and morphological markers. *Hereditas* 147:27–33
- Parra-González LB, Aravena-Abarzúa GA, Navarro-Navarro CS, Udall J, Maughan J, Peterson LM et al (2012) Yellow lupin (*Lupinus luteus* L.) transcriptome sequencing: molecular marker development and comparative studies. *BMC Genomics* 13:425

- Patel RK, Jain M (2012) NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7: e30619
- Przybylski R, Gruczynska E (2009) A review of nutritional and nutraceutical components of buckwheat. *Eur J Plant Sci Biotechnol* 3:10–22
- Rodriguez MC, Edsgard D, Sarfraz HS, Alquezar D, Rasmussen M, Gilbert T et al (2010) Transcriptomes of the desiccation tolerant resurrection plant *Craterostigma plantagineum*. *Plant J* 63:212–228
- Sharma R, Jana S (2002) Species relationships in *Fagopyrum* revealed by PCR based DNA fingerprinting. *Theor Appl Genet* 105:306–312
- Somta P, Seehalak W, Srinives P (2009) Development, characterization and cross-species amplification of mungbean (*Vigna radiata*) genic microsatellite markers. *Conserv Genet* 10: 1939–1943
- Song JY, Lee GA, Yoon MS, Ma HK, Chio YM, Lee JR et al (2011) Analysis of genetic diversity and population structure of buckwheat (*Fagopyrum esculentum* Moench) landraces of Korea using SSR markers. *Korean J Plant Res* 24(6):702–711
- Souframanien J, Reddy KS (2015) De novo assembly, characterization of immature seed transcriptome and development of genic-SSR markers in black gram [*Vigna mungo* (L.) Hepper]. *PLoS One* 10(6):e0128748
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.) *Theor Appl Genet* 106:411–422
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC et al (2012) Primer3-new capabilities and interfaces. *Nucleic Acids Res* 14(15):e115
- Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002) In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett* 7:537–546
- Varshney RK, Sigmund R, Börner A, Korzum V, Stein N, Sorrells ME et al (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Sci* 168:195–202
- Yabe S, Ohsawa RS, Iwata H (2014) Genomic selection for the traits expressed after pollination in allogamous plants. *Crop Sci* 54:1448–1457
- Yasui Y, Wang Y, Ohnishi O, Campbell CG (2004) Amplified fragment length polymorphism linkage analysis of common buckwheat (*Fagopyrum esculentum*) and its wild self-pollinated relative *Fagopyrum homotropicum*. *Genome* 47: 345–351
- Yasui Y, Mori M, Aii J, Abe T, Matsumoto D, Sato S et al (2012) S-LOCUS EARLY FLOWERING 3 is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. *PLoS One* 7(2):e31264
- Yasui Y, Hirakawa H, Ueno M, Matsui K, Katsube-Tanaka T, Yang SJ et al (2016) Assembly of the draft genome of buckwheat and its applications in identifying agronomically useful genes. *DNA Res* 23(3):215–224
- Zhang HY, Wei LB, Miao HM, Zhang TD, Wang CY (2012) Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics* 13:316
- Zhang L, Yan HF, Wu W, Yu H, Ge HX (2013) Comparative transcriptome analysis and marker development of two closely related primrose species (*Primula poissonii* and *Primula wilsonii*). *BMC Genomics* 14:329
- Zhu HF, Wang H, Zhu YF, Zou JW, Zhao FJ, Huang CF (2015) Genome-wide transcriptomic and phylogenetic analyses reveal distinct aluminum-tolerance mechanisms in the aluminum-accumulating species buckwheat (*Fagopyrum tataricum*). *BMC Plant Biol* 15:16

Table S1 The common buckwheat accessions used in this study

Code	Accession	Origin	Code	Accession	Origin
1	Pinglutianqiao	Pinglu Shanxi China	13	Pingqiao2	Pingliang Gansu China
2	Shouyangqiao	Shouyan Shanxi China	14	Yuqiao1	Yulin Shaanxi China
3	Qujingtianqiao	Qujing Yunnan China	15	Yuqiao2	Yulin Shaanxi China
4	Datian1	Guiyang Guizhou China	16	Yuqiao3	Yulin Shaanxi China
5	Fengtian1	Guiyang Guizhou China	17	Yuqiao4	Yulin Shaanxi China
6	Dingtianqiao1	Dingxi Gansu China	18	Chuntian1	Guiyang Guizhou China
7	Dingtianqiao2	Dingxi Gansu China	19	Chuntian2	Guiyang Guizhou China
8	Xinnong1	Guyuan Ningxia China	20	Chuntian3	Guiyang Guizhou China
9	Bendidaliquiao	Tongliao Inner Mongolia China	21	Chuntian4	Guiyang Guizhou China
10	Ningqiao1	Guyuan Ningxia China	22	Chuntian5	Guiyang Guizhou China
11	Qiaoza2	Yulin Shaanxi China	23	Mogamiwase	Japan
12	Limashengdao	Inner Mongolia China	24	Kitawasesoba	Japan

Table S2 Registered tartary buckwheat varieties used in this study

Code	Variety	Registered time	Code	Variety	Registered time
1	Chuanqiao2	2012	19	Liuku2	2006
2	Chuanqiao3	2010	20	Liuku3	2011
3	Chuanqiao1	2000	21	Yunqiao2	2012
4	Chuanqiao4	2009	22	Rice buckwheat 1	2009
5	Chuanqiao5	2009	23	Ningqiao2	2005
6	Diku1	2010	24	Qiankuqiao2	2004
7	Zhaoqiao1	2008	25	Qiankuqiao3	2008
8	Fenghuangkuqiao	2001	26	Qiankuqiao4	2004
9	Ganqiao1	1990	27	Qiankuqiao5	2010
10	Zhaoqiao2	2010	28	Qiankuqiao6	2011
11	Heifeng1	1999	29	Tangwankuqiao	2001
12	Jinqiao2	2010	30	Qianheiqiao1	2002
13	Jinqiao4	2007	31	Xinong9920	2004
14	Jinqiao5	2011	32	Xinong9909	2008
15	Jinqiao6	2011	33	Xiqiao1	2000
16	Jiujiangkuqiao	2000	34	Xiqiao2	2008
17	Yu6-21	1996	35	Xiqiao3	2008
18	Yunqiao1	2010			

**Table S3 Assembly and annotation results in seed transcriptome and comparison with the
floral transcriptome of common buckwheat**

		<i>F. esculentum</i> Seed	<i>F. esculentum</i> Inflorescence (Logacheva et al. 2011)
Raw reads	Length of reads	101 X 101	42 - 3,607
	Total number of reads	21186802	-
	Total size (bp)	2,139,867,002	-
	GC content (%)	42.68	-
Clean reads	Total number of clean reads	20,508,824	266,782
	Total size (bp)	1,889,111,004	93,106,918
	GC content (%)	43.59	42.09
Assembly	Total number of transcripts	54,947	82,199
	Total length of transcripts (bp)	46,149,736	43,732,664
	Size of transcripts (bp)	201-7,231	42-3,607
	Mean length of transcripts (bp)	840	532
	Media length of transcripts (bp)	599	490
	N50 value of transcripts (bp)	1,235	533
	Number of reads mapped on transcripts	15,586,706	266,782
	Total bases mapped on transcripts	1,460,282,806	93,106,918
	Mean reads per transcript	31.6	3.25
	Number of unigenes	36,133	-
Annotation	Transcripts annotated by Nr	41,412 (75.37%)	33,382 (40.61%)
	Transcripts annotated by InterPro	16,246 (29.57%)	-
	Transcripts annotated by Gene Ontology	43,647 (79.43%)	25, 418 (30.92%)
	Transcripts annotated by KEGG	12,677 (23.07%)	9,382 (11.41%)
	Total Annotated	46,126 (83.95%)	33,382 (40.61%)

Table S4 Coverage of previously reported common buckwheat genes with full-length coding sequences in the seed and floral transcriptome assembly

Accession number	Putative function of gene product	Coverage (%)	
		Seed transcriptome	Floral transcriptome (Logacheva et al. 2011)
GU169469	dihydroflavonol-4-reductase	100	100
AY826351	aspartic proteinase 9	96	64
GU172166	chalcone synthase	100	40
D87984	thioredoxin	100	100
GU363529	phenylalanine ammonia-lyase	100	57
D87982	legumin-like protein	101	78
D87980	legumin-like protein	100	74
AF152003	major allergenic storage protein	102	79
EF488807	BW10KD allergen protein	100	100
AB055892	BW8KD allergen protein	76	100
AY256960	13S globulin	100	47
FJ430161	16 kDa major allergen protein	100	100
DQ304682	16 kDa allergen protein	99	98
DQ681064	metallothionein MT10	99	99
DQ681063	metallothionein MT5	97	97
AF056203	metallothionein-like protein	87	87
AB232935	S-like RNase	47	44
AB232934	S-like RNase	76	89
AB355645	phytochelatin synthase	90	73
DQ241824	aspartic proteinase-like protein	54	16
AY536047	aspartic protease	99	31
AB306326	multidrug resistance-associated protein	24	21
DQ289792	proteinase inhibitor BTIw1	99	99
D87983	protein declined during seed development	100	56
AF191099	peroxiredoxin	95	100
AB290016	autophagy-related protein 8	62	91
AB290015	hypothetical protein	63	49
AY126718	fagopyritol synthase 1	75	29

Table S7 Primer sequences of 150 genic-SSR markers used for marker validation

Marker	Transcript ID	Length of transcript (bp)	SSR motif	Start	End
2218	comp57914_c2_seq9	1707	(CA)9	1438	1455
5176	comp57865_c0_seq1	1531	(AAG)5	1390	1404
6819	comp53706_c0_seq1	1540	(AT)6	198	209
1722	comp32140_c1_seq1	612	(GAT)5	110	125
2158	comp48089_c0_seq1	964	(A)11	590	600
2204	comp52021_c0_seq1	393	(AG)6	300	312
2210	comp46527_c0_seq1	577	(TC)6	259	270
2212	comp53958_c2_seq2	401	(GA)6	291	302
2214	comp199246_c0_seq1	409	(AT)6	303	314
2216	comp98016_c0_seq1	266	(TA)6	106	117
2228	comp46594_c0_seq1	445	(TA)6	301	312
2234	comp54915_c0_seq2	744	(AT)7	371	384
2252	comp45732_c0_seq1	794	(GT)6	73	84
2270	comp43024_c0_seq1	661	(AT)6	338	349
2276	comp44134_c0_seq1	440	(TC)8	140	155
2288	comp12116_c0_seq1	256	(A)11	154	164
2298	comp47174_c0_seq1	1903	(TGT)5	798	812
2310	comp44160_c0_seq1	472	(TTC)5	230	244
2312	comp48275_c0_seq1	514	(CCA)5	241	255
5184	comp44212_c0_seq1	596	(AAG)5	237	251
5206	comp49625_c0_seq1	400	(CT)6	252	263
6783	comp36899_c0_seq1	421	(A)10	175	184
6785	comp114710_c0_seq1	498	(T)11	53	63
6805	comp5442_c0_seq1	380	(CT)6	155	166
6813	comp43759_c1_seq1	280	(AG)8	115	130
6821	comp57859_c0_seq1	1184	(GA)6	997	1008
6827	comp37166_c0_seq1	648	(CT)6	517	528
6835	comp31686_c0_seq1	518	(CA)6	187	198
6841	comp47203_c0_seq1	1014	(TC)7	108	121
6843	comp38272_c1_seq1	661	(GA)6	121	132
6853	comp56406_c0_seq1	1747	(TC)8	1597	1612
6855	comp57370_c0_seq1	363	(AT)7	139	152
6861	comp43482_c0_seq1	780	(CCT)6	94	111
6871	comp45384_c0_seq1	523	(ATC)5	226	240
6873	comp47179_c0_seq1	527	(AGA)5	199	213
6875	comp129035_c0_seq1	371	(CAT)5	176	190
6891	comp53816_c0_seq1	974	(ATG)5	731	745
7570	comp141667_c0_seq1	505	(T)11	198	208
7582	comp53844_c0_seq1	669	(TCA)5	161	175
7586	comp45687_c0_seq1	415	(GGA)5	190	204
7596	comp54687_c0_seq2	466	(TTC)5	193	207
7606	comp50902_c0_seq1	565	(GCA)5	207	221
7614	comp33533_c0_seq1	379	(GTT)6	191	205

7618 comp50595_c0_seq1	510 (TCT)5	295	309
7622 comp52009_c0_seq3	1159 (GGT)5	802	816
7628 comp44821_c1_seq1	368 (ACA)5	142	156
7648 comp50221_c0_seq1	844 (CGG)5	60	74
7662 comp127516_c0_seq1	292 (TCT)5	70	84
7672 comp52476_c1_seq1	516 (T)10	271	280
7678 comp44802_c0_seq1	957 (TGG)5	797	811
7682 comp52653_c0_seq2	1323 (TGA)6	168	185
7684 comp49043_c0_seq1	607 (CAA)5	221	235
8963 comp31655_c0_seq1	379 (T)10	312	321
8983 comp37728_c1_seq1	352 (A)11	259	269
8995 comp243205_c0_seq1	241 (AT)7	45	58
9007 comp56162_c0_seq1	359 (AT)6	43	54
9011 comp38176_c0_seq1	582 (GA)6	146	157
9013 comp57982_c0_seq1	960 (A)14	189	202
9023 comp54381_c0_seq1	764 (CT)8	118	133
9033 comp32784_c0_seq1	458 (TA)6	337	348
9045 comp7031_c0_seq1	293 (AG)6	35	46
9069 comp46352_c0_seq2	502 (CTT)5	135	149
9077 comp47131_c1_seq2	554 (GGA)6	371	388
9079 comp49368_c1_seq1	527 (TGA)5	414	428
9095 comp49452_c1_seq1	1051 (TGC)5	195	209
1720 comp53224_c0_seq7	1017 (GCT)8	371	394
1724 comp58433_c1_seq2	738 (AAC)5	374	388
1726 comp36420_c0_seq1	787 (TGA)7	433	453
1728 comp54906_c0_seq1	1169 (A)12	825	836
1730 comp43333_c0_seq1	626 (A)10	179	188
2150 comp58369_c2_seq6	2403 (T)10	2102	2111
2152 comp46028_c0_seq1	2501 (A)10	146	155
2336 comp67209_c0_seq1	891 (CTT)5	691	705
2338 comp56332_c0_seq1	1611 (TC)6	261	272
2340 comp49781_c0_seq1	1657 (GAA)5	1222	1236
5150 comp120593_c0_seq1	517 (CAG)5	164	178
5152 comp54091_c0_seq2	983 (GTA)7	768	788
5224 comp43050_c0_seq1	1232 (ATA)5	1113	1127
5226 comp51844_c0_seq1	1459 (A)14	209	222
5228 comp58692_c0_seq1	4265 (TGT)5	1862	1876
5230 comp54736_c0_seq1	810 (T)10	170	179
6751 comp52333_c1_seq1	1597 (GAG)5	482	496
6753 comp57064_c0_seq1	1808 (AGC)5	1161	1175
6755 comp37222_c0_seq1	1214 (CCG)6	210	227
6757 comp52430_c0_seq3	1453 (GTT)5	495	509
6759 comp58119_c0_seq1	1718 (TAA)5	97	111
6899 comp56038_c0_seq6	1798 (TGT)6	489	506
6801 comp58195_c0_seq2	1708 (ATC)5	253	267
6903 comp48804_c0_seq1	1320 (A)10	152	161

8921 comp53829_c0_seq2	1648 (T)10	178	187
8923 comp51905_c0_seq1	349 (CCG)5	133	147
8925 comp55456_c0_seq1	769 (A)11	183	193
8927 comp58419_c0_seq1	2035 (CAT)6	429	446
8929 comp57928_c0_seq1	3250 (TCA)6	1044	1061
8931 comp58474_c0_seq1	2425 (T)14	220	233
8933 comp51831_c0_seq1	1330 (A)11	190	200
8935 comp54660_c1_seq4	3509 (T)16	3389	3404
8937 comp44419_c0_seq3	514 (T)13	111	123
8939 comp50615_c0_seq2	1548 (T)11	1309	1319
8941 comp47209_c0_seq1	1150 (TGA)5	228	242
8943 comp51301_c0_seq1	1724 (GAA)5	1538	1552
9105 comp54063_c0_seq1	2040 (T)10	363	372
9107 comp56005_c1_seq2	2395 (TTC)5	1856	1870
9109 comp49178_c0_seq1	1238 (T)10	197	206
9111 comp48562_c2_seq1	278 (TCC)7	114	134
9113 comp43654_c0_seq2	2124 (AGAGG)!	77	101
9115 comp56333_c0_seq2	883 (A)11	152	162
7500 comp57179_c0_seq4	898 (GTT)5	426	440
7502 comp52371_c0_seq1	660 (GTG)6	491	508
7504 comp55105_c0_seq2	989 (T)11	171	181
7506 comp53693_c1_seq7	1729 (A)10	130	139
7508 comp56312_c0_seq3	1563 (T)10	203	212
7510 comp56249_c1_seq3	3093 (TC)8	325	340
7512 comp49466_c0_seq1	669 (GCA)7	173	193
7514 comp52490_c0_seq1	2913 (AT)6	273	284
7516 comp58148_c0_seq2	2357 (CTT)5	189	203
7518 comp58430_c1_seq12	3913 (T)10	1239	1248
7520 comp31776_c0_seq1	883 (T)10aaaat	207	296
7522 comp47286_c0_seq2	969 (T)10	648	657
7524 comp52435_c0_seq1	2423 (AT)8	171	186
7686 comp57504_c0_seq1	1739 (AT)7	1496	1509
7688 comp57216_c0_seq2	1304 (TGA)5	87	101
7690 comp57185_c0_seq6	3810 (CAA)5	2979	2993
7692 comp52779_c0_seq1	1975 (T)13	97	109
7694 comp200649_c0_seq1	321 (T)12	236	247
7696 comp45953_c0_seq1	1234 (CCG)5	225	239
7698 comp52498_c0_seq1	982 (AG)6	868	879
7700 comp46726_c0_seq1	1774 (CTC)5	1287	1301
7702 comp58037_c0_seq5	1891 (AG)6	89	100
7704 comp55401_c0_seq6	1530 (T)10	365	374
7706 comp57956_c3_seq1	2556 (TTC)7	1535	1555
7708 comp58436_c0_seq4	3266 (A)10	873	882
7710 comp57313_c0_seq2	1196 (A)10	160	169
7712 comp58689_c0_seq4	2928 (CTC)6	236	253
7714 comp55250_c0_seq1	1543 (AGA)5	1362	1376

7716 comp50111_c0_seq1	891 (TCT)5	706	720
7718 comp52870_c0_seq1	2025 (TGC)5	304	318
7720 comp49432_c0_seq1	940 (CTT)5	791	805
7722 comp58120_c1_seq1	1213 (A)10	306	315
7724 comp53285_c0_seq2	1285 (T)11	1136	1146
7726 comp52027_c0_seq1	1449 (GA)6	1294	1305
7728 comp35115_c0_seq1	363 (A)10	127	136
7730 comp57256_c0_seq3	3392 (AGC)5	287	301
7732 comp57978_c0_seq3	1497 (T)10	285	294
7734 comp58954_c0_seq4	3168 (AAT)5	2439	2453
7736 comp52013_c0_seq2	2245 (A)10	116	125
7738 comp57705_c1_seq1	533 (T)11	376	386
7740 comp93375_c0_seq1	404 (ACC)5	112	126
7742 comp57725_c0_seq2	2605 (T)10	189	198
7744 comp50426_c0_seq1	1035 (AT)6	124	135

Cluster	Product Size (bp)	Start	Length (bp)	TM (°C)	GC%	Any_th	Fo
comp57914_c2	175	1409	20	59	55,0	0,0	
comp57865_c0	200	1172	22	60	50,0	0,0	
comp53706_c0	190	91	20	61	55,0	0,0	
comp32140_c1	295	28	18	58	55,6	0,0	
comp48089_c0	209	509	20	60	55,0	0,0	
comp52021_c0	144	215	29	60	55,0	0,0	
comp46527_c0	143	213	20	60	55,0	0,0	
comp53958_c2	172	162	20	61	60,0	0,0	
comp199246_c0	102	256	20	60	55,0	14,4	
comp98016_c0	180	11	20	60	55,0	0,0	
comp46594_c0	185	155	20	60	55,0	0,0	
comp54915_c0	168	260	20	60	55,0	0,0	
comp45732_c0	117	1	20	60	50,0	0,0	
comp43024_c0	273	216	20	60	55,0	8,4	
comp44134_c0	165	97	20	60	60,0	0,0	
comp12116_c0	113	79	21	60	47,6	0,0	
comp47174_c0	175	726	20	60	55,0	28,1	
comp44160_c0	129	198	20	60	55,0	6,7	
comp48275_c0	131	155	20	60	55,0	0,0	
comp44212_c0	131	154	24	60	37,5	11,2	
comp49625_c0	272	88	22	55	45,5	0,0	
comp36899_c0	147	126	20	60	60,0	20,4	
comp114710_c0	190	33	20	61	55,0	0,0	
comp5442_c0	136	90	20	60	55,0	0,0	
comp43759_c1	172	23	20	60	55,0	0,0	
comp57859_c0	200	896	21	60	47,6	0,0	
comp37166_c0	138	413	20	60	55,0	0,0	
comp31686_c0	200	55	20	59	50,0	0,0	
comp47203_c0	195	58	21	61	57,1	0,0	
comp38272_c1	194	276	25	60	36,0	0,0	
comp56406_c0	109	1533	20	60	55,0	0,0	
comp57370_c0	118	76	21	60	47,6	3,0	
comp43482_c0	109	46	21	60	52,4	0,0	
comp45384_c0	177	88	20	60	55,0	0,0	
comp47179_c0	170	112	21	57	47,6	0,0	
comp129035_c0	141	114	20	59	55,0	0,0	
comp53816_c0	142	654	25	60	32,0	0,0	
comp141667_c0	250	53	20	60	50,0	0,0	
comp53844_c0	226	71	20	60	55,0	0,0	
comp45687_c0	201	117	20	60	55,0	0,0	
comp54687_c0	222	129	20	60	55,0	0,0	
comp50902_c0	239	164	20	60	50,0	0,0	
comp33533_c0	222	31	25	60	36,0	0,0	

comp50595_c0	222	130	20	60	55,0	0,0
comp52009_c0	218	727	20	60	55,0	0,0
comp44821_c1	216	25	20	60	55,0	0,0
comp50221_c0	121	34	21	58	52,4	1,5
comp127516_c0	191	51	20	60	50,0	0,0
comp52476_c1	293	116	20	60	55,0	0,0
comp44802_c0	194	727	20	60	50,0	0,0
comp52653_c0	185	54	20	60	55,0	0,0
comp49043_c0	214	63	20	60	55,0	1,2
comp31655_c0	122	241	20	59	55,0	3,9
comp37728_c1	150	187	20	60	50,0	0,0
comp243205_c0	187	26	20	61	60,0	0,0
comp56162_c0	103	10	22	58	42,9	0,0
comp38176_c0	176	42	20	60	55,0	5,3
comp57982_c0	192	75	20	60	50,0	1,0
comp54381_c0	136	90	20	60	55,0	0,0
comp32784_c0	164	262	20	59	55,0	0,0
comp7031_c0	274	8	20	60	50,0	0,0
comp46352_c0	144	98	20	60	55,0	0,0
comp47131_c1	205	276	20	60	55,0	0,0
comp49368_c1	128	361	20	60	55,0	0,0
comp49452_c1	175	152	20	60	50,0	0,8
comp53224_c0	225	299	21	59	52,4	5,9
comp58433_c1	223	297	20	60	50,0	0,0
comp36420_c0	299	242	20	60	50,0	1,9
comp54906_c0	250	673	20	60	55,0	0,0
comp43333_c0	224	40	20	58	45,0	0,0
comp58369_c2	292	1980	20	60	50,0	17,3
comp46028_c0	300	61	20	60	55,0	0,0
comp67209_c0	296	570	20	60	55,0	0,0
comp56332_c0	232	160	20	60	55,0	0,0
comp49781_c0	216	1093	20	60	50,0	12,7
comp120593_c0	198	52	20	60	60,0	0,0
comp54091_c0	230	643	20	60	55,0	0,0
comp43050_c0	254	967	20	60	55,0	0,0
comp51844_c0	220	109	21	59	52,4	3,8
comp58692_c0	282	1759	20	60	50,0	5,8
comp54736_c0	261	45	21	59	47,6	0,0
comp52333_c1	290	380	20	60	55,0	0,0
comp57064_c0	288	1060	20	60	55,0	5,6
comp37222_c0	283	50	20	60	55,0	0,0
comp52430_c0	296	300	22	60	45,5	17,8
comp58119_c0	265	8	20	60	55,0	0,0
comp56038_c0	277	387	20	60	55,0	0,0
comp58195_c0	272	142	21	59	42,9	0,0
comp48804_c0	207	59	20	60	50,0	0,0

comp53829_c0	294	21	20	59	50,0	8,5
comp51905_c0	268	32	20	60	60,0	0,0
comp55456_c0	195	71	22	57	40,9	18,3
comp58419_c0	248	318	20	60	55,0	0,0
comp57928_c0	238	957	20	59	50,0	0,0
comp58474_c0	300	62	21	60	47,6	0,0
comp51831_c0	266	115	20	60	60,0	0,0
comp54660_c1	241	3239	20	60	50,0	17,7
comp44419_c0	286	37	20	59	55,0	0,0
comp50615_c0	207	1231	20	60	55,0	0,0
comp47209_c0	277	140	21	59	52,4	0,0
comp51301_c0	206	1453	20	60	55,0	0,0
comp54063_c0	293	211	20	60	55,0	1,2
comp56005_c1	281	1670	20	60	50,0	0,0
comp49178_c0	229	59	21	59	52,4	0,0
comp48562_c2	231	21	20	60	55,0	0,0
comp43654_c0	263	4	21	58	42,9	0,0
comp56333_c0	265	3	20	60	55,0	0,0
comp57179_c0	247	298	20	60	60,0	2,9
comp52371_c0	267	326	20	60	55,0	7,4
comp55105_c0	194	63	22	59	40,9	0,0
comp53693_c1	157	58	22	57	40,9	0,0
comp56312_c0	283	68	20	60	50,0	38,2
comp56249_c1	215	244	20	59	55,0	0,0
comp49466_c0	239	74	20	60	55,0	0,0
comp52490_c0	227	200	21	60	47,6	0,0
comp58148_c0	239	65	20	60	50,0	12,2
comp58430_c1	297	1151	20	59	50,0	0,0
comp31776_c0	267	102	19	60	63,2	0,0
comp47286_c0	201	554	20	60	60,0	1,9
comp52435_c0	223	52	20	60	50,0	2,4
comp57504_c0	208	1410	20	60	55,0	0,0
comp57216_c0	169	8	20	60	55,0	0,0
comp57185_c0	190	2886	20	60	55,0	0,0
comp52779_c0	234	18	21	57	38,1	0,0
comp200649_c0	290	31	22	60	50,0	0,0
comp45953_c0	278	103	20	60	50,0	0,0
comp52498_c0	247	710	20	60	50,0	5,8
comp46726_c0	273	1151	20	60	55,0	0,0
comp58037_c0	291	20	20	59	55,0	0,0
comp55401_c0	287	270	20	60	55,0	0,0
comp57956_c3	283	1362	20	60	55,0	0,0
comp58436_c0	287	785	20	58	45,0	0,0
comp57313_c0	259	58	22	60	50,0	17,2
comp58689_c0	266	138	20	60	55,0	0,0
comp55250_c0	169	1290	20	60	60,0	0,0

comp50111_c0	239	598	21	58	42,9	13,8
comp52870_c0	217	192	20	60	55,0	5,0
comp49432_c0	239	648	20	60	60,0	0,0
comp58120_c1	285	147	20	60	50,0	0,0
comp53285_c0	242	982	20	60	55,0	3,5
comp52027_c0	248	1140	20	60	60,0	28,9
comp35115_c0	218	10	22	59	40,9	12,9
comp57256_c0	291	208	20	59	55,0	0,0
comp57978_c0	288	103	20	60	55,0	19,5
comp58954_c0	284	2263	20	60	55,0	0,0
comp52013_c0	264	2	22	60	45,5	0,0
comp57705_c1	264	199	20	60	55,0	0,0
comp93375_c0	178	24	20	60	55,0	0,0
comp57725_c0	287	14	21	60	52,4	0,0
comp50426_c0	288	29	20	59	55,0	3,1

Forward Primer

3'_th	Hairpin	Sequence	End	Length (bp)
0,0	0,0	TATCCGGCTAACCCCTACCA	1583	22
0,0	0,0	GCTTTGGAGAAGGGACTTTTGC	1271	21
0,0	0,0	CAGTTTGAAGGGCAGCAGGA	280	20
0,0	0,0	AAGCAAAGGCAAGTGGGC	322	22
0,0	43,3	GCTAGTGAGGCAGCTGAGAA	717	20
0,0	0,0	TGGCGCCTTTGCCTAGAATC	358	20
0,0	0,0	TCTAACATGTGCCCCTGGTG	355	20
0,0	0,0	GGGTCTGAGAACGAGATGCG	333	19
0,0	40,8	GTGAGTCGGAAACCGTAGCA	357	20
0,0	0,0	TTTTCCACCTCAGAGGGCAG	190	20
0,0	0,0	ATGGCGTAGGGCATAACAGAG	339	25
0,0	36,3	CTTGCCCAGAGCCAAGGTAT	427	23
0,0	0,0	CACGCCCTCCACTTTTTCTT	117	21
6,8	0,0	CGTGCAACTACCACACTTGC	488	20
0,0	0,0	CCGGCACTACTCCATTTCCC	261	20
0,0	0,0	CCCCAACAAACAGAGCCAAAA	191	20
0,0	0,0	CCCAGTTAAGGCCGGAATGA	900	20
0,0	0,0	GGGAATCCTCCACCGTTGAT	326	20
0,0	0,0	ATGAGTGCACCACATCCACC	285	20
11,2	0,0	AAGCATTCTGAGTTTATCAGCTAC	284	25
0,0	0,0	TGGTTAGCTTGTATGATGGTGG	359	22
0,0	0,0	ACTCCCCGATACCTGTACCC	272	20
0,0	0,0	GGTTCGGGAGAACTGAGCCAA	222	20
0,0	0,0	AAGACAGTGGTGGTAGTGGC	225	20
0,0	0,0	CGGCTGATCGGTTAGGGTTT	194	20
0,0	0,0	ACCTGCAAAGCAACCTAGTGA	1095	20
0,0	0,0	AAACCACGAGCTTCCTCTCC	550	20
0,0	0,0	TGTTGAAGCCCAATGGTCA	254	20
0,0	0,0	CACCGCTCTCCAGATCCATAC	252	21
0,0	46,8	TGAGTAATTGAAAATGGAGATCGAC	469	20
0,0	0,0	GGCCTCCTTGAACAGCTAGA	1641	20
0,0	0,0	TGTACGTACGCTGGAGAAACT	193	20
0,0	36,0	CCAGCGTTCTCAAGTTTACGC	154	20
0,0	0,0	GAGAGATGAGCACAGCGTCA	264	20
0,0	42,4	TTGATGGACACTAGGCAAGGT	281	20
0,0	0,0	GAATCCACCATCACCAGGCA	254	20
0,0	40,4	TGAGGGGCCAATGCATGAGA	795	20
0,0	0,0	ACAAGCAACTCCACAAGTGC	302	20
0,0	36,2	ATCCGAAACCGCCTCCTTAC	296	20
0,0	0,0	GCATCAACCTTGGTAGCCGA	317	20
0,0	0,0	GCTTGAAATGGCCTTGAGCC	350	20
0,0	0,0	AGTTCGGCGGGAAGATGAAT	402	20
0,0	0,0	TGAAAATTGACTTAAACTCGGGACT	252	20

0,0	0,0	CGGGAGGCCAGTGTA AAAAGT	351	21
0,0	0,0	GGACTGGCCAGCTGATTTCT	944	21
0,0	0,0	TCCAGATCGAGCAACTGAGC	240	20
0,0	0,0	CCTTTAAGCCTCCGATGTCGT	154	20
0,0	33,7	AATGGCGGCAAATCTCGCTA	241	20
0,0	0,0	TGCCTGGAAACCTCTTCGTC	408	20
0,0	43,9	GCCAACATGCAAATGACCCA	920	20
0,0	0,0	CCACGTTCCCACCATACGAA	238	20
0,0	39,8	CAGATGGTACAAGGACGGCA	276	21
0,0	0,0	GAGGATTGGCTGTCTGCTCT	362	20
0,0	0,0	TAGAAGATTTGTGGCCGCCT	336	25
0,0	0,0	CCAAACGGCTAGTCTCGGTC	212	20
0,0	0,0	AGTTCACGTGAGATCCTAGTGT	112	20
0,0	41,8	GGCTCTTCTTTAGCACCAGC	217	20
1,0	43,5	TGTAGCTGCAGTTGAAGACG	266	20
0,0	0,0	TCGGGGCCACAAGTCAATAC	225	20
0,0	0,0	AGGGTGGTGGTAGCCTTAGT	425	20
0,0	0,0	ACGGTTTGCCTTTGGATTGC	281	20
0,0	35,9	TTCGATACCTCATCGTCGGC	241	20
0,0	0,0	GAAGGCCGATTACGCTTCCT	480	20
0,0	0,0	TTCCATTTCCCCGCTTCCTG	488	20
0,0	44,4	AGCTGCAGCCATGTCCATAA	326	20
3,7	40,6	CCCTAAGTGTCAACTGGGACA	523	20
0,0	0,0	TTGACCCTGCTGTGCTCATT	519	20
0,0	0,0	TGTCATAGTTGCTGCCGGTT	540	20
0,0	0,0	AGTGCAGGTGTGGGATTACG	922	20
0,0	0,0	ACCCCAATTTTCAATGGCGA	263	20
16,0	0,0	TGCTGATCATGGTCTTGGA	2271	20
0,0	0,0	CCCGGTTCTTTCCTGCATCT	360	20
0,0	40,2	CGGATCGGGTCATTCGAACT	865	20
0,0	41,3	GGAAGTAGAAGCCCAAGCCA	391	20
11,4	0,0	ACAATTCTCGCCCGGATTGT	1308	20
0,0	44,8	GCAGACGAAGACCAGAAGGG	249	20
0,0	0,0	TCTCCTTAGGCCAGCATTGC	872	20
0,0	0,0	GAGCCCACAACATGCAAAGG	1220	22
0,0	0,0	GGATCCTAAACTGCTCCCTCA	328	20
0,0	0,0	TGTTGGA ACTGGATGCGTCT	2040	20
0,0	0,0	GCAAGAAGATCGAGCAAGTGT	305	20
0,0	0,0	CGCTTATTTTGGCAGCCTCC	669	20
0,0	33,2	CCATCTCGTTGGGCAGAGAA	1347	20
0,0	42,9	GATGTCAAACAGCAGGACGC	332	20
0,0	36,9	ACCAAACATGCCATGAGGATCA	595	20
0,0	0,0	CCATCAAATGCCGCTACTGC	272	20
0,0	0,0	CTGGTGTATTGCCCCCTTGT	663	20
0,0	0,0	ACAACACAAGAACGACAAGCT	413	20
0,0	0,0	AAAGCTGCTCACTCGAACCA	265	20

0,0	38,9	TCAAGGGAGCCCATTCTTCA	314	20
0,0	0,0	GAGGGGAGATGACGGGTAGT	299	20
0,0	0,0	AAGAGACTGCAGATTTCAAGGA	265	20
0,0	39,4	TGCTTGCTTTACCTCCTCGG	565	20
0,0	0,0	AAGCGAGCAACCAATCAAGG	1194	20
0,0	0,0	TGCAACTACCAGTCCACAAGT	361	20
0,0	0,0	TCCCTCTACCCAAGGCTACC	380	20
17,7	0,0	TCTTTTGGATTTTCGCGGCG	3479	22
0,0	0,0	CAGAGTTGGAGCCCGTGTAT	322	22
0,0	0,0	AACACCTTACCCACTCCAGC	1437	20
0,0	37,4	AGTTAGTCAGGACCTCCCAGA	416	20
0,0	0,0	GTGGTGTCTGTGGGTGGAT	1658	20
1,2	45,4	TTCGATTCACCTTGGGGTCG	503	20
0,0	0,0	TGTCTACGGGTGGCAAGTT	1950	20
0,0	0,0	TGTCTGCCCTGTAATACACCC	287	20
0,0	0,0	GATCCGCCAAGCACATTAGC	251	20
0,0	0,0	ACGCCATTAAAGAAGGTTGGA	266	20
0,0	38,1	CGCATCCGCCATGTGTAATC	267	20
0,0	0,0	CCGAGATCCATGAGTGCTCC	544	20
0,0	33,9	CGAACTCCGGCGTTACTCAT	592	21
0,0	0,0	TCCACATGAAAAAGGGATTGCA	256	20
0,0	0,0	AGGAAAAATAGCCTCACTCACA	214	20
0,0	45,2	CCGAAGAAAAAGCAGCTGCA	350	20
0,0	0,0	AACCCAAATCTCCTTCCCCC	458	20
0,0	0,0	CAGTACCCTTCACACGCCAT	312	20
0,0	0,0	TCAACACAAACCACCTTCCCA	426	20
0,0	0,0	AGGCTTGCCTTTGGAAGTGA	303	20
0,0	0,0	CAAACCTTGTGTGCAGCTCA	1447	21
0,0	45,8	GATGGAGCCCTCGTGAAGG	368	21
0,0	37,7	GATGAAGACGACGAGGTCCC	754	20
0,0	44,8	AACATGCCATGTCAGCAACG	274	20
0,0	38,4	GGTTCGTCCGTATCTCGCTT	1617	22
0,0	0,0	TCTCAATCGCCGAATCCTC	176	20
0,0	38,5	ATTCGTAAGGCTGGTGGTGG	3075	20
0,0	43,4	AACGATTGCTGAATGCTTTGA	251	20
0,0	0,0	GCCCTTCCTTTGTGTTATTCCC	320	22
0,0	0,0	TCCCTCAAATCCACGCAACT	380	20
5,8	32,6	AACGACGAAAGAGCCATCGT	956	20
0,0	33,4	AAAGTGCAGGGTCTGTCTGG	1423	20
0,0	0,0	GGGTTTTAAGAGGGGTGGC	310	20
0,0	0,0	CACATCCCACACTCGTCACA	556	20
0,0	37,9	CCACACCAGGGGGCATAAAA	1644	20
0,0	0,0	TCTGTGATTTTGGCCCCTTT	1071	20
2,8	35,7	GCTGGCAGTACAATAATGAGCC	316	20
0,0	35,6	GAAAAAGCGGACCACCACAC	403	20
0,0	39,4	GAAGAGAGGGTTCAGACGCC	1458	20

0,0	0,0	AGGCTATTCAAGTTCAGCCAA	836	20
0,0	40,9	CCCTTGGCCGAGATTCGTTA	408	20
0,0	36,6	GAGAACAACTCTCCTGGGC	886	20
0,0	43,0	TCTCCTTCATGCCTTGACA	431	20
3,5	0,0	GCGATGAAAACCCTAACGCG	1223	20
3,1	39,3	GAACCCTCAAACCCTAGGGC	1387	20
3,0	38,2	AGAAACAATCGCTGCAATTGGA	227	20
0,0	0,0	CGTCGTCATCCTCCTCATCA	498	20
0,0	39,1	AGTCGACAACCGGAGAAACC	390	21
0,0	34,7	CCCAGCCAGGTCCAAGTTTT	2546	20
0,0	0,0	TCCACACCAAATCGCAAATCC	265	20
0,0	0,0	GGGGAGAGACTTTGGGCAAT	462	22
0,0	0,0	CTACCCACGTTGCCCATCAA	201	20
0,0	0,0	GCCCCATAATGCAACTCAAGG	300	20
3,1	0,0	CAAGGCAAGTGTACATGGGC	316	20

TM	Reverse Primer				Sequence
	GC%	Any_th	3'_th	Hairpin	
60	50,0	0,0	0,0	0,0	ACACCTTCACCCTTGAATCTCG
60	52,4	0,0	0,0	0,0	GATCGAAGGACCAAGCCCTTA
59	50,0	0,0	0,0	0,0	TCTCCACCAGCAACAACGTA
60	57,0	15,8	0,0	0,0	TTGGGTTGGTCTTCAGCATCTA
60	55,0	0,0	0,0	0,0	TCTGCACGGACTAGTGAAGC
60	55,0	0,0	0,0	0,0	CCTTAACTCACCAGACCCCA
60	55,0	0,4	0,0	42,9	CTCCTCAAGACGCGTGAAC
61	68,4	0,0	0,0	0,0	GACCCCCACCCAGACCAAC
59	50,0	7,3	0,0	0,0	ACCCTATCGCAGGAAAGCAA
59	45,0	10,6	0,0	0,0	TTGCACATTGTCGTTTCCCA
55	40,0	0,4	0,0	0,0	TGCAGTTCAAGTCCATAAACAAGTC
60	39,1	21,0	10,4	45,5	AGCAAAACCTATGCTTTTACTGC
59	52,4	0,0	0,0	0,0	TGCAAGCGATAGTAGCGATCC
60	50,0	0,0	0,0	0,0	ACGTTACTCGTGCAAGAAGC
60	50,0	0,0	0,0	0,0	TTGTGACTTCTGATCCGCCA
58	55,0	0,0	0,0	0,0	CAGGGGAGGGTTGTTGTTGT
60	55,0	0,0	0,0	44,9	TAAGGCTGGTGGTGGGGATA
60	55,0	0,0	0,0	46,6	TCCTGAAGCCGAGTCTCCTA
60	55,0	0,0	0,0	0,0	ATTGCAGCATTAGGAGGCGG
60	36,0	0,0	0,0	0,0	ACTAGTGACCATTCAATTTGAAACCT
55	45,5	0,0	0,0	0,0	AGGGGAAAAGAAGGAATGTCTC
60	55,0	0,0	0,0	0,0	CTGAAGTGGCCGTCACAGAT
60	50,0	1,3	0,0	0,0	AGTTCGGACAATCCAAGTGC
60	55,0	11,0	11,0	40,8	TAGAGGAAAGTAGTGCGGCG
61	55,0	0,0	0,0	0,0	CCGTCGACGCCTTTGTTTTTC
60	55,0	0,0	0,0	0,0	GCCCCAATCGAAGTTTCACC
59	50,0	2,7	2,7	36,6	AGAATTTGCGGGAAGGGTGA
58	45,0	0,0	0,0	0,0	TGCGTTCTTGTTTGGCTACA
61	52,4	0,0	11,2	0,0	CGTCATCTGCTCTGAATCCCA
60	55,0	0,0	0,0	0,0	TACTCCCAACCAGCAATCCG
60	55,0	0,0	0,0	0,0	GCAGTGACCGAAGTGCAGAT
60	55,0	0,0	0,0	0,0	CAAGAGCTGGCAAGGTTAGC
60	55,0	10,6	1,4	39,2	CTGTGTGCGACTCTTTCCCT
60	50,0	0,0	0,0	41,2	TCGATGGCGAAACCCAAGTT
60	50,0	29,4	8,3	0,0	GTGAGGAGCAATTGGCGTTT
59	50,0	1,3	0,0	0,0	GTTTTTCGATGCCAACTGCGA
60	50,0	0,0	0,0	0,0	GGAGGCTTCTTAGTTCTTTTCCC
60	55,0	0,0	0,0	0,0	GGCTTATGCACCCTACCCAA
60	55,0	0,0	0,0	0,0	GGGGTTTTGGTGCAGGTACT
60	55,0	0,0	0,0	0,0	CTCCGCCACTTTCTCTCGAT
60	55,0	0,0	0,0	46,0	TTCTCGTGGCTGAGAAGTCC
60	55,0	5,4	0,0	0,0	AGTAGTCCCGGAGCTCACAT
60	55,0	7,8	7,8	41,8	CTAAGGTGGCTGCAGCAGAA

60	47,6	0,0	0,0	41,2	TTCCTCCATTGAAGCACCACT
60	52,4	0,0	0,0	0,0	GTGAGGGCTGGTGTGTAAGA
60	50,0	0,9	0,0	0,0	CCTGGCACTGCAAAATGCTT
61	55,0	0,0	0,0	0,0	GGAATCGTGGTGGATTCCGA
60	55,0	9,7	5,8	0,0	TGAGGATTTTCAGGGGCTTGG
60	55,0	0,0	0,0	0,0	TGGAAGATGAGCGCAGGAAG
60	60,0	20,9	0,0	0,0	CGGTGGATACGGGTAGGTTG
60	50,0	0,0	0,0	42,4	CGGCATTACCAAGAAACCCA
60	47,6	0,0	0,0	0,0	ACTGCAACTGAGCTGATGACA
60	50,0	0,0	0,0	0,0	TCAGCTCAATGCAAACCTCC
60	36,0	0,1	0,0	43,6	GGGAAATCACAATCATAGTAAAGCA
60	55,0	0,0	0,0	42,7	ATGGTCCTTGGCTGAAGCTC
60	55,0	3,5	0,0	0,0	CAACAGGGGAGAATGACGCT
60	50,0	13,3	0,0	45,0	ACTGGACATGGTGAAATGGG
60	55,0	3,9	0,0	46,1	CTTCCCGGTTTTCTCTGGTGA
60	55,0	0,0	0,0	0,0	TCTGGATAAGGGTTGGGGGT
59	50,0	0,0	0,0	0,0	TGACATGTGCTCCAAACGCT
60	55,0	10,2	10,2	0,0	CACCGACAACAACCGTTTCC
60	55,0	10,8	10,8	45,6	GCCGGCATCTAAAATGGACG
61	60,0	0,0	0,0	34,4	GTCTCCGTAGCAGCAAGAGG
60	55,0	0,0	0,0	0,0	GACGCAAGAAAATGGCCCAC
60	60,0	0,0	0,0	41,4	CATCATGAGACTCGGGACGG
60	55,0	0,0	0,0	0,0	GGCCAGAGAGAAGACGTTGT
60	50,0	0,0	0,0	34,0	TGCTGATCTTTTCCGCCCTT
60	55,0	6,0	4,2	0,0	TCAGCCCAAGAATCCGGAAC
60	50,0	8,4	0,0	38,7	ACTGGAACAACGATGCCCAA
60	55,0	0,0	0,0	39,6	CCTTCTCTTCAAGGGCGTGT
60	60,0	0,0	0,0	0,0	GCCCATCTCTAAGAGCCCAC
60	55,0	0,0	0,0	42,0	GTATGACATGCCGCTTGCTG
60	55,0	0,0	0,0	0,0	ATGGCGACTCTCTCCCTTCT
60	50,0	0,0	0,0	0,0	ATGCCGCCACCATTTTCATG
60	55,0	7,6	0,0	0,0	GCCTTCGTCGACTAGCTTGT
60	55,0	0,0	0,0	33,7	AGGCTTTCCAGGGGTTGATG
59	55,0	0,0	0,0	0,0	CCATCAAAGCCCTGTTTCGG
58	45,5	0,0	0,0	41,4	ACACACCTAAAGAGAGGGTACA
60	55,0	0,5	0,0	0,0	TTGCTTTCGAGACGGCTAGG
59	55,0	0,0	0,0	31,3	CCCTCAAGCCCTTCTTCGAT
60	55,0	0,0	0,0	0,0	CAAATTGTGGCTGCTGGTGG
60	60,0	0,0	0,0	0,0	GCCTCCTCCAGACACAGTTC
60	60,0	21,6	0,0	45,0	CCCTACCGGCCAGGAATTAC
60	55,0	0,0	0,0	0,0	CCTCTTCGGCAACAGCAGTA
60	55,0	0,0	0,0	0,0	GACGTGGTGAAGTTGGAGGT
60	55,0	0,0	0,0	35,9	CAACCACGTCCAGAAGGTCA
60	55,0	0,0	0,0	46,1	CGCAACTTCGTGGTGACAAG
60	60,0	9,6	0,0	34,6	GAAGGAGCTCGACTTGAGG
60	55,0	0,0	0,0	0,0	TTACTTCACTTGCCCCCACC

60	60,0	0,0	0,0	0,0	CCCGTCACGAGGAGATTAGC
60	55,0	0,0	0,0	0,0	AAACACCAGTCGCCACCTAG
60	60,0	0,0	0,0	0,0	CTACAGCCGTTTCCAGGGAG
60	55,0	2,1	0,0	42,7	GGCTCCTGCTCCAGCTAAAA
60	50,0	5,8	0,0	45,2	ACTCCTTGTTGGATCGCGTT
59	50,0	0,0	0,0	0,0	TCGTCCAATCACAGCTCGAA
60	55,0	0,0	0,0	0,0	AGATCGTGGATGAGTTCGCC
57	36,4	1,8	0,0	39,7	ACACAATCTTGAAATTGCCACA
60	45,5	2,6	0,0	37,2	TCCTGCTCCTTTTTCAAGCTGA
60	55,0	0,0	0,0	0,0	GACTCACCACTGCACCTTGA
59	50,0	0,0	0,0	44,2	GCCATTTGCCAGAAGCTGAA
61	55,0	0,0	0,0	0,0	TGCCTCAAAGTCACAACCGG
60	50,0	0,0	0,0	43,6	AAACAGAGCAACCCCAGCTT
59	50,0	0,0	0,0	0,0	TTTCAGGACAGTGAGCAGGA
60	60,0	0,0	0,0	0,0	GGGGAGGGGGAAAACATCAG
60	55,0	0,0	0,0	42,4	CTCACGCATCCTACTGCACA
60	55,0	24,0	0,0	0,0	GGGACGTCGCTGGTAGAAAA
60	55,0	0,0	0,0	35,0	GGGTTTTTCGGGTTCGGAGAT
60	55,0	0,0	0,0	0,0	AAGCCGAGCAAGAGAAGGTC
60	47,6	4,2	0,0	42,1	CCAATCAATCGTCTTGCGTGT
60	55,0	4,1	4,1	0,0	AACACTTGCTCCTCCCATGG
60	55,0	15,5	3,9	39,4	CATGGA ACTCCGGCTCCTTT
60	55,0	0,0	0,0	37,4	CGCTGGTAGAAAGCCACAGA
59	50,0	0,0	0,0	0,0	TCCACCTTGATTGCTCCAGA
60	55,0	0,0	0,0	36,0	CTTCGGCGGAACATCCTTCT
60	50,0	4,9	0,0	40,8	AATTTGCCGTAGCTTTTCGCC
60	60,0	4,6	0,0	41,2	GAGGGTCCCCATGAAGGAGA
60	47,6	6,1	0,0	41,0	CAAAGCAGCTTCCTGTTTGCT
60	52,4	0,0	0,0	0,0	GACTCTTGGA CTTTGCAGCC
60	55,0	6,6	0,0	0,0	CCGCTACATTGAGCTCGTCA
60	55,0	0,0	0,0	0,0	AGATGGAAGGTTGGAAGGCG
58	36,4	0,0	0,0	0,0	ACAGCAAACCCAAAACAAAACA
60	55,0	0,0	0,0	0,0	TCGTCCAGCTCTCCAAAACC
60	55,0	0,0	0,0	36,2	GAATGCTCCCAGTCAAGGCT
60	55,0	0,0	0,0	38,7	TGGGGACTCGGAAAGTAGGA
57	40,9	0,0	0,0	0,0	ATTGACGTAGAAAACATCCCCT
60	55,0	0,0	0,0	44,6	AGATGAAGCAGAGGCAGCAG
60	55,0	3,7	0,0	0,0	CAACTTCCCTGCGCTGAATC
59	50,0	4,6	4,6	44,0	AAAATGCGGGGTCATGTGAC
60	50,0	0,0	0,0	0,0	ATTCCTTCCGCCATCAACGT
60	60,0	0,0	0,0	0,0	GATGATGATGGAGGAGGCGG
60	55,0	0,0	0,0	0,0	TCCCCCAAGACTGAGATCGT
60	55,0	0,0	0,0	37,6	CACGACTCCCTGAGGCATTT
60	55,0	4,0	0,0	39,5	GGTGCGAGTAGGCTCTTGTT
60	55,0	0,0	0,0	0,0	GGCGGTTTTGGAGCATCATC
60	55,0	0,0	0,0	0,0	TTTCATCTCTCTACGCCGCC

60	55,0	0,0	0,0	37,6	CACGGGATGGAAAAGCCTCT
60	55,0	14,2	14,2	46,3	AGGGGGAACCTCCTTGAAGGA
60	55,0	18,8	0,0	40,9	GCTCGAGATGGAAGTGTCGA
60	55,0	0,0	0,0	46,0	TGGTCTCGTGGGATTCCAAC
58	45,0	0,0	0,0	0,0	AAACCACGACCGTAACAACA
60	60,0	2,1	0,0	36,1	GACCGGCGACTATAGCACTC
60	55,0	0,0	0,0	0,0	GGTTCCCGGCTTTGGTTTTTC
60	55,0	28,2	0,0	45,1	AGCCGGCTTTATCTGAGACG
59	57,1	0,0	0,0	0,0	CCTCCCATTCTAACCCCTACC
60	55,0	5,2	0,0	41,3	ACTCTGATGGCCTCGTAGGA
60	50,0	0,0	0,0	0,0	TCTTCGCTTGACACCCACAA
58	45,5	0,0	0,0	0,0	ACCAGATTGACAGACAGTTTCC
60	60,0	0,0	0,0	0,0	GTGCCCCGATAGTCGTACTG
60	50,0	0,0	0,0	0,0	TGCAGTTGGAAGTAGCACGT
60	55,0	0,0	0,0	0,0	GATGAACAACGACTTGGCCG

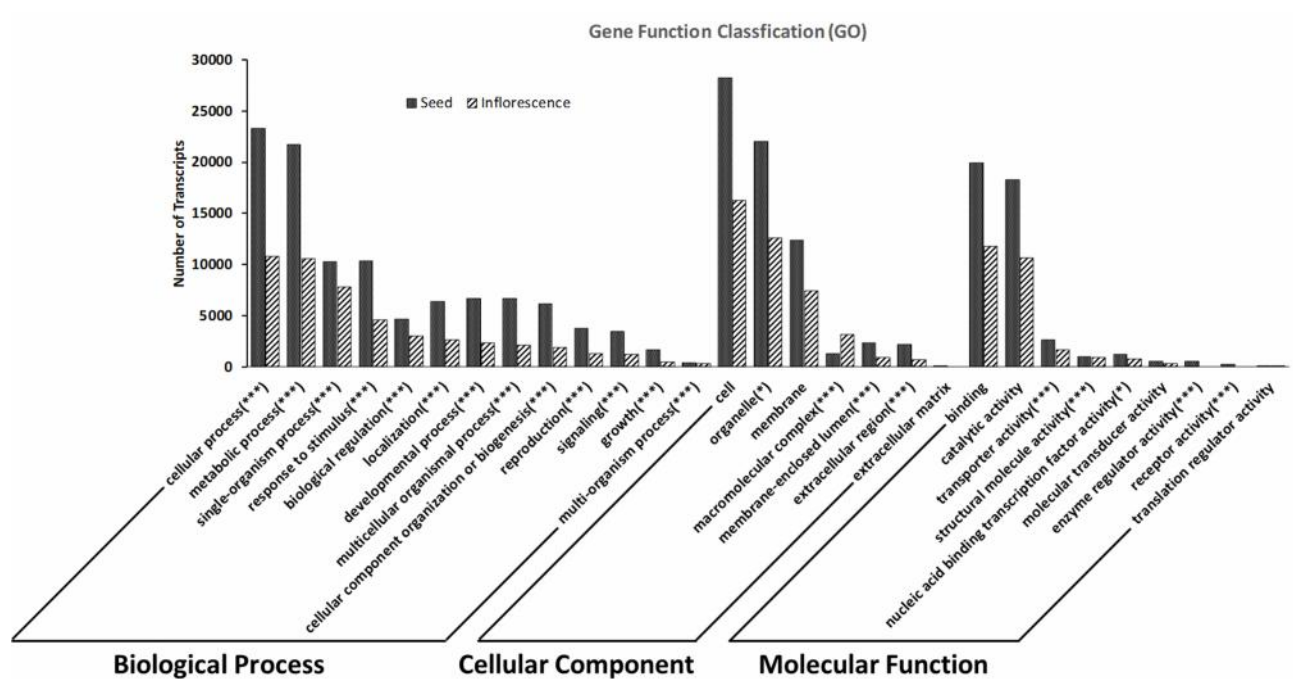


Fig.S1 Gene ontology (GO) classification of assembled transcripts and comparison with the floral

transcriptome (Logacheva et al. 2011) of common buckwheat. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.005$