


RESEARCH ARTICLE

Open Access



Transcriptome-wide mining, characterization, and development of microsatellite markers in *Lychnis kiusiana* (Caryophyllaceae)

Seongjun Park^{1,2} , Sungwon Son³, Myungju Shin², Noriyuki Fujii⁴, Takuji Hoshino⁵ and SeonJoo Park^{2*}

Abstract

Background: *Lychnis kiusiana* Makino is an endangered perennial herb native to wetland areas in Korea and Japan. Despite its conservational and evolutionary significance, population genetic resources are lacking for this species. Next-generation sequencing has been accepted as a rapid and cost-effective solution for the identification of microsatellite markers in nonmodel plants.

Results: Using Illumina HiSeq 2000 sequencing technology, we assembled 67,498,600 reads into 91,900 contigs and identified 11,403 microsatellite repeat motifs in 9563 contigs. A total of 4510 microsatellite-containing transcripts had Gene Ontology (GO) annotations, and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis identified 124 pathways with significant scores. Many microsatellites in the *L. kiusiana* leaf transcriptome were linked to genes involved in the plant response to light intensity, salt stress, temperature stimulus, and nutrient and water deprivation. A total of 12,486 single-nucleotide polymorphisms (SNPs) were identified on transcripts harboring microsatellites. The analysis of nucleotide substitution rates for 2389 unigenes indicated that 39 genes were under strong positive selection. The primers of 6911 microsatellites were designed, and 40 of 50 selected primer pairs were consistently and successfully amplified from 51 individuals. Twenty-five of these were polymorphic, and the average number of alleles per SSR locus was 6.96, with a range from 2 to 15. The observed and expected heterozygosities ranged from 0.137 to 0.902 and 0.131 to 0.827, respectively, and locus-specific F_{IS} estimates ranged from -0.116 to 0.290 . Eleven of the 25 primer pairs were successfully amplified in three additional species of *Lychnis*: 56% in *L. wilfordii*, 64% in *L. cognata* and 80% in *L. fulgens*.

Conclusions: The transcriptomic SSR markers of *Lychnis kiusiana* provide a valuable resource for understanding the population genetics, evolutionary history, and effective conservation management of this species. Furthermore, the identified microsatellite loci linked to the annotated genes should be useful for developing functional markers of *L. kiusiana*. The developed markers represent a potentially valuable source of transcriptomic SSR markers for population genetic analyses with moderate levels of cross-taxon portability.

Keywords: *Lychnis*, RNA-seq, Single-nucleotide polymorphisms, Transcriptomic SSRs

* Correspondence: sjpark01@ynu.ac.kr

²Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea

Full list of author information is available at the end of the article



Background

Microsatellites, or simple sequence repeats (SSRs), are tandem repeats of mono- to hexa-nucleotide sequence motifs that generally vary in length between five and 40 repeats. They are codominant, abundant, multiallelic and locus-specific markers and are found at high frequency in eukaryotic genomes [1, 2]. Microsatellites are powerful markers for population genetic studies on breeding systems, genetic diversity, and conservation genetics [3, 4]. However, traditional approaches used to identify microsatellite markers are usually time-consuming, labor-intensive, and costly processes [4]. Recently, next-generation sequencing (NGS) has been accepted as a rapid and cost-effective solution for the identification of microsatellite markers in nonmodel plants [5, 6]. Microsatellites can be identified in both genomic and transcriptomic sequences with contrasting features. For example, genomic SSRs have an uncertainty associated with protein-coding genes, while transcriptomic SSRs are potentially linked with protein-coding genes and their untranslated regions (UTRs) [4]. Genomic SSRs are highly polymorphic and are randomly dispersed throughout the genome, while transcriptomic SSRs exhibit relatively low polymorphism and occur infrequently [4, 7]. However, transcriptomic SSRs facilitate the understanding of their association with functional genes or phenotypic variation [8]. Transcriptomic SSRs are transferable among closely related species, whereas genomic SSRs have relatively low transferability [4]. Furthermore, transcriptomic SSRs may play an important role in adaptive evolution by generating genetic variation [9].

The angiosperm genus *Lychnis* (Caryophyllaceae) is well studied with respect to the genetic effects of habitat fragmentation on population structure [10–12]. Many *Lychnis* species are drought-tolerant perennials, but some of them require water to survive. Previous studies have shown that drought stress, mating history and nutrient availability influence inbreeding depression in *Lychnis* populations [13, 14]. This genus exhibits diversity in their sexual and mating systems, breeding system and host-pathogen dynamics, sex chromosome evolution, genomic conflict and speciation, and biological invasions [15]. *Lychnis kiusiana* Makino is an endangered perennial herb native to wetland areas in Korea and Japan [16]. The loss of wet meadow habitats has resulted in decreases in the size and number of populations of this species [17]. Fragmentation of habitats may influence the breeding system and may lead to a reduction in genetic variation, resulting in diminished population capacity to adapt to environmental changes [18–21]. Information on genetic diversity is important for planning the management of *L. kiusiana*. Despite its conservation and evolutionary significance, population genetic resources are lacking for *L. kiusiana*. Only five

polymorphic genomic SSR markers have been identified using traditional approaches, and an applied genetic diversity study has been conducted in Japanese populations of the species [17]. However, enriched genomic resources and genetic markers are needed to interpret genetic variation within populations and to investigate the genetic diversity of *L. kiusiana*. Transcriptomic sequencing has been applied for mining microsatellite markers in related species, i.e., *Silene vulgaris* [22], and the results indicated that transcriptomic SSRs provide potential for genomic and population genetic approaches.

In this study, we developed microsatellites from the *L. kiusiana* transcriptome and characterized their frequency, distribution, and function, which promotes an understanding of microsatellite evolution in the *L. kiusiana* transcriptome. We estimated nonsynonymous (d_N) and synonymous (d_S) substitution rates of open reading frames (ORFs) that contained at least one single-nucleotide polymorphism (SNP). In addition, we designed primers for amplifying microsatellite loci and validated the availability of selected primers. Our results offer a valuable resource for studies of population genetics in *L. kiusiana* as well as three other *Lychnis* species.

Results

Characteristics of microsatellites in the *Lychnis kiusiana* transcriptome

Using the Illumina HiSeq 2000 sequencing technology tool, we assembled 67,498,600 reads into 91,900 contigs with an average length of 800 bp. A total of 11,403 microsatellite repeat motifs were identified from 9563 contigs (10.4%; Additional file 1: Table S1). Of the total 11,403 SSRs, tri-nucleotide microsatellites were the most abundant motif (42.2%), followed by mono- (35.3%), di- (13.7%), hexa- (9.2%), penta- (3.6%), and tetra-nucleotide (1.7%) types (Additional file 2: Figure S1). A/T motifs (97.7%) were the most abundant in mono-nucleotide repeats (Additional file 2: Figure S1). With respect to the di-nucleotide motif, AG/CT was the most abundant type, with a total of 800 (51.4%), while there was only one CG/GC. Ten tri-nucleotide motif types were identified with AAT/ATT accounting for approximately 21.4%, while CCG/CGG only accounted for 1.7% (Additional file 2: Figure S1). The hexa-nucleotide motif accounted for the highest number (77) across all types (Additional file 2: Figure S1).

The average length of the microsatellites from the *L. kiusiana* transcriptome was 19.75 bp (Additional file 2: Figure S2A). The length variation of microsatellites was significantly affected by the repeat motif size (Kruskal–Wallis rank sum test, $p < 1 \times 10^{-15}$; Additional file 2: Figure S2B). The length differences between each motif size class were statistically significant (pairwise Wilcoxon

rank-sum test, $p < 1 \times 10^{-15}$ after Bonferroni correction; Additional file 2: Figure S2B), except for two comparisons between mono- and di-nucleotide and between tetra- and penta-nucleotide motifs. The mono-nucleotide motifs had the shortest average length (16.7 bp), while hexa-nucleotide motifs were the longest, with an average of 27.4 bp. The longest microsatellite identified was 96 bp, which was composed of a 16-fold repetition of a hexa-nucleotide motif.

Microsatellite distribution in the *Lychnis kiusiana* transcriptome

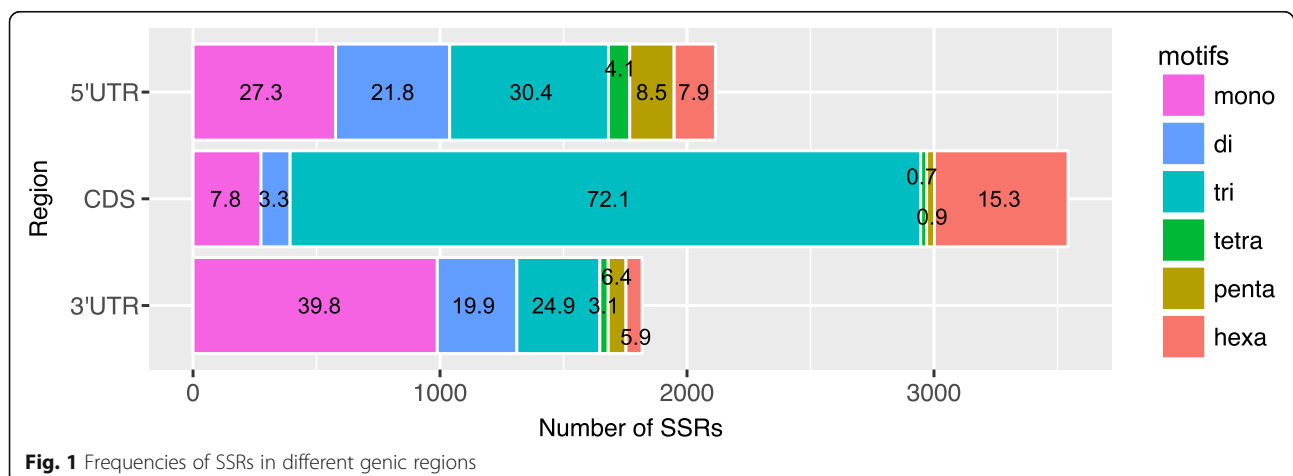
The distribution of microsatellites in the 5' UTR, CDS region, and 3' UTR was investigated. Of 11,403 microsatellites, 2115, 3543 and 1818 were located in the 5' UTR, CDS region, and 3' UTR, respectively (Additional file 1: Table S2). The remaining 3927 microsatellites were excluded from analyses because the transcripts lacked information to determine the CDS region. The frequency of microsatellites in the CDS region was similar to that in the UTRs (Additional file 1: Table S2). However, the frequencies of the motif size classes were significantly affected by the location (Kruskal-Wallis rank sum test, $p < 1 \times 10^{-15}$; Fig. 1). Microsatellites located in the CDS regions were mostly tri-nucleotide motifs (72.1%). Mono- and di-nucleotide microsatellites dominated in the UTRs, showing that the proportions of mono- (mean 39.8%) and di-nucleotide motifs (mean 19.9%) in the UTRs were much higher than those in the CDS regions (Fig. 1). The mean lengths of microsatellites differed significantly between the CDS regions and each UTR (Kruskal-Wallis rank sum test, $p < 1 \times 10^{-15}$; Additional file 2: Figure S3A). The mean microsatellite lengths of only two motif size classes were significantly affected by location (Kruskal-Wallis rank sum test, mono-: $p < 1 \times 10^{-13}$, di-: $p = 0.009301$; Additional file 2: Figure S3B).

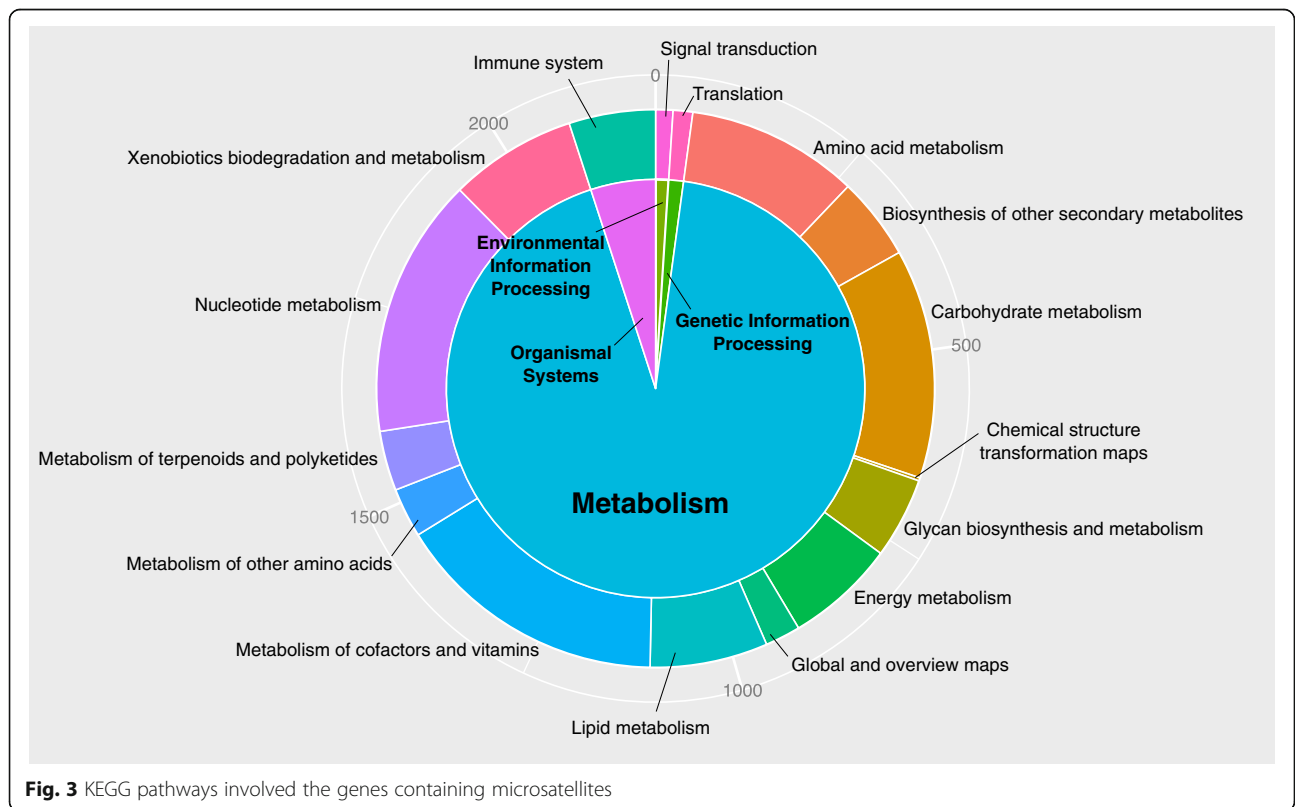
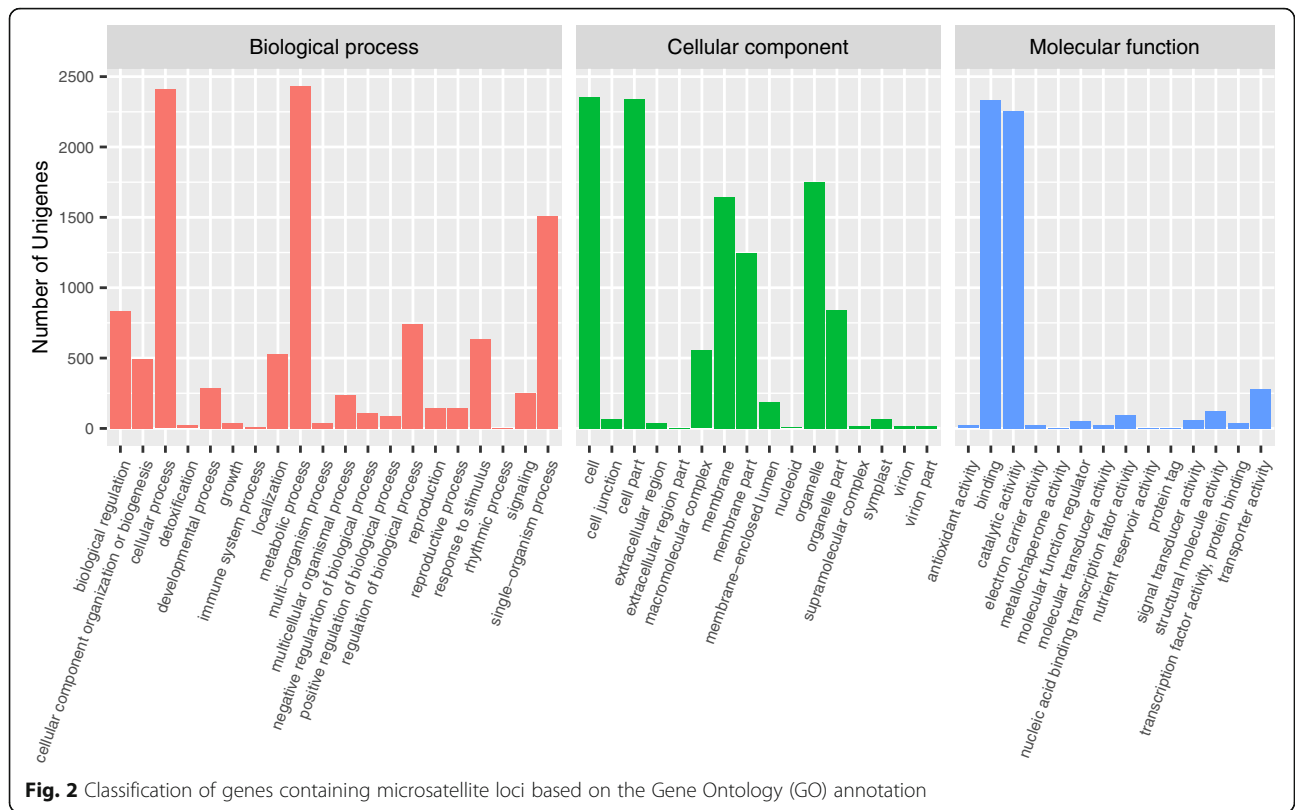
Functional annotation of genes containing microsatellites

Gene Ontology (GO) assignment was used to classify the transcripts according to their function. Of the 9563 contigs in *L. kiusiana*, 6560 (68.6%) had blast hits against “land plants (taxa: 3193)” of the nonredundant (nr) database. In total, 4510 (47.2%) microsatellite-containing transcripts had GO annotations and were categorized into three functional groups and 50 subgroups (Fig. 2). *Cellular* (2401) and *metabolic processes* (2429) were the top two subgroups that involved the most genes in the “biological process” group. In the “cellular component” group, the *cell* (2259) and *cell parts* (2246) were the top two subgroups that involved the most genes. Two subgroups, *catalytic activity* (2241) and *binding* (2239), involved the most genes among the “molecular function” group. A suite of transcripts was annotated as transcription factors (TFs) that contain typical DNA binding motifs, such as AP2/ERF, basic leucine zipper (bZIP), MADS-box, MYB, and WRKY (Additional file 1: Table S3). The functions of microsatellite-containing transcripts were further surveyed by KEGG pathway analysis. The results showed that the transcripts were involved in a total of 124 pathways (Additional file 1: Table S4). The KEGG pathways, including metabolism (92.9%), organismal systems (5.0%), genetic information processing (1.1%), and environmental information processing (1.0%), were grouped into 16 functional categories (Fig. 3). The most represented pathways included purine, pyrimidine and thiamine metabolism, biosynthesis of antibiotics, aminobenzoate degradation, the T-cell receptor signaling pathway, Th1 and Th2 cell differentiation, and starch and sucrose metabolism (Additional file 1: Table S4).

Identification of selection signatures in transcriptomic SSR markers

To identify transcriptomic SSR markers under selection, we calculated SNPs on transcripts and





investigated their effect on nucleotide substitution rates. A total of 67,498,600 reads were mapped onto the 9563 transcripts that contained microsatellites, generating 12,486 SNPs after quality control and filtration. The proportions of transition substitutions were 30.5% for A/G and 29.9% for C/T, compared with smaller proportions of transversion for A/C (9.5%), A/T (13.1%), C/G (7.3%), and G/T (9.7%). Among all SNPs detected, 3087 were in the putative 2389 ORF regions, of which 1943 were nonsynonymous (62.9%) and 1144 were synonymous (37.1%). We estimated nonsynonymous (d_N) and synonymous (d_S) substitution rates of the ORFs that contained at least one SNP. The results showed that the d_N/d_S ratios of 39 ORFs were greater than one, indicating that the loci are putatively under positive selection (Fig. 4). Investigation of these ORFs may represent genes with interesting evolutionary histories that have led to high levels of accumulated polymorphism. Functional annotations for 39 ORFs showed that some of the ORFs included genes encoding Culmins, ABC transporters, serine/threonine-protein kinase SRK2E, and synaptotagmin-5 (Additional file 1: Table S5).

Validation of transcriptomic SSR markers

The primers of 6911 microsatellites were designed based on 9563 transcripts containing microsatellites. Apart from transcriptomic SSR markers under selection, a total of 50 pairs of primers were selected (except mononucleotide repeats). Ten primer pairs gave no amplicon, but 40 pairs were consistently and successfully amplified from 51 individuals of the Minamioguni (MG, Kumamoto, Japan) population. Of the 40 primer pairs, 38 produced an amplicon of the expected size, whereas the others produced amplicons larger than the expected size. Twenty-five microsatellite loci were polymorphic and were successfully genotyped in the 51 individuals of MG, generating a total of 174 alleles. The average number of alleles per SSR locus was 6.96 (Table 1). The estimated null allele frequency for *LKI20* was greater than 0.1 (Table 1), and MICRO-CHECKER also indicated a higher frequency of null alleles at the same locus. No significant deviation from Hardy-Weinberg equilibrium (HWE) was found except for one (*LKI20*), following the correction based on the false discovery rate ($p < 0.05$). A linkage disequilibrium (LD) test revealed that most of the loci were in linkage equilibrium except three pairs of loci

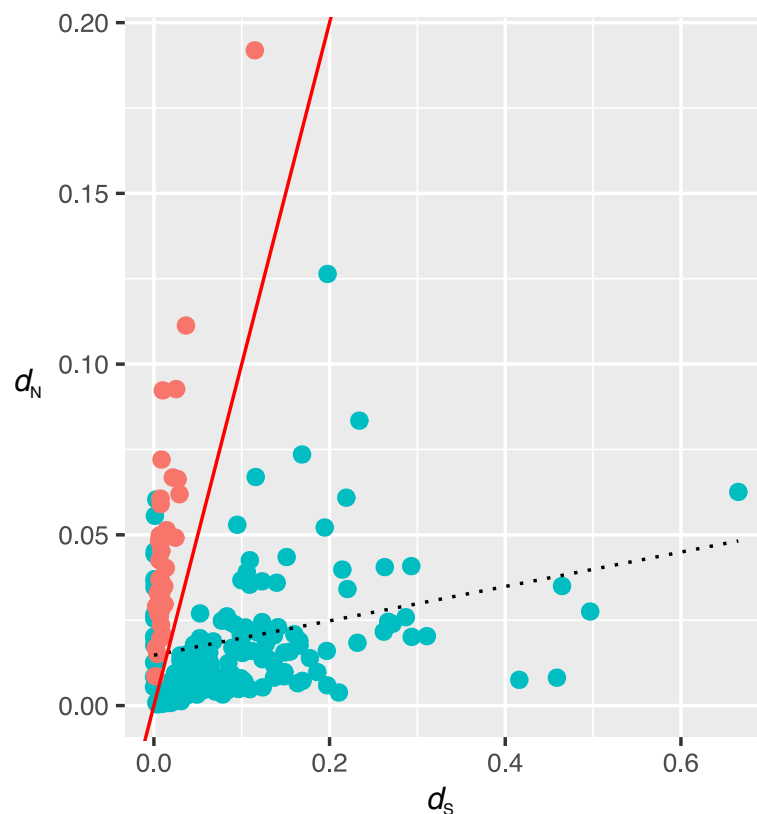


Fig. 4 Correlation of synonymous and nonsynonymous substitution rates of SNPs. The solid red line indicates the d_N/d_S ratio is equal to one. The dashed line represents the regression, which was analyzed using d_N and d_S for all unigenes. The red circles represent unigenes with d_N/d_S ratios > 1

Table 1 Summary statistics for 25 polymorphic microsatellite loci developed for *Lychnis kiusiana*. Number of alleles (N_A), observed (H_O) and expected (H_E) heterozygosities, inbreeding coefficient (F_{IS}), Hardy-Weinberg equilibrium (HWE), and polymorphism information content (PIC). Significant values after false discovery rate correction are indicated with asterisks (* $p < 0.05$, ** $p < 0.01$)

Locus	Size range (bp)	N_A	F_{null}	H_O	H_E	F_{IS}	HWE	PIC
LKI04	201–217	4	0.00000	0.37255	0.36905	-0.010	0.2921	0.341
LKI05	230–260	6	0.02708	0.68627	0.72821	0.058	0.0108*	0.672
LKI08	211–229	6	0.08018	0.52941	0.64221	0.177	0.0942	0.576
LKI09	119–173	15	0.03127	0.76471	0.77927	0.019	0.0253*	0.755
LKI10	146–176	7	0.00000	0.82353	0.73986	-0.114	0.3221	0.695
LKI12	124–157	10	0.00000	0.84314	0.80431	-0.049	0.8835	0.767
LKI15	198–240	8	0.01274	0.78431	0.82702	0.052	0.7256	0.798
LKI16	222–240	5	0.02202	0.68627	0.68744	0.002	0.1274	0.625
LKI17	183–219	12	0.00013	0.72549	0.75306	0.037	0.6855	0.718
LKI20	241–268	8	0.12785	0.54902	0.77053	0.290 [†]	0.0006**	0.742
LKI28	158–188	8	0.00000	0.82353	0.75791	-0.074	0.0502	0.725
LKI29	211–253	10	0.00000	0.90196	0.80897	-0.116	0.0756	0.777
LKI32	226–244	4	0.00003	0.64706	0.64007	-0.011	0.5823	0.572
LKI33	126–152	8	0.05123	0.64706	0.75908	0.149	0.0593	0.711
LKI34	181–208	9	0.05066	0.7451	0.82411	0.097	0.0215*	0.797
LKI35	185–197	5	0.00001	0.56863	0.56824	-0.046	0.6829	0.414
LKI38	151–169	6	0.00000	0.7451	0.70511	-0.057	0.6523	0.653
LKI39	227–236	4	0.00002	0.56863	0.56824	-0.001	0.3841	0.510
LKI41	195–210	5	0.00679	0.70588	0.70258	-0.005	0.5343	0.649
LKI42	127–130	2	0.05875	0.39216	0.48146	0.187	0.2406	0.363
LKI43	197–203	3	0.00001	0.13725	0.13143	-0.045	1.0000	0.126
LKI44	245–248	2	0.00005	0.5098	0.50476	-0.010	1.0000	0.375
LKI45	168–192	9	0.00000	0.84314	0.76199	-0.108	0.8721	0.718
LKI46	213–237	7	0.00000	0.60784	0.60435	-0.006	0.2555	0.559
LKI48	251–279	10	0.07888	0.62745	0.70879	0.116	0.0291*	0.656
Total		174						
Mean		6.96		0.6455	0.6617	0.023		0.6118

(LKI05 and LKI17, LKI16 and LKI17, and LKI20 and LKI34) (Additional file 2: Figure S4). Observed (H_O) and expected (H_E) heterozygosities ranged from 0.1373 to 0.9020 and 0.1314 to 0.8270, respectively, and locus-specific F_{IS} estimates ranged from -0.116 (LKI29) to 0.290 (LKI20). The positive F_{IS} value of LKI20 was significantly higher than zero (Wilcoxon signed-rank test; $p < 0.05$). The polymorphism information content (PIC) of 25 SSR markers ranged from 0.126 to 0.798. At the population level, the MG population exhibited high genetic diversity ($H_O = 0.6455$, $H_E = 0.6617$) and a low inbreeding coefficient ($F_{IS} = 0.023$). The mean value of the Garza-Williamson ($G-W$) index for the MG population was 0.8278. Twenty-one of the identified microsatellite markers had GO annotations (Additional file 1: Table S6). In particular, LKI09 which exhibited a high level of

polymorphism (Table 1) was associated with WRKY transcription factor 57 (WRKY57) (Fig. 5).

To examine cross-species transferability, 25 newly developed microsatellite markers were tested on the other *Lychnis* species (*L. cognate*, *L. fulgens*, and *L. wilfordii*). Eleven markers were successfully amplified in all three *Lychnis* species (44%; Table 2). Twenty and 16 loci were amplified in *L. fulgens* and *L. cognate*, respectively, while only 14 loci were amplified in *L. wilfordii* (Table 2).

Discussion

Deep transcriptome sequencing technologies are rapid and cost-effective tools to characterize gene content and identify polymorphic markers in nonmodel plants [6]. In this study, we generated a leaf transcriptome from the angiosperm *L. kiusiana* and characterized microsatellites in the transcriptome. The results showed that

Table 2 Amplification success of 25 microsatellites developed for *Lychnis kiusiana* across three other *Lychnis* species. Gray shading indicates a locus that was successfully amplified in all three *Lychnis* species

Locus	<i>Lychnis cognata</i>	<i>Lychnis wilfordii</i>	<i>Lychnis fulgens</i>
LKI04	○	○	○
LKI05			
LKI08	○	○	○
LKI09	○	○	○
LKI10			
LKI12	○	○	○
LKI15	○		○
LKI16		○	○
LKI17	○	○	○
LKI20	○	○	○
LKI28			○
LKI29	○	○	○
LKI32			
LKI33	○	○	○
LKI34	○		○
LKI35	○		
LKI38			○
LKI39	○	○	○
LKI41		○	○
LKI42	○		○
LKI43			○
LKI44	○	○	○
LKI45	○	○	○
LKI46	○		
LKI48		○	○

stresses, leaf senescence and seed dormancy, germination and developmental processes [30]. In addition to WRKY TFs, AP2/ERF or MYB TFs also regulate various biological processes, including development, defense, and biotic and abiotic stress responses [31–34]. Many transcripts associated with the AP2/ERF or MYB TF family were identified in the *L. kiusiana* leaf transcriptome (Additional file 1: Table S3). These microsatellites related to TFs may function as an important “tuning knobs” for the expression of genes [35, 36]. Examination of microsatellite polymorphisms in *Lychnis* transcripts linked to TFs requires further study to provide valuable insights into understanding environmental stresses.

For genetic diversity study, including structure analysis, it is critical to identify loci affected by selection in order to exclude them. We applied outlier tests based on

detecting SNP loci under selection. Examination of the rate of variation in the ORFs that contained at least one SNP revealed the acceleration of the evolution of some genes, probably to adapt to the extreme environmental stress in wetlands of mountain areas. For example, the d_N/d_S ratios of two transcripts, Cullins and ABC transporters, were higher than one (Additional file 1: Table S5). Cullins are a family of hydrophobic proteins that provide a scaffold for ubiquitin ligases (E3). E3 is a member of a ubiquitination system that operates in conjunction with an E1 (ubiquitin activating enzyme; UBA) and an E2 (ubiquitin conjugating enzyme; UBC) and functions in regulating abiotic stress (e.g., drought, salinity, cold and nutrient deprivation) responses [37]. In plants, ATP-binding cassette transporters (ABC transporters) are involved in essential aspects of a terrestrial

plant's lifestyle [38]. ABC transporters also play an important role in organ growth, plant nutrition, plant development, response to abiotic stress, and the interaction of the plant with its environment [39].

The genetic variability of microsatellites is extensively exploited in evolutionary studies of a wide variety of plant species [40]. Genetic diversity and fitness play an important role in species conservation, especially in rare plants and small populations [41, 42]. *Lychnis kiusiana* is an endangered perennial herb native to wetland areas, and the sizes and numbers of populations of this species are decreasing [17]. A previous study showed high genetic diversity (mean $H_E = 0.791$ and mean $N_A = 12.0$) within the seven *L. kiusiana* populations based on five genomic SSR markers [17]. Although these novel genomic SSR markers are useful for investigating genetic diversity and molecular-assisted breeding, these markers are still insufficient for genetic population study. Herein, we identified applicable microsatellites from the *L. kiusiana* leaf transcriptome, and 25 transcriptomic SSR markers have been proven to be efficient genetic markers. Our results showed that the genetic diversity of the MG population with the 25 transcriptomic SSR markers was higher ($N_A = 6.96$ and $H_E = 0.6617$), although the five genomic SSRs revealed higher genetic variations ($N_A = 14.2$ and $H_E = 0.817$) than the transcriptomic SSR in the MG population. This is because genomic SSR markers generally have a higher level of polymorphism compared with transcriptomic SSR markers [7]. Compared with other transcriptomic SSR markers [43–46], the developed *L. kiusiana* transcriptomic SSR markers have a high level of genetic diversity. Full understanding of the genetic diversity and structure of *L. kiusiana* requires evaluation of genetic variation among the other Korean and Japanese populations. SSR neutrality tests in the populations of *L. kiusiana* are also needed to identify either neutral or outlier loci among the 25 newly developed markers.

Given the high transferability of transcriptomic SSR markers, the microsatellites identified from *L. kiusiana* will have wide application in other *Lychnis* species. We applied 25 newly developed transcriptomic SSR markers to three species of the genus *Lychnis* to evaluate the transferability of these markers. The results showed that the transferability ratios were 56% in *L. wilfordii*, 64% in *L. cognate*, and 80% in *L. fulgens*, consistent with the estimation (46.8–100%) for a genus [7]. The high transferability of *L. kiusiana* transcriptomic SSRs to *L. fulgens* may be due to a high similarity of the sequences flanking the SSR between two species. In particular, the 14 amplified SSR markers in *L. wilfordii* will be a valuable source for its population genetic analyses because *L. wilfordii* is an endangered species in Korea.

Conclusions

In this study, we characterized a large number of transcriptome-derived microsatellites from *Lychnis kiusiana*, and functional annotation of microsatellite-containing transcripts provides new insights into the evolution of microsatellites. Several microsatellites in the *L. kiusiana* leaf transcriptome were linked to the genes involved in plant response to biotic and abiotic stresses. The microsatellites located in these genes may be crucial for *L. kiusiana* to evaluate adaptive genetic variations to adapt to wetlands in mountain areas. Our results showed that the identified microsatellite loci that have high allele numbers and heterozygosity could provide genetic markers in *L. kiusiana* populations. Furthermore, many microsatellites were transferable across the other species of the genus.

Methods

Sample collection, RNA extraction and sequencing

Leaf samples of *Lychnis kiusiana* were collected from a single individual in the Korea National Arboretum, South Korea. Total RNA was extracted from fresh leaf tissue (100 mg) using the methods of Ghawana et al. [47] and was treated with DNase I (Invitrogen, California, USA). The *Lychnis* RNA (29.7 μ g) was sequenced on the Illumina HiSeq 2000 platform at LabGenomics (Seongnam, South Korea), generating 6 Gb of 100 bp of paired-end reads.

Transcriptome assembly and microsatellite identification

RNA sequence reads were assembled with Trinity v2.2.0 [48] on a 12-core 3.33-GHz Linux Workstation with 192 GB memory. These assembled contigs were used to detect microsatellite loci using the MlCroSatellite identification tool (MISA) [49] with the following criteria: mono-nucleotide repeat motifs with at least 12 repeats, di-nucleotide repeat motifs with at least six repeats, tri- and tetra-nucleotide repeat motifs with at least five motifs, and penta- and hexa-nucleotide repeat motifs with at least four repeats. The criterion for compound microsatellites was that the interval between two repeat motifs was shorter than 100 nt. The assembled contigs were deposited in the Dryad Digital Repository (doi:<https://doi.org/10.5061/dryad.rc47tt4>).

To investigate the distribution of microsatellites in the *L. kiusiana* leaf transcriptome, candidate coding regions within transcript sequences were identified using TransDecoder v3.0.1 (<http://transdecoder.github.io/>). The location of microsatellite was determined based on the predicted CDS region, 5' UTR and 3' UTR. Characteristics of motif type were determined and compared with each other among microsatellite loci located in the CDS region, 5' UTR and 3' UTR.

Functional annotation

To understand the possible function of microsatellites, all the transcripts harboring a microsatellite were searched against the GenBank nr protein database using BLASTx with an e -value cut-off of 10^{-5} using Blast2GO v4.1.9 [50]. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were used.

SNP identification and estimation of synonymous and nonsynonymous sites

To identify putative SNPs in the transcripts containing SSRs, all reads were mapped onto the assembled transcripts using Bowtie v2.2.9 [51] with default parameters. The SAMtools “mpileup” utility and BCFtools [52] were used to identify SNPs. The SNPs with a read depth ≥ 10 and mapping quality ≥ 20 were retained. To detect transcriptomic SSRs under directional selection, nonsynonymous (d_N) and synonymous (d_S) rates of ORFs that contained at least one SNP were estimated by KaKs_Calculator v2.0 [53]. A d_N/d_S ratio greater than one suggests positive selection and less than one indicates purifying selection [54]. We estimated that ORFs with $d_N/d_S > 1$ and $p < 0.05$ (Fisher’s exact test) were under positive selection.

Microsatellite detection and genotyping

Previously extracted genomic DNA [17] for 51 individuals collected from the population of Minamioguni (MG, Kumamoto, Japan) was used. To explore amplification and polymorphism of the microsatellite loci identified in the *L. kiusiana* leaf transcriptome, the primers were designed based on the sequences flanking the microsatellite loci using Primer3 [55]. Apart from transcriptomic SSR markers under selection, 50 primer pairs were selected from the transcripts with a large number of repeat sizes and then used to amplify the microsatellite loci. Polymerase chain reaction (PCR) was performed in a 25- μ l reaction volume containing 10 ng genomic DNA, 0.2 mM of each dNTP, 0.25 μ M of each primer, 5 μ l 10 \times *h-Taq* reaction buffer, and 0.25 units of *h-Taq* polymerase (Solgent Co., Daejeon, South Korea). PCR was performed under the following conditions: an initial denaturation at 95 °C for 15 min followed by 30 cycles each of 95 °C for 20 s, 58 °C or 60 °C for 40 s, and 72 °C for 30 s, and a final extension at 72 °C for 5 min. Electrophoresis of the amplified fragments was performed on 2% agarose gel to confirm the success of amplification and polymorphism. Forward or reverse primers selected for the population genetic analyses (Table S7) were labeled with 6-FAM, HEX, or TAMRA fluorescent dyes (Macrogen Inc.; Seoul, South Korea). For fragment analyses, PCR products amplified with fluorescent primers were genotyped on an ABI 3730xl DNA Analyzer

(Applied Biosystems, California, USA) at Macrogen Inc. (Seoul, South Korea).

Microsatellite polymorphisms of *L. kiusiana* were calculated based on the average number of alleles per locus (N_A), observed (H_O) and expected (H_E) heterozygosities and inbreeding coefficient (F_{IS}) [56] in FSTAT 2.9.3 [57] and in GENEPOP v4.2 [58]. Departure from genotypic proportions expected under Hardy-Weinberg equilibrium (HWE) was examined by Fisher’s exact test using permutations [59] implemented in GENEPOP. The linkage disequilibrium between loci was also detected using Fisher’s exact test under the Markov chain algorithm implemented in GENEPOP. PIC was calculated using CER-VUS 3.0 [60]. Critical significance levels were adjusted for multiple comparisons using the false discovery rate [61]. The statistical analysis was conducted using R v.3.3.3 [62].

Cross-species amplification

Three species of the genus *Lychnis* (*L. cognate*, *L. fulgens*, and *L. wilfordii*) were chosen to evaluate the transferability of these newly developed microsatellite markers to other related species. Total genomic DNA was isolated from herbarium specimens using the methods of Allen et al. [63]. PCR was performed as described above with the primer pairs. Electrophoresis of PCR products was performed on 2% agarose gel to check the success of amplification. The experiments were repeated three times with the same controls to confirm reproducibility and consistency.

Additional files

Additional file 1: Table S1. General information for the microsatellite analysis. **Table S2.** Distribution and characteristics of the microsatellites in different transcript regions. **Table S3.** Results of the KEGG pathway analysis. **Table S4.** Blast results of 9563 transcripts that contain SSRs in *Lychnis* leaf transcriptome. **Table S5.** Blast results of the 39 ORFs showing positive selection ($d_N/d_S > 1$). **Table S6.** Blast results of the 25 newly developed transcriptomic SSR markers. **Table S7.** NCBI accession numbers, primer sequences and characterization of the 25 microsatellite loci developed for *Lychnis kiusiana*. (DOCX 1498 kb)

Additional file 2: Figure S1. The distributions of the major repeat types in the *Lychnis kiusiana* leaf transcriptome. **Figure S2.** Box plots of the sizes of different repeat motifs. **Figure S3.** Box plots of the sizes of different repeat motifs in different genic regions. **Figure S4.** Linkage disequilibrium (LD) map for the 25 transcriptomic SSR markers. (PDF 556 kb)

Abbreviations

d_N : Number of substitutions per nonsynonymous site; d_S : Number of substitutions per synonymous site; F_{IS} : the inbreeding coefficient within individuals relative to the subpopulation; GO: Gene Ontology; H_E : expected heterozygosity; H_O : observed heterozygosity; HWE: Hardy-Weinberg equilibrium; KEGG: Kyoto Encyclopedia of Genes and Genomes; LD: Linkage disequilibrium; ORF: Open reading frame; PIC: Polymorphism information content; SNP: Single-nucleotide polymorphism; SSR: Simple sequence repeat

Acknowledgements

The authors thank Han-Gyu Bae for assistance with measures of genetic diversity and anonymous reviewers for valuable comments on an earlier version of the manuscript.

Funding

This research was supported by the “Constructure of Infrastructure on Conservation and Restoration of Rare and Endemic Species” Program through the Korea National Arboretum (KNA1–2-10, 10–1). The funding body had no role in the design, collection, analysis, or interpretation of the data or in the writing of the manuscript or the decision to submit the manuscript for publication.

Availability of data and materials

The data sets supporting the results of this article are included in additional files. The assembled contigs generated and analyzed during the current study are available in the Dryad Digital Repository [doi:<https://doi.org/10.5061/dryad.rc47tt4>].

Authors' contributions

SP contributed to the design of the project, assembled, finished, and annotated the transcriptome, performed all analyses, prepared the figures and tables, and drafted the manuscript; SS contributed to the design of the project, assisted in collecting *Lychnis kiusiana* for RNA isolation, and read/edited the manuscript; MS performed the experiments and read/edited the manuscript; NF and TH provided *Lychnis kiusiana* DNA samples from the MG population and read/edited the manuscript; and SJP contributed to the design of the project and read/edited the manuscript. All authors read and approved the final draft of the manuscript.

Ethics approval and consent to participate

The *Lychnis kiusiana* leaf material used for RNA-seq was collected from a single individual in the Korea National Arboretum, South Korea. Experimental studies on the plant, including collection of plant material, comply with institutional, national, or international guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Institute of Natural Science, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea. ²Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea. ³Plant Conservation Division, Korea National Arboretum, Pocheon, Gyeonggi 11186, South Korea. ⁴Department of Biological Science, Graduate School of Science and Technology, Kumamoto University, Chuo-ku, Kumamoto 860-8555, Japan. ⁵Faculty of Biosphere-Geosphere Science, Okayama University of Science, Kita-ku, Okayama 700-0005, Japan.

Received: 22 October 2017 Accepted: 27 December 2018

Published online: 08 January 2019

References

- Powell W, Machray G, Provan J. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* 1996;1:215–22.
- Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* 1984;12:4127–38.
- Slate J, Pemberton JM. Comparing molecular measures for detecting inbreeding depression. *J Evol Biol.* 2002;15:20–31.
- Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 2005;23:48–55.
- Eklom R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity.* 2011;107:1–15.
- Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am J Bot.* 2012;99:193–208.
- Ellis JR, Burke JM. EST-SSRs as a resource for population genetic analyses. *Heredity.* 2007;99:125–32.
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol.* 2002;11:2453–65.
- Kashi Y, King DG. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 2006;22:253–9.
- Galeuchet DJ, Perret C, Fischer M. Microsatellite variation and structure of 28 populations of the common wetland plant, *Lychnis flos-cuculi* L., in a fragmented landscape. *Mol Ecol.* 2005;14:991–1000.
- Bowman G, Perret C, Hoehn S, Galeuchet DJ, Fischer M. Habitat fragmentation and adaptation: a reciprocal replant-transplant experiment among 15 populations of *Lychnis flos-cuculi*. *J Ecol.* 2008;96:1056–64.
- Aavik T, Holderegger R, Bolliger J. The structural and functional connectivity of the grassland plant *Lychnis flos-cuculi*. *Heredity.* 2014;112:471–8.
- Hauser TP, Loeschcke V. Drought stress and inbreeding depression in *Lychnis flos-cuculi* (Caryophyllaceae). *Evolution.* 1996;50:1119–26.
- Mustajärvi K, Siikamäki P, Akerberg A. Inbreeding depression in perennial *Lychnis viscaria* (Caryophyllaceae): effects of population mating history and nutrient availability. *Am J Bot.* 2005;92:1853–61.
- Bernasconi G, Antonovics J, Biere A, Charlesworth D, Delph LF, Filatov D, Giraud T, Hood ME, Marais GA, McCauley D, Pannell JR, Shykoff JA, Vyskot B, Wolfe LM, Widmer A. *Silene* as a model system in ecology and evolution. *Heredity.* 2009;103:5–14.
- Akiyama S, Silene L. In: Iwatsuki K, Boufford DE, Ohba H, editors. *Flora of Japan vol. Ila, Angiospermae, Dicotyledonae, Archichlamydeae(a)*. Tokyo: Kodansha; 2006. p. 202–10.
- Yamasaki T, Ozeki K, Fujii N, Takehara M, Yokogawa M, Kaneko S, Isagi Y. Genetic diversity and structure of *Silene kiusiana* (Caryophyllaceae) in the Aso region, Kyushu, Japan, revealed by novel nuclear microsatellite markers. *Acta Phytotax Geobot.* 2013;63:107–20.
- Young A, Boyle T, Brown T. The population genetic consequences of habitat fragmentation for plants. *Trends Ecol Evol.* 1996;11:413–8.
- Honnay O, Jacquemyn H. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conserv Biol.* 2007;21:823–31.
- Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y, Lobo J. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Mol Ecol.* 2008;17:5177–88.
- Jacquemyn H, Meester LD, Jongejans E, Honnay O. Evolutionary changes in plant reproductive traits following habitat fragmentation and their consequences for population fitness. *J Ecol.* 2012;100:76–87.
- Sloan DB, Keller SR, Berardi AE, Sanderson BJ, Karpovich JF, Taylor DR. De novo transcriptome assembly and polymorphism detection in the flowering plant *Silene vulgaris* (Caryophyllaceae). *Mol Ecol Resour.* 2012;12:333–43.
- Kantety RV, La Rota M, Matthews DE, Sorrells ME. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol.* 2002;48:501–10.
- Li YC, Korol AB, Fahima T, Nevo E. Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol.* 2004;21:991–1007.
- Pryde DC, Dalvie D, Hu Q, Jones P, Obach RS, Tran TD. Aldehyde oxidase: an enzyme of emerging importance in drug discovery. *J Med Chem.* 2010;53:8441–60.
- Puinean AM, Foster SP, Oliphant L, Denholm I, Field LM, Millar NS, Williamson MS, Bass C. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLoS Genet.* 2010;6:e1000999.
- Irving HR, Gehring CA, Parish RW. Changes in cytosolic pH and calcium of guard cells precede stomatal movements. *Proc Natl Acad Sci U S A.* 1992;89:1790–4.
- Doebley J, Lukens L. Transcriptional regulators and the evolution of plant form. *Plant Cell.* 1998;10:1075–82.
- Lawson MJ, Zhang L. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biol.* 2006;7:R14.
- Rushton PJ, Somssich IE, Ringler P, Shen QJ. WRKY transcription factors. *Trends Plant Sci.* 2010;15:247–58.
- Licausi F, Giorgi FM, Zenoni S, Osti F, Pezzotti M, Perata P. Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. *BMC Genomics.* 2010;11:719.

32. Dossa K, Wei X, Li D, Fonceca D, Zhang Y, Wang L, Yu J, Boshou L, Diouf D, Cissé N, Zhang X. Insight into the AP2/ERF transcription factor superfamily in sesame and expression profiling of DREB subfamily under drought stress. *BMC Plant Biol* 2016;16:171.
33. Shu Y, Liu Y, Zhang J, Song L, Guo C. Genome-wide analysis of the AP2/ERF superfamily genes and their responses to abiotic stress in *Medicago truncatula*. *Front Plant Sci*. 2016;6:1247.
34. Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Plants*. 2013; 19:307–21.
35. King DG, Soller M, Kashi Y. Evolutionary tuning knobs. *Endeavour*. 1997;21: 36–40.
36. Trifonov EN. Tuning function of tandemly repeating sequences: a molecular device for fast adaptation. In: Wasser SP, editor. *Evolutionary theory and processes: modern horizons papers in honour of Eviatar Nevo*. Massachusetts: Kluwer Academic Publishers; 2004. p. 115–38.
37. Stone SL. The role of ubiquitin and the 26S proteasome in plant abiotic stress signaling. *Front Plant Sci*. 2014;5:135.
38. Hwang JU, Song WY, Hong D, Ko D, Yamaoka Y, Jang S, Yim S, Lee E, Khare D, Kim K, Palmgren M, Yoon HS, Martinoia E, Lee Y. Plant ABC transporters enable many unique aspects of a terrestrial plant's lifestyle. *Mol Plant*. 2016; 9(3):338–55.
39. Kang J, Park J, Choi H, Burla B, Kretschmar T, Lee Y, Martinoia E. *Arabidopsis* Book. 2011;9:e0153.
40. Hodel RG, Gitzendanner MA, Germain-Aubrey CC, Liu X, Crowl AA, Sun M, Landis JB, Segovia-Salcedo MC, Douglas NA, Chen S, Soltis DE, Soltis PS. A new resource for the development of SSR markers: millions of loci from a thousand plant transcriptomes. *Appl Plant Sci*. 2016;4:1600024.
41. Ellstrand NC, Elam DR. Population genetic consequences of small population size: implications for plant conservation. *Annu Rev Ecol Syst*. 1993;24:217–42.
42. Leimu R, Mutikainen P, Koricheva J, Fischer M. How general are positive relationships between plant population size, fitness and genetic variation? *J Ecol*. 2006;94:942–52.
43. Mathithumilan B, Kadam NN, Biradar J, Reddy SH, Ankaiah M, Narayanan MJ, Makarla U, Khurana P, Sreeman SM. Development and characterization of microsatellite markers for *Morus spp.* and assessment of their transferability to other closely related species. *BMC Plant Biol*. 2013;13:194.
44. Jia H, Yang H, Sun P, Li J, Zhang J, Guo Y, Han X, Zhang G, Lu M, Hua J. De novo transcriptome assembly, development of EST-SSR markers and population genetic analyses for the desert biomass willow, *Salix psammophila*. *Sci Rep*. 2016;6:39591.
45. Liu H, Tan W, Sun H, Liu Y, Meng K, Liao W. Development and characterization of EST-SSR markers for *Artocarpus hypargyreus* (Moraceae). *Appl Plant Sci*. 2016;4:1600113.
46. Zhang Y, Zhang X, Wang YH, Shen SK. De Novo assembly of transcriptome and Development of novel EST-SSR markers in *Rhododendron rex* Lévl. Through Illumina sequencing. *Front Plant Sci*. 2017;8:1664.
47. Ghawana S, Paul A, Kumar H, Kumar A, Singh H, Bhardwaj PK, Rani A, Singh RS, Raizada J, Singh K, Kumar S. An RNA isolation system for plant tissues rich in secondary metabolites. *BMC Res Notes*. 2011;4:85.
48. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52.
49. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet*. 2003;106:411–22.
50. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36:3420–35.
51. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9:357–9.
52. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 genome project data processing subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
53. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*. 2010;8:77–80.
54. Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet*. 2005; 36:197–218.
55. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012;40:e115.
56. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984;38:1358–70.
57. Goudet J. FSTAT, version 2.9.3, A program to estimate and test gene diversities and fixation indices. <https://www2.unil.ch/popgen/softwares/fstat.htm> (2002). Accessed 5 Feb 2002.
58. Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Hered*. 1995;86:248–9.
59. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. 1992;48:361–72.
60. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol*. 2007;16:1099–106.
61. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57:289–300.
62. R Development Core Team. R: A Language and Environment for statistical computing. Vienna, Austria: the R Foundation for Statistical Computing; 2017.
63. Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat Protoc*. 2006;1:2320–5.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

