Genetic Variation and Association Mapping in the F2 Population of the *Perilla* Crop (*Perilla Frutescens* L.) using new Developed *Perilla* SSR Markers

Ju Keon Kim
  Kangwon National University

Kyu Jin Sa
  Kangwon National University

Ye Ju Ha
  Kangwon National University

Ju Kyong Lee (✉ jukyonglee@kangwon.ac.kr)
  Kangwon National University   https://orcid.org/0000-0002-2769-0799

Research Article

Abstract

The transcriptome sequencing approach RNA-seq represents a powerful tool for transcriptional analysis and development of SSR markers for nonmodel crop. In the *Perilla* crop, analysis of the distribution of different repeat motifs showed that dinucleotide repeats were the most abundant (62.0%) type, followed by trinucleotide repeats (35.3%), with the two together comprising 97.3% of the eSSR repeats. In this study, we developed 39 new SSR primer sets by the transcriptome sequencing approach RNA-sEq. A total of 130 alleles were detected segregating in nine *Perilla* accessions with an average of 3.3 alleles per locus, ranging from 125 to 360 bp. The number of alleles per locus ranged from two to six. To detect SSR markers associated with morphological characteristics of *Perilla* crop, a total of 40 individuals from the F_2 population of *Perilla* were selected for association analysis based on their leaf- and plant-related characteristics. In an association analysis of 37 SSR markers and 9 leaf- and plant-related traits in the 40 individuals of the F_2 population, we identified 12 and 11 SSR markers associated with leaf-related traits and plant-related traits, respectively. Therefore, the new *Perilla* SSR primers described in this study could be helpful in identifying genetic diversity and genetic mapping, designating important genes/QTLs for *Perilla* crop breeding programs, and allowing *Perilla* breeders to improve leaf and plant quality through marker-assisted selection (MAS) breeding programs.

Introduction

Perilla frutescens Britt. is a self-fertilizing species of Labiatae. *Perilla* species is widely distributed and cultivated in East Asia, especially in South Korea, Japan, and China. There are two varieties (or two cultivated types) based on their uses and morphology: *P. frutescens* var. *frutescens* and var. *crispa*. Generally in East Asia, var. *frutescens* is used as an oil crop (dlggae in Korean, egoma in Japanese, and ren in Chinese), and var. *crispa* is used as a vegetable crop or herbal medicine crop (jaso in Korean, shiso in Japanese, and zisu in Chinese) (Lee and Ohnishi 2001, 2003; Lee et al. 2002; Nitta and Ohnishi 1999; Nitta et al. 2003). These two cultivated types of *Perilla* crop have a long history of cultivation and are very important as oil crop and vegetable or medicine crop in East Asia (Lee and Ohnishi 2001, 2003; Nitta et al. 2003). The two cultivated types of *Perilla* crop have different morphological characteristics, although var. *frutescens* and var. *crispa* can be crossed with each other by artificial pollination (Honda et al. 1990, 1994; Lim et al. 2019). For example, *Perilla frutescens* var. *frutescens* is taller, larger in seed size (above 2 mm), has either soft or hard seeds, has green leaves and stems, and has non-wrinkled leaves and a fragrance specific to the var. *frutescens*. Meanwhile, *Perilla frutescens* var. *crispa* is smaller in plant height and seed size (below 2 mm), has only hard seeds, has red or green coloration in the leaves and stem, and has wrinkly or non-wrinkly leaves and a fragrance specific to the var. *crispa* (Lee and Ohnishi 2001, 2003; Lee et al. 2002; Nitta et al. 2003).

Many taxonomic studies have been conducted to distinguish between the two cultivated types of *Perilla* crop using morphological characteristics and molecular markers. Their chromosome number is the same, 2n = 40 (Honda et al. 1994; Yamane 1950). A cross experiment between the two cultivated types by Nagai (1935), Honda et al. (1990, 1994) and Lim et al. (2019) showed that they are cross-fertile with each other by artificial pollination. Furthermore, studies by Koezuka et al. (1985, 1986), Honda et al. (1990), and Lee and Ohnishi (2001, 2003) failed to distinguish the two cultivated types of *Perilla* crop because of the existence of intermediate types. That is, essential oils, anthocyanins, and the color and hardness of the seeds cannot be used as key characteristics. It is still difficult to distinguish the two cultivated types of *Perilla* crop. In addition, weedy plants of *Perilla* crop were first reported in the two cultivated types of *P. frutescens* by Lee and Ohnishi (2001, 2003), Nitta and Ohnishi

(1999), and Nitta et al. (2003). They reported that analyses of random amplification of polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) markers showed that weedy plants could be grouped into two types: one belonging to the var. *frutescens* group and another belonging to the var. *crispa* group. However, the analysis does not provide concrete evidence of the origin of the weedy plants of the *Perilla* crop. Therefore, further research is needed to distinguish between cultivated and weedy types of *Perilla* crop.

Understanding the molecular genetic mechanism of phenotypic variation remains a great challenge because most complicated traits, such as yield, quality and resistance, are controlled by quantitative trait loci (QTLs) (i.e. quantitative and qualitative traits) (Mazzcucato et al. 2008). In crop breeding studies, genetic mapping of functional loci for quantitative and qualitative related traits facilitates marker-assisted breeding. Two of the most commonly used tools for analyzing complex traits are linkage analysis and association analysis or association mapping. That is, linkage disequilibrium (LD)-based methods are used to detect associations between phenotypic variations and genetic polymorphisms. Thus, association mapping analysis complements traditional linkage analysis in understanding the genetic basis of complex traits and offers several advantages over linkage analysis, such as much higher mapping resolution, large number of alleles, wide reference population, and less research time in establishing an association (Buckler and Thornsberry 2002; Darvasi and Shifman 2005; Flint-Garcia et al. 2003; Mackay 2001). Furthermore, Shi et al. (2011) stated that using breeding populations may be a more practical approach for crop variety development, considering that markers linked to major QTLs can immediately be utilized in marker-assisted selection (MAS). Therefore, a thorough understanding of genetic diversity, population structure, and family relationships in a given panel is necessary for successful association analysis.

There are two methods to identify the genomic regions associated with important agricultural traits: 1) QTL mapping based on the results of crosses between two parents with linked and contrasting phenotypes and genotypes within segregated populations (Skot et al. 2005), and 2) association mapping using linkage disequilibrium (LD) between molecular markers and agronomic traits of interest (Pritchard et al. 2000, Zhang et al. 2012). Many population structures have been used for crop QTL detection and mapping. For example, backcross populations, F_2 populations, doubled haploid populations, recombinant inbred line populations, and near-isogenic line populations have all proven useful in identifying and confirming QTLs using various molecular marker systems (Choi et al. 2019; Farre et al. 2016; Park et al. 2013; Ramekar et al. 2018; Sa et al. 2015, Tanksley and Nelson 1996). However, for *Perilla* crop, QTL analysis using genetic maps of segregated populations is very difficult because various molecular markers, such as simple sequence repeat (SSR) markers for each chromosome, have not yet been developed for *Perilla* crop, in contrast with other major crops, such as rice, wheat and maize. In addition, more than 500 seeds of the F_2 population are produced in F_1 plant. For this reason, genetic analysis of the entire F_2 population is not possible (Lim et al. 2019). Therefore, in this study, some lines of the F_2 population were selected based on leaf characteristics such as leaf color (green or purple) and leaf shape (non-wrinkled or wrinkled) (see Fig. 1) to identify morphological differences between the two cultivated types of *Perilla* crop using molecular markers.

Meanwhile, molecular marker-based techniques in genetic research have been used in many crop species to identify the genomic regions associated with important agricultural traits. Among the various DNA molecular marker systems, SSRs are codominant markers, which are considered to be one of the most suitable means for assessing genetic diversity, genetic relationships, population structure, QTLs, and association mapping because of their reliability, reproducibility, and discrimination (Park et al. 2009; Pejic et al. 1998; Sa et al. 2018; Vathana et al. 2019). Thus, SSRs are widely used in many major crop species, such as rice (Pradhan et al. 2016), maize (Kim et

al. 2017), wheat (Crossa et al. 2007), and soybean (Qin et al. 2016), for assessing population structure and association analysis.

In *Perilla* crop, although SSR markers have recently been developed by several researchers, such as Kwon et al. (2005), Park et al. (2008), and Sa et al. (2018, 2019), the research is still insufficient. In our previous study, we sequenced and assembled one cultivated type (PF98095) of *P. frutescens* var. *frutescens* using transcriptome sequencing by RNA-seq (Tong et al. 2015), and we also obtained 15,991 SSR loci (Sa et al. 2018). The transcriptome sequencing approach RNA-seq represents a powerful tool for transcriptional analysis, novel gene discovery, and the development of SSR markers for nonmodel crops (Mutz et al. 2012, Wang et al. 2015, Yang et al. 2018). In particular, SSR markers developed from RNA-seq can enable marker-assisted selection for crop breeding programs because these regions are coding sequences and may be close to or within functionally transcribed genes (Mutz et al. 2012, Yang et al. 2018). In this study, we successfully developed SSR primers from *Perilla* crop, and these novel additional SSR markers can be used to analyze the genetic diversity, genetic relationships, population structure, QTL mapping, and association analysis in *Perilla* species.

Our study reports the results of the new *Perilla* SSR primers developed by the transcriptome sequencing approach RNA-seq and the results of association analysis in the F_2 population using these novel additional SSR primers and previously developed SSR primers. The results of this study are expected to provide useful information for future *Perilla* crop breeding programs.

Materials And Methods

Plant materials and morphological characteristics of the F_2 population

A total of 40 individuals from the F_2 population of the *Perilla* crop were selected from a cross between PF13-110 (female parent) and PF13-160 (male parent) (Fig. 1, Table 1). The female parent, PF13-110, is a var. *frutescens* that has non-wrinkled leaves with a green leaf surface and reverse side, green stem, and a fragrance specific to the var. *frutescens* (Fig. 1a, Table 1). The male parent, PF13-160, is a var. *crispa* that has wrinkled leaves with a purple leaf surface and reverse side, purple stem, and a fragrance specific to the var. *crispa* (Fig. 1b, Table 1).

Table 1
Morphological characteristics of the two parental lines used for crossing between *Perilla frutescens* var. *frutescens* (PF13-110, ⚥, Parent A) and var. *crispa* (PF13-160, ⚥, Parent B)

Morphological trait	Parent A	Parent A
QL1 (leaf size)	Large	Small
QL2 (color of leaf surface)	Green	Purple
QL3 (color of reverse side leaf)	Green	Purple
QL4 (degree of pubescence)	Heavily pubescent	Slightly pubescent
QL5 (number of leaf teeth)	Many	Few
QL6 (flower color)	White	Purple
QL7 (Stem color)	Green	Purple
QL8 (leaf shape)	Non-wrinkled	Wrinkled
QL9 (plant fragrance)	Var. *frutescens*	Var. *crispa*

To detect SSR markers associated with morphological characteristics, 40 individuals of the F_2 population were selected based on leaf characteristics such as leaf color, leaf size, and leaf shape and plant characteristics such as pubescence degree, plant fragrance, and flower and stem colors (Tables 1, 2). Of the total individuals of the F_2 population used in this study, approximately 20 individuals generally showed typical var. *frutescens* characteristics in leaf color and leaf shape, and the other 20 individuals showed a typical shape of var. *crispa* (Fig. 1, Table 2). To evaluate the morphological variation among the 40 individuals of the F_2 population and two parental lines of *Perilla* crop, each line was grown in a greenhouse at the College of Agriculture and Life Sciences, Kangwon National University, Chuncheon, Gangwon-do, Korea. We examined nine leaf- and plant-related traits, as shown in Tables 1 and 2, that were selected based on a previous report by Lee and Ohnishi (2001).

Table 2
Morphological characteristics in two parental lines and 40 individuals of the F_2 population developed from cross between *Perilla frutescens* var. *frutescens* and var. *crispa*

F_2 population line	QL1	QL2	QL3	QL4	QL5	QL6	QL7	QL8	QL9
1	Large	Green	Green	S/P	Few	White	Green	None	Var. *frutescens*
2	Small	Green	Green	S/P	Medium	Pink	Green	Wrinkle	Var. *crispa*
3	Medium	Green	Green	S/P	Few	White	G/P	Wrinkle	Var. *frutescens*
4	Small	Purple	Purple	Medium	Medium	Pink	Purple	Medium	Var. *crispa*
5	Large	Green	G/P	H/P	Medium	White	Green	Medium	Var. *crispa*
6	Large	G/P	G/P	Medium	Medium	Pink	G/P	Medium	Var. *crispa*
7	Large	G/P	G/P	Medium	Few	Pink	G/P	Medium	Var. *crispa*
8	Medium	Green	Green	S/P	Medium	Pink	G/P	Wrinkle	Var. *crispa*
9	Medium	Green	G/P	S/P	Medium	White	Purple	Medium	Var. *crispa*
10	Small	Green	Green	S/P	Few	Pink	Green	None	Var. *frutescens*
11	Large	Green	Green	S/P	Few	Purple	Purple	Medium	Var. *crispa*
12	Medium	G/P	G/P	Medium	Few	Purple	Purple	None	Var. *crispa*
13	Small	Green	Green	S/P	Medium	White	Purple	Wrinkle	Var. *crispa*
14	Medium	Green	Green	Medium	Medium	White	Green	None	Var. *crispa*
15	Medium	G/P	G/P	S/P	Few	Pink	G/P	None	Var. *frutescens*
16	Small	Green	G/P	Medium	Few	White	Purple	None	Var. *frutescens*
17	Small	G/P	G/P	Medium	Medium	Pink	Purple	None	Medium
18	Medium	G/P	G/P	Medium	Medium	Purple	G/P	Medium	Var. *crispa*
19	Small	Green	Green	H/P	Few	White	G/P	None	Var. *crispa*
20	Medium	Purple	Purple	S/P	Medium	Pink	Purple	Medium	Var. *crispa*
21	Medium	Green	Green	H/P	Many	Pink	Purple	Wrinkle	Var. *crispa*
22	Medium	Green	G/P	S/P	Medium	White	Purple	Medium	Var. *frutescens*

* Leaf color: G/P (Green/Purple)

** Pubescence: S/P (Slightly Pubescent), H/P (Heavily Pubescent)

F$_2$ population line	QL1	QL2	QL3	QL4	QL5	QL6	QL7	QL8	QL9
23	Small	Green	Green	H/P	Many	White	G/P	Medium	Var. *frutescens*
24	Medium	Green	Green	H/P	Medium	White	G/P	Wrinkle	None
25	Small	Green	Green	S/P	Many	White	Green	Medium	Var. *frutescens*
26	Large	Purple	G/P	S/P	Many	Purple	G/P	Medium	Var. *crispa*
27	Large	Green	Green	S/P	Many	Pink	G/P	Wrinkle	Var. *crispa*
28	Large	Green	Green	Medium	Many	White	G/P	Wrinkle	Var. *crispa*
29	Medium	Green	Green	S/P	Many	White	G/P	Wrinkle	None
30	Small.	Green	G/P	Medium	Many	Pink	G/P	Wrinkle	None
31	Large	Green	Green	S/P	Many	White	G/P	Wrinkle	Var. *frutescens*
32	Large	Purple	G/P	H/P	Medium	White	G/P	Wrinkle	Var. *crispa*
33	Medium	Green	Green	H/P	Medium	White	G/P	None	None
34	Small	Green	Green	H/P	Many	White	Green	Wrinkle	Var. *crispa*
35	Small	Purple	Purple	Medium	Many	Pink	Purple	Medium	None
36	Small	Green	Green	Medium	Many	White	Green	Medium	None
37	Large	Purple	Purple	S/P	Medium	Pink	G/P	None	Var. *crispa*
38	Large	Green	Green	Medium	Medium	White	Green	Medium	Var. *frutescens*
39	Small	Green	G/P	H/P	Medium	Pink	G/P	Medium	Var. *frutescens*
40	Medium	Green	G/P	S/P	Many	White	Purple	Wrinkle	Var. *crispa*
* Leaf color: G/P (Green/Purple)									
** Pubescence: S/P (Slightly Pubescent), H/P (Heavily Pubescent)									

DNA extraction and SSR primer development

Total DNA was extracted from young leaf tissues following the Plant DNAzol reagent protocols (GibcoBRL Inc., Grand Island, NY, USA). For construction of the transcriptome reference set in a previous study (Tong et al. 2015), *de novo* assembly of the PF98095 RNA-seq data was performed using Trinity software (http://TrinityRNASeq.sourceforge.net). The raw reads from next-generation sequencing (NGS) with a Phred quality score of at least 20 and a read length of at least 50 bp of HiSeq 2000 data were filtered before assembly. A Perl script MISA tool (http://pgrc.ipk-gatersleben.de/misa) was used to search microsatellite sites in the assembled transcriptome sequences of PF98095. The SSRs with dinucleotide, trinucleotide, and tetranucleotide

repeat units were identified. Based on the SSR flanking sequences, PRIMER 3 software was employed to design the primer pairs. Then, we searched all unigenes in the cultivated type of var. *frutescens* (PF98095). In a previous study, Sa et al. (2018) detected a total of 15,991 SSR loci. All SSRs can be classified according to the number of repeating units: dinucleotide SSRs had the highest number of SSRs (9,910), followed by trinucleotide SSRs (5,652) and tetranucleotide SSRs (429). In this study, we designed 148 SSR primer sets based on the di-, tri-, and tetranucleotide types and the number of repeat units.

SSR analysis and silver-staining

SSR amplifications were conducted in a total volume of 20 µl consisting of 20 ng genomic DNA, 1× PCR buffer, 0.5 µM of forward and reverse primers, 0.2 mM dNTPs, and 1 unit of *Taq* polymerase (Biotools, Madrid, Spain). The PCR profile consisted of an initial denaturation at 95 ºC for 3 minutes followed by 36 cycles of 95 ºC for 30 seconds, 55 ºC for 30 seconds, and 72 ºC for 1 minute 30 seconds with a final extension step of 5 minutes at 72 ºC. After PCR, 5 µl of the final product was mixed with 10 µl of electrophoresis loading buffer (98% formamide, 0.02% BPH, 0.02% xylene C, and 5 mM NaOH). After denaturing and immediate cooling, 2 µl from each sample was loaded onto a 6% denaturing (7.5 M urea) acrylamide-bis acrylamide gel (19:1) in 1× Tris-borate-EDTA (TBE) buffer and then electrophoresed at 1,800 V and 60 W for 130 minutes. The separated fragments were then visualized using a silver staining kit (Promega, Madison, WI, USA).

Data analysis

TASSEL 3.0 (Bradbury et al. 2007) was used to evaluate marker-trait associations using a general linear model (GLM). The number of permutation runs was set to 10,000 to obtain a marker significance value of $P \leq 0.05$. Power Marker version 3.25 (Liu and Muse 2005) was applied to obtain information on the number of alleles, major allele frequency (MAF), genetic diversity (GD), and polymorphic information content (PIC). GD was calculated for each pair of accessions using the Dice similarity index (Dice 1945). To examine the clustering pattern of all individuals of the F_2 population by construction of a phylogenetic tree, Nei's distance was calculated and used for the construction of unrooted phylogeny using a neighbor-joining (NJ) method, bootstrapping of the data (10,000 permutations) was implemented in POWERMARKER, and MEGA X was used to visualize the tree (Nei 1973, Kumar et al. 2018).

Results

SSR identification and polymorphisms

In our study, we surveyed the 148 newly developed SSR primer sets using nine *Perilla* accessions (3 cultivated var. *frutescens*, 3 weedy var. *frutescens*, 3 weedy var. *crispa*). The results showed that the 39 new SSR primer sets exhibited good amplification patterns and polymorphisms among the nine *Perilla* accessions (Supplement Tables 1, 2, Supplement Fig. 1). However, the remaining SSR primer pairs exhibited a monomorphic band (73) or ambiguous band pattern (24) and poor or no amplification (12). The 39 new SSR primer sets were used to measure polymorphisms in terms of the number of alleles, MAF, GD, and PIC among the nine *Perilla* accessions (Supplement Table 2). A total of 130 alleles were detected segregating in the nine *Perilla* accessions with an average of 3.3 alleles per locus, ranging from 125 to 360 bp. The number of alleles per locus ranged from two (KNUPE88, KNUPE90, KNUPE94, KNUPE100, KNUPE113, KNUPE114, KNUPE117, KNUPE118, KNUPE120, KNUPE122, KNUPE124, KNUPE125) to six (KNUPE105, KNUPE166). The GD ranged from 0.198 (KNUPE100, KNUPE113, KNUPE118, KNUPE124, KNUPE125) to 0.815 (KNUPE116), with an average of 0.552. The average PIC

value was 0.469, ranging from 0.178 (KNUPE100, KNUPE113, KNUPE118, KNUPE124, KNUPE125) to 0.788 (KNUPE116). The MAF per locus varied from 0.222 (KNUPE116) to 0.889 (KNUPE100, KNUPE113, KNUPE118, KNUPE124, KNUPE125), with an average of 0.590 (Supplement Table 2).

Meanwhile, we surveyed the 39 newly developed SSR primer sets between two parental lines of the F_2 population of the *Perilla* crop. From the results, among the 39 newly developed SSR primer sets, we selected 12 SSR primer sets that exhibited good amplification patterns and polymorphisms between the two parental lines of the F_2 population of the *Perilla* crop (Tables 3, 4). In addition, we surveyed a total of 105 SSR primer sets, which were developed in previous research, and selected 25 SSR primer sets that showed polymorphism between the two parental lines of the F_2 population (Table 4). Therefore, in our study, a total of 37 SSR primer sets, including 12 new SSR primer sets and 25 previous SSR primer sets, were used to measure polymorphisms in terms of GD, PIC, MAF, and separation patterns of allele bands (SPAB) among 40 individuals of the F_2 population of the *Perilla* crop (Table 4). In the results, the GD ranged from 0.185 (KNUPE95) to 0.708 (KNUPE102), with an average of 0.627. The average PIC value was 0.564, ranging from 0.177 (KNUPE95) to 0.650 (KNUPE102). The MAF per locus varied from 0.325 (KNUPE102) to 0.900 (KNUPE95), with an average of 0.481. In the case of SPAB, six SSR primer sets (KNUPF74, KNUPF79, KNUPF88, KNUPF89, KNUPF95, KNUPF103) were skewed towards AA (parent A), and four SSR primer sets (KNUPF5, KNUPF107, KNUPF108, KNUPF109) were skewed towards BB (parent B). Therefore, these SSR primer sets are considered to deviate from the expected Mendelian segregation ratio of 1:2:1 (or AA: AB:BB) in the F_2 population because more than 50% of the allele bands amplified in the F_2 population appeared to be biased towards one of the two parents (A or B). The remaining SSR primer sets showed a tendency towards a Mendelian segregation ratio of 1:2:1 (or AA: AB: BB) in 40 individuals of the F_2 population, although this does not represent the entire F_2 population. Meanwhile, among the SSR primer sets used in the analysis, 21 SSR primer sets showed a null band pattern in 40 individuals of the F_2 population (Table 4).

Table 3
Information of the 37 SSR primer sets used in the study

SSR loci	Forward Primer(5'-3')	Reverse Primer(3'-5')	Repeat motif
KNUPF4	TTTCAAAAATCTTACCAACGCT	TTCGTTTTTGCATCTAATTATTCA	$(AT)_{10}$
KNUPF5	TCCATCTCATCTCATTCAAACA	ATGGATCGGAAATCTAAAAACA	$(AT)_{10}$
KNUPF11	TTGCAAGGTAAGATGATGATGA	TTGAGGATTGACAATGTTCGT	$(CA)_{10}$
KNUPF12	AATTCAATCTCGCCTCCATATT	TTCTGAATCTTGAAGCTTTGGT	$(CA)_{11}$
KNUPF15	CCACACGTAAACCTCATAAACC	TTATCTCTAAAGAAATCGGGCA	$(CT)_{16}$
KNUPF16	CCTGTATCTCTCCCCGATAAAT	TGGATTTAATGCAGTTGAGTTG	$(CT)_{22}$
KNUPF21	ATCTCATGAGTGATGGAGGTTC	AACCATGGCTATCATCTACAGG	$(TG)_{10}$
KNUPF23	TTGCAAGTTCTTGAATTGTGAC	CACTCCTTCCCTCCTCTTTAAT	$(TG)_{11}$
KNUPF25	GCTTAGTGTGAGGAATTATGTAGGA	ACTCAGCATGCTTGAATTCTC	$(AAG)_{12}$
KNUPF29	CTGGAAGTTTCAGAGGAAAATG	GTCTAATCCGAACGAGAATCTG	$(AGT)_{10}$
KNUPF30	AACTAGTATATATGGCCTGCAAAAA	GACCTCTATCTCCCACATCCTA	$(ATC)_{10}$
KNUPF39	TCACCTTCCCCTTCATTTATTA	AGGATCGAACAGAACAAACTGT	$(TCT)_{13}$
KNUPF49	CTAGGTGTGGGTGATTTTCAAT	AAACTACCTACCACCATTTCCC	$(AG)_{17}$
KNUPF53	GATTCATCATTCAGCTCTCTCC	ATGACCAATGGATTAAACAAGG	$(CT)_{17}$
KNUPF54	GCCATTTGGAGATGGAATG	ATTTCGAGACAAAGCAACAAT	$(GC)_5$
KNUPF59	AATCTCGATGCCTAACAACAGT	TTCCTTGTAAATCCAGCTAAGG	$(CAG)_7$
KNUPF60	GCAATGGACATCTGTGAGAGTA	AATTGTGGTAATCATAGGGCAG	$(CAG)_7$
KNUPF61	GGGATACCCAAATTTCTACCAT	TCATGAAAAATCCAAACATTCA	$(CAG)_7$
KNUPF72	TAATTTGAGGGATTCCTTTCCT	CGCCACCCTTACTACTTCATAC	$(TCGA)_4$
KNUPF73	CCATTCTTCAATTCGATCAACTA	TCTGCAAATCATCCAGTTAAAA	$(CTAG)_4$
KNUPF74	TTGACTGTACCAGAGCATCAAG	GGGTACACTCACAACTCTACCAA	$(AAAT)_6$
GBPFM111	ATCATGGATGAATCGCACTT	CATTCTCCAAATGTTACTCTATTT	$(ACACA)_8$
KWPE25	ACATTTAAGAGAGAGAGCAAG	ACGAACGGGCTTCAATCTT	$[(GT)_8(GA_{14})]$
KNUPF79	CCTTTTCCCCTTTTCTCTTTTA	CAACAATAGATCGCACTGAATG	$(AG)_{26}$

* New SSR primer sets

SSR loci	Forward Primer(5'-3')	Reverse Primer(3'-5')	Repeat motif
KNUPF81	TTAAGCAACCAATTGCAGGTA	GTTGTGCAAAATTTGGTGATTT	$(AAT)_7$
KNUPF82	AAACCAAGGAACTCGTCAACTA	CGCTTCGTCTTTATTGTGTGTA	$(AGA)_7$
*KNUPF88	ATTCTTTTGAGTTGTGGGAGAA	CGTCTTTTGTGTGTAGCGATAA	$(GTAC)_4$
*KNUPF89	ATATCCCATTTTCTGATGCAAG	CCTTTTTCTTGCCTACTCATCA	$(AGTG)_6$
*KNUPF95	AGGATCAGGAGATCAGATGTGT	GAATATTTAGATTTCGTTATTTCCCA	$(AT)_9$
*KNUPF102	AAGCTTATGCAACCTATGAAGG	ATGACTATCAGCAGATCTCGAA	$(AT)_9$
*KNUPF103	CAAAAGTACTTGTTCCTGCCTC	TGAAGCCCTCAAGCAACTATTA	$(CA)_9$
*KNUPF104	ATGCTTGTTTGATTGTCGTCTT	GCCAATCTTACAATGTCCAAAA	$(TG)_9$
*KNUPF105	TCCAAAACCCTGATTCTGTAAC	AATTTGATCCATGGGATCTTC	$(TG)_9$
*KNUPF107	TAAGGTTGCCGATTGTACTTTT	ACAGAATGGCTCACAAGGATAG	$(AG)_{25}$
*KNUPF108	CCTCCTCTAATACATGTTTCTGC	TGCAGCTTCTGTTATCTTGAAA	$(AG)_{22}$
*KNUPF109	CTGAAAACTTCTTGAATGAGGG	GGCCTATTTTTGTAGCTGTTTG	$(AG)_{19}$
*KNUPF111	CTAGGTGTGGGTGATTTTCAAT	AAACTACCTACCACCATTTCCC	$(AG)_{17}$
*KNUPF112	GACTTCTCCGAGCATATTCAAC	ATTGTTTCTGCTGAAATGTGTG	$(CT)_4$
* New SSR primer sets			

Table 4

Estimates of GD, PIC, MAF, and separation pattern of allele bands (SPAB) of 37 SSR primer sets in 40 individuals of the F_2 population.

Marker	GD	PIC	MAF	Separation of the F2 population			
				AA	AB	BB	null
KNUPF4	0.666	0.599	0.425	12	17	10	1
KNUPF5	0.629	0.565	0.500	2	13	20	5
KNUPF11	0.636	0.564	0.475	12	17	11	0
KNUPF12	0.624	0.553	0.500	9	20	11	0
KNUPF15	0.694	0.636	0.400	14	14	9	3
KNUPF16	0.674	0.622	0.475	5	20	11	4
KNUPF21	0.645	0.587	0.500	12	16	10	2
KNUPF23	0.654	0.580	0.425	11	17	11	1
KNUPF25	0.661	0.587	0.375	14	14	12	0
KNUPF29	0.679	0.618	0.425	10	17	11	2
KNUPF30	0.634	0.561	0.475	9	18	13	0
KNUPF39	0.670	0.610	0.450	9	19	11	1
KNUPF49	0.596	0.530	0.550	8	23	9	0
KNUPF53	0.686	0.625	0.400	11	16	12	1
KNUPF54	0.629	0.572	0.525	7	21	10	2
KNUPF60	0.669	0.608	0.450	11	18	10	0
KNUPF61	0.651	0.577	0.425	10	17	13	0
KNUPF72	0.539	0.481	0.625	8	25	7	0
KNUPF73	0.646	0.571	0.425	14	18	8	0
KNUPF74	0.656	0.596	0.475	21	9	8	2
KNUPF79	0.601	0.516	0.450	17	18	5	0
KNUPF81	0.676	0.613	0.375	15	15	8	2
KNUPF82	0.690	0.628	0.350	10	14	15	1
KNUPF88	0.656	0.587	0.425	17	14	8	1
KNUPF89	0.681	0.634	0.475	19	8	6	7
KNUPF95	0.185	0.177	0.900	36	2	2	0

*GD: Genetic Diversity, PIC: Polymorphism Information Content, MAF: Major Allele Frequency

Marker	GD	PIC	MAF	Separation of the F2 population			
				AA	AB	BB	null
KNUPF102	0.708	0.650	0.325	13	11	13	3
KNUPF103	0.514	0.438	0.625	25	12	3	0
KNUPF104	0.591	0.522	0.550	11	22	7	0
KNUPF105	0.635	0.573	0.500	6	20	13	1
KNUPF107	0.644	0.579	0.475	2	13	24	1
KNUPF108	0.619	0.568	0.550	9	6	22	3
KNUPF109	0.543	0.489	0.625	9	5	25	1
KNUPF111	0.645	0.572	0.450	10	18	12	0
KNUPF112	0.645	0.572	0.450	12	18	10	0
GBPFM111	0.651	0.583	0.450	8	19	12	1
KWPE25	0.586	0.506	0.525	14	20	6	0
Mean	0.627	0.564	0.481				

*GD: Genetic Diversity, PIC: Polymorphism Information Content, MAF: Major Allele Frequency

In addition, a phylogenetic tree was constructed using a total of 37 selected SSR markers to elucidate the phylogenetic relationship between the 40 individuals of the F_2 population and their two parental lines (Fig. 2). The 40 individuals of the F_2 population and two parental lines were clustered into two major groups: Group I included 18 individuals of the F_2 population, including Parent B (PF13-160); Group II included 22 individuals of the F_2 population, including Parent A (PF13-110). In the results, with the exception of exceptional individuals, Group I mainly contained many individuals with similar characteristics to Parent B, and Group II included many individuals with similar characteristics to Parent A (Table 2, Fig. 2).

Phenotypic variation and association analysis of leaf- and plant-related traits

Information on the distribution of leaf- and plant-related traits in 40 individuals of the F_2 population is shown in Table 5. In the case of QL1, among the 40 individuals of the F_2 population, 12 individuals showed large, 14 individuals showed medium, and 14 individuals showed small. For QL2, 28 individuals showed green, 6 individuals showed mixed green/purple, and 6 individuals showed purple. For QL3, 21 individuals showed green, 15 individuals showed mixed green/purple, and 4 individuals showed purple. For QL4, 9 individuals showed heavily pubescent, 13 individuals showed medium pubescent, and 18 individuals showed slightly pubescent. For QL5, 13 individuals showed many numbers, 18 individuals showed medium numbers, and 9 individuals showed few numbers. For QL6, 21 individuals were white, 15 individuals showed pink, and 4 individuals showed purple. For QL7, 9 individuals showed white, 19 individuals showed mixed green/purple, and 12 individuals showed purple. For QL8, 10 individuals showed non-wrinkles, 16 individuals showed medium wrinkles, and 14 individuals showed wrinkles. For QL9, 11 individuals showed specific to var. *frutescens*, 7 individuals showed medium, and 22 individuals showed specific to var. *crispa*. In addition, we performed a correlation analysis to detect the

relationship between leaf- and plant-related traits in the 40 individuals of the F_2 population (Table 6). The results showed significant variation among nine leaf- and plant-related traits (QL1, QL2, QL3, QL4, QL5, QL6, QL7, QL8, and QL9) in the 40 individuals of the F_2 population (Table 6). There were highly statistically significant positive correlations of QL2 with QL3 (0.682**) and QL6 (0.429**), of QL3 with QL6 (0.390*) and QL7 (0.462**), and of QL5 with QL8 (0.489**) and QL9 (0.313*) at P < 0.01 or P < 0.05.

Table 5

Distribution of morphological variation in 40 individuals of the F_2 population developed from cross between *Perilla frutescens* var. *frutescens* and var. *crispa*

Morphological character	Category	F_2 population (N = 40)
QL1 (leaf size)	Large (> 14cm^2)-1, medium (= 14cm^2)-2, small (< 14cm^2)-3	1- (12)* 2- (14) 3- (14)
QL2 (color of leaf surface)	Green-1, green/purple-2, purple-3	1- (28) 2- (6) 3- (6)
QL3 (color of reverse side leaf)	Green-1, green/purple-2, purple-3	1- (21) 2- (15) 3- (4)
QL4 (degree of pubescence)	Heavily pubescent-1, medium-2, slightly pubescent – 3	1- (9) 2- (13) 3- (18)
QL5 (number of leaf teeth)	Many (> 24 number)-1,medium-2, Few (< 17 number)-3	1- (13) 2- (18) 3- (9)
QL6 (flower color)	White-1, pink-2, purple-3	1- (21) 2- (15) 3- (4)
QL7 (stem color)	Green-1, green/purple-2, purple-3	1- (9) 2- (19) 3- (12)
QL8 (leaf wrinkle)	Non-wrinkle-1, medium-2, wrinkle – 3	1- (10) 2- (16) 3- (14)
QL9 (plant fragrance)	Specific to var. *frutescens*-1, non-specific – 2, specific to var. *crispa*-3	1- (11) 2- (7) 3- (22)
* The number of individuals in the F_2 population for each trait type.		

Table 6
Pearson correlation coefficients among 9 qualitative traits in F_2 population

	QL2	QL3	QL4	QL5	QL6	QL7	QL8	QL9
QL1	-0.151	-0.039	-0.175	-0.037	-0.100	0.144	-0.089	0.103
QL2		0.682**	0.047	0.126	0.429**	0.269	-0.165	0.191
QL3			0.005	-0.024	0.390*	0.462**	-0.262	0.095
QL4				-0.114	0.182	0.129	0.045	-0.243
QL5					-0.142	-0.073	0.489**	0.313*
QL6						0.244	-0.161	0.130
QL7							-0.014	0.150
QL8								0.167
** Significance at P < 0.01								
* Significance at P < 0.05								

Meanwhile, to select SSR markers associated with leaf- and plant-related characteristics in the 40 individuals of the F_2 population, genotypes of the 37 SSR markers and the nine leaf- and plant-related traits were used to confirm significant marker-trait associations (SMTAs) using TASSEL software (Table 7). In the results, we detected 24 SMTAs involving 21 SSR markers associated with nine leaf- and plant-related traits using GLM at a significance level of P ≤ 0.05 (Table 7). Among the 12 SSR markers for leaf-related traits, KWPF25 and KNUPF108 were associated with QL2, KNUPF11 was associated with QL3, KNUPF74 and KNUPF89 were associated with QL5, and seven SSR markers (KNUPF21, KNUPF49, KNUPF60, KNUPF81, KNUPF103, KNUPF109, GBPEM111) were associated with QL8. In the case of plant-related traits, 11 SSR markers were associated with plant-related traits. Among them, two SSR markers (KNUPF15 and KNUPF21) were associated with QL4, KNUPE11 was associated with QL6, four SSR markers (KNUPF11, KNUPF54, KNUPF82, and KNUPF88) were associated with QL7, and five SSR markers (KNUPF5, KNUPF29, KNUPF39, KNUPF61, and GBPEM111) were associated with QL9 (Table 7).

Table 7. List of significant markers detected with the Q GLM model

Trait	Marker	P value	Marker R^2
Leaf-related traits			
QL2 (Color of leaf surface)	KWPE25	0.02	0.19
	KNUPF108	0.02	0.22
QL3 (Color of reverse side leaf)	KNUPF11	0.02	0.20
QL5 (Number of leaf teeth)	KNUPF74	0.01	0.22
	KNUPF89	0.01	0.26
QL8 (Leaf wrinkle)	KNUPF21	0.04	0.16
	KNUPF49	0.05	0.15
	KNUPF60	0.02	0.21
	KNUPF81	0.03	0.18
	KNUPF103	0.01	0.21
	KNUPF109	0.04	0.17
	GBPFM111	0.04	0.16
Plant-related traits			
QL4 (Degree of pubescence)	KNUPF15	0.03	0.19
	KNUPF21	0.04	0.17
QL6 (Flower color)	KNUPF11	0.02	0.19
QL7 (Stem color)	KNUPF11	0.03	0.20
	KNUPF54	0.04	0.17
	KNUPF82	0.01	0.17
	KNUPF88	0.02	0.21
QL9 (Plant fragrance)	KNUPF5	0.03	0.20
	KNUPF29	0.03	0.19
	KNUPF39	0.04	0.17
	KNUPF61	0.00	0.33
	GBPFM111	0.00	0.32

Discussion

Genetic variation between individuals within a population or between populations of crop species because of either or both genetic and environmental influences can be easily assessed using various molecular markers. Among various molecular markers, SSR markers in particular have many advantages compared with other marker

systems as follows: the first advantage is high reproducibility, which is most important in genetic analysis; the second advantage of the SSR marker system is the polymorphic genetic information contents, and the hypervariable nature of SSRs produces very high allelic variations even among very closely related varieties; and the third advantage relates to the codominant nature of SSR polymorphisms, as the codominant nature of SSRs is suitable for genetic analysis in segregation of F_2 populations or pedigree analysis in hybrids (Park et al. 2009; Pejic et al. 1998; Sa et al. 2018).

In plants, the presence of SSRs was first demonstrated by the hybridization of poly (G-T) and poly (A-G) oligonucleotide probes on phage libraries of tropical tree genomes (Condit and Hubbell 1991). A search of published DNA sequences revealed that SSRs are also abundant in diverse plant genomes (Morgante and Olivieri 1993; Wang et al. 1994). Therefore, SSRs have become the preferred molecular marker system for the analysis of plant genetics and ecology. However, SSR markers can only be applied to crop species with large amounts of EST or cDNA sequences that are freely accessible to the public. Recently, the transcriptome sequencing approach RNAseq has become a powerful tool for novel gene discovery and the development of molecular markers for nonmodel crop, such as *Perilla* crop. In the case of the *Perilla* crop, in a previous study, we sequenced and assembled one cultivated type (PF98095) of *P. frutescens* var. *frutescens* using transcriptome sequencing by RNA-seq (Tong et al. 2015). This information will help develop SSR primers for *Perilla* crop. As explained in the Introduction, Sa et al. (2018) detected a total of 15,991 SSR loci based on the SSR flanking sequences via *de novo* assembly of the PF98095 RNA-seq data (Tong et al. 2015), which were classified based on the number of repeating units. Dinucleotide SSRs had the highest number of SSRs (9,910), followed by trinucleotide SSRs (5,652) and tetranucleotide SSRs (429) (Sa et al. 2018). In the *Perilla* crop, analysis of the distribution of different repeat motifs showed that dinucleotide repeats were the most abundant (62.0%) type, followed by trinucleotide repeats (35.3%), with the two together comprising 97.3% of the eSSR repeats. This finding is important for developing a large number of effective eSSR primers in *Perilla* crop, as the markers associated with smaller SSR motif lengths (di- and tri-) are more variable than other motif lengths (Ellegren 2004). Our results also showed that dinucleotide repeats were more abundant than other repeat units in the *Perilla* crop. While an almost equal distribution of di- and trinucleotide repeats was reported in ramie (Liu et al. 2013), a high abundance of trinucleotide repeats was reported in flax (Cloutier 2009), jute (Saha et al. 2017), and kenaf ESTs (Zhang et al. 2015).

In this study, a total of 130 alleles with 39 SSRs were detected segregating in the 9 *Perilla* accessions, which yielded an average of 3.3 alleles per locus (Supplement Tables 1, 2). Depending on the number of repeat units in the 39 SSR primer sets, the dinucleotide SSRs such as KNUPF87, KNUPF97, KNUPF103, KNUPF105, KNUPF110, and KNUPF116 showed comparatively more allelic bands and genetic diversity values than the tri- and tetranucleotide SSRs in *Perilla* accessions. In our study, although we analyzed only nine *Perilla* accessions, the allele number, GD, PIC, and MAF values for 39 new *Perilla* EST-SSR markers were found to determine the unique genetic profiles of individual genotypes of the *Perilla* crop. Therefore, these new SSR primer sets may be useful in the future for studying GD for *Perilla* germplasm resources by association mapping and designating important genes/QTLs for future genetic and breeding programs in *Perilla* crop.

In addition, analysis of genetic diversity and relationships and association mapping of plant breeding materials will help develop new molecular markers associated with target morphological traits or useful varieties and lines in *Perilla* crop breeding programs. In our study, we developed an F_2 population from a cross between PF13-110 (female parent, cultivated var. *frutescens*) and PF13-160 (male parent, weedy var. *crispa*) to develop molecular

markers for traits showing morphological differences between two cultivated types of var. *frutescens* and var. *crispa* through 40 individuals of the F_2 population of the *Perilla* crop (Fig. 1, Tables 1, 2). According to our results, the 40 individuals of the F_2 population showed discontinuous variations within the range of morphological characteristics of their parental lines, as shown in Tables 2 and 5. That is, some morphological traits investigated in our study showed a Mendelian segregation ratio of 1:2:1 (or AA:AB:BB), while other traits showed different separation patterns in 40 individuals of the F_2 population. For example, in the case of QL1, of the 40 lines of the F_2 population, 12 lines were large, 14 lines were medium, and 14 lines were small; whereas in the case of QL2, 28 lines were green, 6 lines were mixed green/purple, and 6 lines were purple. However, unfortunately, as mentioned in the Introduction, the F_2 population lines used in this study did not represent the entire population, so it was not possible to predict the exact segregation ratio of the F_2 population.

In our study, we also performed a correlation analysis for nine leaf- and plant-related traits in 40 individuals of the F_2 population (Table 6). The results showed that there were highly statistically significant positive correlations of QL2 with QL3 (0.682**) and QL6 (0.429**), of QL3 with QL6 (0.390*) and QL7 (0.462**), of QL5 with QL8 (0.489**) and QL9 (0.313*) at $P < 0.01$ or $P < 0.05$. Among the morphological traits examined in our study, leaf and plant color-related traits, such as QL2, QL3, QL6, and QL7, and leaf shape-related traits, such as QL5 and QL8, were closely related to each other. In addition, to better understand the genetic relationships among 40 individuals of the F_2 population, we analyzed the phylogenetic relationships of the 40 individuals of the F_2 population and their two parental lines using 37 SSR markers. In the phylogenetic tree analysis, the 40 individuals of the F_2 population and two parental lines were clustered into two major groups. Except for exceptional individuals, Group I mainly contained many individuals with similar characteristics to Parent B, while Group II contained many individuals with similar characteristics to Parent A (Table 2, Fig. 2). Therefore, these separated individuals and leaf- and plant-related traits of the F_2 population are considered to be useful genetic materials for finding molecular markers associated with the leaf and plant morphological characteristics of *Perilla* crop through association mapping analysis.

Association mapping has been proposed as a selection method to identify loci involved in the inheritance of complex traits (Risch and Merikangas 1996). In this study, we developed an F_2 population for association mapping analysis to find molecular markers associated with leaf- and plant-related traits of *Perilla* crop. Additionally, in our study, we used SSR markers for association mapping analysis because SSR markers are highly polymorphic and generally codominant (Park et al. 2009; Powell et al. 1996; Sa et al. 2018, 2019). Therefore, the SSR marker is considered to be a useful analysis method for identifying separation patterns of allelic bands in the F_2 population.

To select the SSR markers associated with leaf- and plant-related traits of the F_2 population, we analyzed marker-trait associations (SMTAs) between 37 SSR markers and 9 leaf- and plant-related characteristics in 40 individuals of the F_2 population using TASSEL software. From the results, we identified 12 and 11 SSR markers associated with leaf-related traits and plant-related traits, respectively (Table 7). In the case of leaf-related traits, 12 SSR markers, namely, KWPF25, KNUPF11, KNUPF21, KNUPF49, KNUPF60, KNUPF74, KNUPF81, KNUPF89, KNUPF103, KNUPF108, KNUPF109, and GBPEM111, were associated with QL2, QL3, QL5, and QL8. In addition, in the case of plant-related traits, 11 SSR markers, namely, KNUPF5, KNUPF11, KNUPF15, KNUPF21, KNUPF29, KNUPF39, KNUPF54, KNUPF61, KNUPF82, KNUPF88, and GBPEM111, were associated with QL4, QL6, QL7, and QL9 (Table 7). Our results showed that with the exception of QL1, most leaf- and plant-related traits were associated

with 1 to 7 SSR markers, depending on the leaf- and plant-related characteristics. Additionally, some SSR markers (KNUPF11, KNUPF21, GBPEM111) were associated with leaf- or plant-related characteristics, such as QL3, QL4, QL6, QL7, and QL9. For example, KNUPE11 was found to be associated with QL3, QL6, and QL7 traits; GBPEM111 was associated with QL8 and QL9 traits; and KNUPE21 was associated with QL4 and QL8 traits (Table 7).

In a previous study by Lim et al. (2000), bulk segregant analysis (BSA) was performed to identify SSR markers linked to leaf- and seed-related traits in *Perilla* crop. They reported that four SSR markers (KNUPF15, KNUPF21, KNUPF29, and KNUPF60) were associated with the color of the leaf surface, four SSR markers (KNUPF11, KNUPF15, KNUPF21, and KNUPF60) were associated with the color of the leaf reverse side, and three SSR markers (KNUPF25, KNUPF61, and GBPEM111) were associated with the depth of leaf teeth. In addition, five SSR markers (KNUPF11, KNUPF12, KNUPF16, KNUPF29, and KNUPF42) were associated with seed-related traits. According to our results, among the SSR markers associated with the leaf- and plant-related traits, eight SSR markers (KNUPE11, KNUPF15, KNUPF21, KWPE25, KNUPF29, KNUPE60, KNUPE61, GBPEM111) appeared to overlap with leaf-related traits previously reported by Lim et al. (2000). Therefore, these SSR markers are thought to be useful molecular markers for selecting leaf- and plant-related traits in *Perilla* crop. Additionally, these SSR markers are thought to be useful molecular markers for distinguishing between two cultivated types of *Perilla* crop. However, in *Perilla* crop, it is difficult to compare genetic characteristics because genomic information on SSR markers associated with leaf- and plant-related characteristics is still lacking.

In the future, if studies such as genome analysis for each chromosome are actively conducted in *Perilla* crop, it will be possible to analyze genomic information of specific *Perilla* SSR markers related to leaf- and plant-related characteristics. As explained in the Introduction, many taxonomic studies have been conducted to distinguish between two cultivated and weedy types of *Perilla* crop using morphological characteristics and molecular markers, but they failed to distinguish the two cultivated types of *Perilla* crop because of the existence of intermediate types, such as weedy plants. Therefore, the SSR markers associated with leaf- and plant-related characteristics selected in this study are expected to provide useful information for distinguishing the two cultivated types of *Perilla* crop and their weedy types in the future. In addition, the new *Perilla* SSR primers described in this study could be helpful in identifying genetic diversity and genetic mapping, designating important genes/QTLs for *Perilla* crop breeding programs, and allowing *Perilla* breeders to improve leaf and plant quality through MAS breeding programs.

Declarations

Author contributions J.K. Lee wrote the manuscript and designed the experiments. J.Y. Kim and K.J. Sa performed the experiment and analyzed the data, and Y.J. Ha helped to draft the manuscript. All authors commented on previous versions of the manuscript and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicting interests.

Availability of data and material All data generated or analyzed during this study are included in this published article and its supplementary information files.

Ethical approval This article does not contain any studies with human subjects or animals performed by any of the above authors.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

References

1. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635. https://doi.org/10.1093/bioinformatics/btm308

2. Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. Curr Opin Plant Biol 5:107-111. https://doi.org/10.1016/s1369-5266(02)00238-8.

3. Choi, J-K., Sa KJ, Park DH, Lim SE, Ryu S-H, Park JY, Park KJ, Rhee H-I, Lee M, Lee JK (2019) Construction of genetic linkage map and identification of QTLs related to agronomic traits in DH population of maize (*Zea mays* L.) using SSR markers. Genes Genom 41: 667-678. https://doi.org/10.1007/s13258-019-00813-x

4. Cloutier S, Niu ZX, Datla R, Duguid S (2009) Development and analysis of EST-SSRs for flax (*Linum usitatissimum* L.). Theor Appl Genet 119:53-63. https://doi.org/10.1007/s00122-009-1016-3

5. Condit R, Hubbell SP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34:66-71. https://doi.org/10.1139/g91-011

6. Crossa J, Burgueño J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J, Reynolds M, Crouch JH, Ortiz R (2007) Association Analysis of Historical Bread Wheat Germplasm Using Additive Genetic Covariance of Relatives and Population Structure. Genetics 177:1889-1913. https://doi.org/10.1534/genetics.107.078659

7. Davasi A, Shifman S (2005) The beauty of admixture. Nature Genet 37:118-119. https://doi.org/10.1038/ng0205-11

8. Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297-302. https://doi.org/10.2307/1932409

9. Ellegren H (2004) Microsatellites: simple sequences with complex evolution, Nat Rev Genet 5:435-445. https://doi.org/10.1038/nrg1348

10. Farré A, Sayers L, Leverington-Waite M, Goram R, Orford S, Wingen L, Mumford C, Griffiths S (2016) Application of a library of near isogenic lines to understand context dependent expression of QTL for grain

yield and adaptive traits in bread wheat. BMC Plant Biology 16:161. https://doi.org/10.1186/s12870-016-0849-6

11. Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Genet 54:357-374. https://doi.org/10.1146/annurev.arplant.54.031902.134907.

12. Honda G, Yuba A, Kojima T, Tabata M (1994) Chemotaxonomic and cytogenetic studies on *Perilla Frutescens* var. *citiodora* ('Lemon Egoma'). Natural Medicine 48:185-190.

13. Honda G, Koezuka Y, Tabata M (1990) Genetic studies of fruit color and hardness in *Perilla frutescens*. Japan J Breed 40:469-474.

14. Kim OG, Sa KJ, Lee JR, Lee JK (2017) Genetic analysis of maize germplasm in the Korean Genebank and association with agronomic traits and simple sequence repeat markers. Genes Genom 39:843-853. https://doi.org/10.1007/s13258-017-0547-2

15. Koezuka Y, Honda G, Sakamoto S, Tabata M (1985) Genetic control of anthocyanin production in *Perilla frutescens*. The Jap J pharm 39:228-231.

16. Koezuka Y, Honda G, Tabata M (1986) Genetic control of the chemical composition of volatile oils in *Perilla frutescens*. Phytochemistry 26:859-863.

17. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547-1549. https://doi.org/10.1093/molbev/msy096

18. Kwon SJ, Lee JK, Kim NS, Yu JW, Dixit A, Cho EG, Park YJ (2005) Isolation and characterization of SSR markers in *Perilla frutescens* Britt. Mol Eco Notes 5:454-456.

19. Lee JK, Nitta M, Kim NS, Park CH, Yoon KM, Shin YB, Ohnishi O (2002) Genetic diversity of *Perilla* and related weedy types in Korea determined by AFLP analyses. Crop Sci 42:2161-2166. https://doi.org/10.2135/cropci2002.2161

20. Lee JK, Ohnishi O (2001) Geographical differentiation of morphological characters among *Perilla* crop and their weedy types in East Asia. Breed Sci 51:247-255. https://doi.org/10.1270/jsbbs.51.247

21. Lee JK, Ohnishi O (2003) Genetic relationships among cultivated types of *Perilla Frutescens* and their weedy types in East Asia revealed by AFLP markers. Genet Resour Crop Evol 50:65-74. https://doi.org/10.1023/A:1022951002271

22. Lim SE, Sa KJ, Ha YJ, Lee JK (2019) Genetic analysis of F$_2$ population derived from the cross between *Perilla frutescens* var. *frutescens* and var. *crispa*. Korean J Breed Sci 51:184-189. https://doi.org/10.9787/KJBS.2019.51.3.184

23. Lim SE, Sa KJ, Ha YJ, Lee JK (2021) Bulk segregant analysis identifies SSR markers associated with leaf- and seed-related traits in *Perilla* crop (*Perilla frutescens* L.). Genes Genom online https://doi.org/10.1007/s13258-021-01056-5

24. Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128-2129. https://doi.org/10.1093/bioinformatics/bti282

25. Mackay TE (2001) The genetic architecture of quantitative traits. Annu Rev Genet 35:303-339. https://doi.org/10.1146/annurev.genet.35.102401.090633

26. Mazzcucato A, Papa R, Bitocchi E, Mosconi P, Nanni L, Negri V, Picarella ME, Siligato F, Soressi GP, Tiranti B, Veronesi F (2008) Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum Lycopersicum* L.) landraces. Theor Appl Genet 116:657-669. https://doi.org/10.1007/s00122-007-0699-6

27. Morgante M, Olivieri AM (1993) PCR-amplified microsatellites as markers in plant genetics. Plant J 3:175-182.

28. Mutz KO, Heilkenbrinker A, Lonne M, Walter JG, Stahl F (2012) Transcriptome analysis using next-generation sequencing. Curr Opin Biotech 24:22-30. https://doi.org/10.1016/j.copbio.2012.09.004

29. Nagai I (1935) On "Shiso" and "Egoma". Agriculture and Horticulture 10:2265-2273 (in Japanese).

30. Nei M (1973) Analysis of gene diversity in subdivided populations. PNAS 70:3321-3323. https://doi.org/10.1073/pnas.70.12.3321

31. Nitta M, Lee JK, Ohnishi O (2003) Asian *Perilla* crop and their weedy forms: their cultivation, utilization and genetic relationships. Econ Bot 57:245-253. https://doi.org/10.1663/0013-0001(2003)057[0245:APCATW]2.0.CO;2

32. Nitta M, Ohnishi O (1999) Genetic relationships among two *Perilla* crop, shiso and egoma, and the weedy type revealed by RAPD markers. Jpn J Genet 74:43-48.

33. Park KJ, Sa KJ, Koh H-J, Lee JK (2013) QTL analysis for eating quality-related traits in an $F_{2:3}$ population derived from waxy corn x sweet corn cross. Breed Sci 63:325-332. https://doi.org/10.1270/jsbbs.63.325

34. Park YJ, Dixit A, Ma KH, Lee JK, Lee MH, Chung CS, Nitta M, Okuno K, Kim TS, Cho EG, Rao VR (2008) Evaluation of genetic diversity and relationships within an on-farm collection of *Perilla frutescens* (L.) Britt. using SSR markers. Genet Resour Crop Evol 55:523-535. https://doi.org/10.1007/s10722-007-9258-x

35. Park YJ, Lee JK, Kim NS (2009) Simple Sequence Repeat Polymorphisms (SSRPs) for Evaluation of Molecular Diversity and Germplasm Classification of Minor Crop. Molecules 14:4546-4569. https://doi.org/10.3390/molecules14114546

36. Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSR, and AFLPs. Theor Appl Genet 97:1248-1255. https://doi.org/10.1007/s001220051017

37. Powell W, Morgante M, Andre C, Hanafey M, Vogelet J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225-238. https://doi.org/10.1007/BF00564200

38. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association Mapping in Structured Populations. Am J Hum Genet 67:170−181. https://doi.org/10.1086/302959

39. Pradhan SK, Barik SR, Sahoo A, Mohapatra S, Nayak DK, Mahender A, Meher J, Anandan A, Pandit E (2016) Population Structure, Genetic Diversity and Molecular Marker-Trait Association Analysis for High Temperature Stress Tolerance in Rice. PLoS ONE https://doi.org/10.1371/journal.pone.0160027

40. Qin J, Zhang J, Wang F, Yang C, Wang J, Zhang B, Wu C, Chen P, Zhang M (2016) Population Structure and Association Analysis of Traits in Summer Soybeans. J Crop Improv 30:572-594. https://doi.org/10.1080/15427528.2016.1200703

41. Ramekar RV, Sa KJ, Park K-C, Roy N, Kim N-S, Lee JK (2018) Construction of genetic linkage map and identification of QTLs related to agronomic traits in maize using DNA transposon-based markers. Breed Sci 68:465-473. https://doi.org/10.1270/jsbbs.18017

42. Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516-1517. https://doi.org/10.1126/science.273.5281.1516

43. Sa KJ, Choi I-K, Park K-C, Lee JK (2018) Genetic diversity and population structure among accessions of *Perilla frutescens* (L.) Britton in East Asia using new developed microsatellite markers. Genes Genom 40:1319-1329. https://doi.org/10.1007/s13258-018-0727-8

44. Sa KJ, Park JY, Woo SY, Ramekar RV, Jang CS, Lee JK (2015) Mapping of QTL traits in corn using a RIL population derived from a cross of dent corn x waxy corn. Genes Genom 37:1-14. https://doi.org/10.1007/s13258-014-0223-8

45. Sa KJ, Lim SE, Choi I-K, Park K-C, Lee JK (2019) Development and Characterization of New Microsatellite Markers for *Perilla frutescens* (L.) Britton. American J Plant Sci 10:1623-1630. https://doi.org/10.4236/ajps.2019.109115

46. Saha D, Rana RS, Chakraborty S, Datta S, Kumar AA, Chakraborty AK, Karmakar PG (2017) Development of a set of SSR markers for genetic polymorphism detection and interspecific hybrid jute breeding. The Crop J 5:416-429. https://doi.org/10.1016/j.cj.2017.02.006

47. Shi C, Navabi A, Yu K (2011) Association mapping of common bacterial blight resistance QTL in Ontario bean breeding populations. BMC Plant Biol 11, https://doi.org/10.1186/1471-2229-11-52

48. Skot L, Humphreys MO, Armstead I (2005) An association mapping approach to identify flowering time genes in natural populations of *Loliumperenne* (L.). Mol Breed 15:233-245. https://doi.org/10.1007/s11032-004-4824-9

49. Tanksley SD, Nelson JC (1996) Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor Appl Genet 92:191-203. https://doi.org/10.1007/BF00223376

50. Tong W, Kwon SJ, Lee JS, Choi IK, Park YJ, Choi SH, Sa KJ, Kim BW, Lee JK (2015) Gene set by *de novo* assembly of *Perilla* species and expression profiling between *P. frutescens* (L.) var. *frutescens* and var. *crispa*. Gene 559:155-163. https://doi.org/10.1016/j.gene.2015.01.028

51. Vathana V, Sa KJ, Lim SE, Lee JK (2019) Genetic Diversity and Association Analyses of Chinese Maize Inbred Lines Using SSR Markers. Plant Breed Biotech 7:186-199. https://doi.org/10.9787/PBB.2019.7.3.186

52. Wang L, Wang Z, Chen J, Liu C, Zhu W, Wang L, Meng L (2015) *De Novo* Transcriptome Assembly and Development of Novel Microsatellite Markers for the Traditional Chinese Medicinal Herb, *Veratrilla baillonii* Franch (Gentianaceae). Evol Bioinform online 11(S1):39-45. https://doi.org/10.4137/EBO.S20942.

53. Wang Z, Weber JL, Zhong G, Tanksley SD (1994) Survey of plant short tandem DNA repeats. Theor Appl Genet 88:1-6. https://doi.org/10.1007/BF00222386

54. Yamane Y (1950) Cytogenetic studies in *Perilla* and *Coleus*. I. Chromosome numbers. Jpn J Genet 25:220 (in Japanese).

55. Yang M, Han N, Li H, Meng L (2018) Transcriptome Analysis and Microsatellite Markers Development of a Traditional Chinese Medicinal Herb *Halenia elliptica* D. Don (Gentianaceae). Evol Bioinform 14:1-6. https://doi.org/10.1177/1176934318790263

56. Zhang LW, Wan XB, Xu JT, Lin LH, Qi JM (2015) *De novo* assembly of kenaf (*Hibiscus cannabinus*) transcriptome using illumine sequencing for gene discovery and marker identification. Mol Breed 35:192. https://doi.org/10.1007/s11032-015-0388-0

57. Zhang Q, Wu C, Ren F, Li Y, Zhang C (2012) Association analysis of important agronomical traits of maize inbred lines with SSRs. Aus J Crop Sci 6:1131-1138.
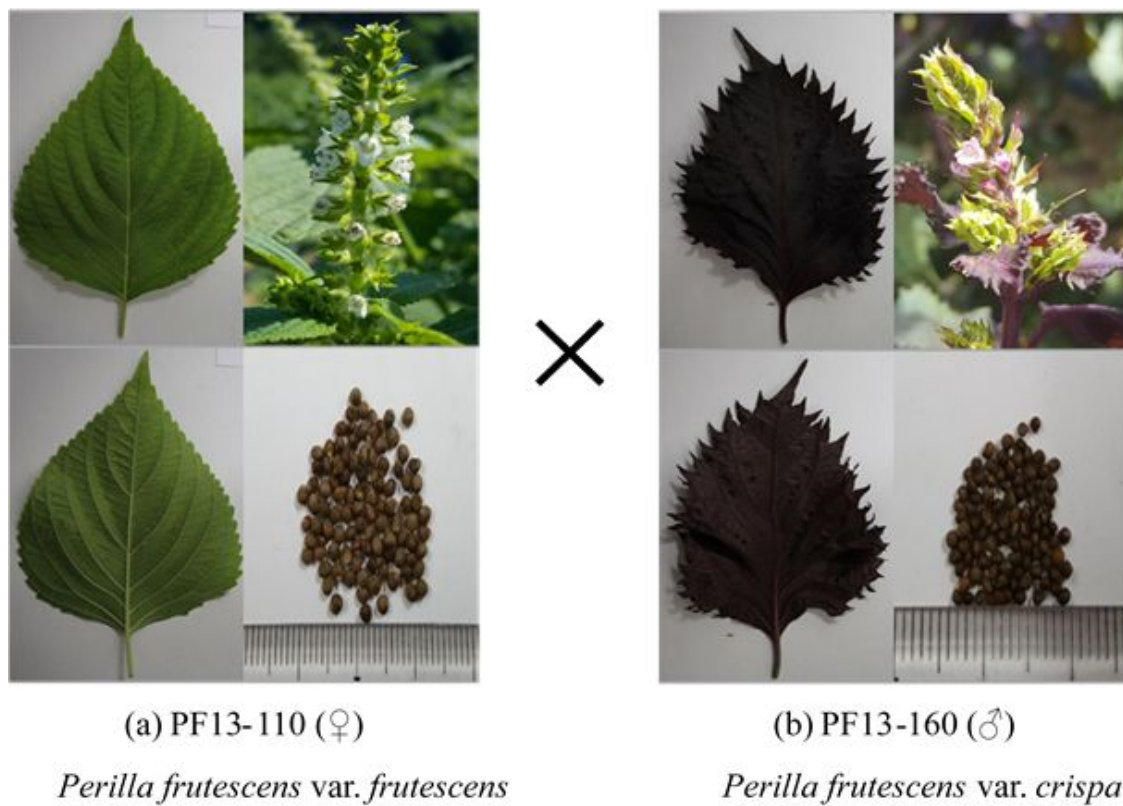
Figures

Figure 1

Morphological features of two parental lines of the F2 population used for crossing between (a) Perilla frutescens var. frutescens (parent A: PF13-110, ♀) and (b) P. frutescens var. crispa (parent B: PF13-160, ♂).
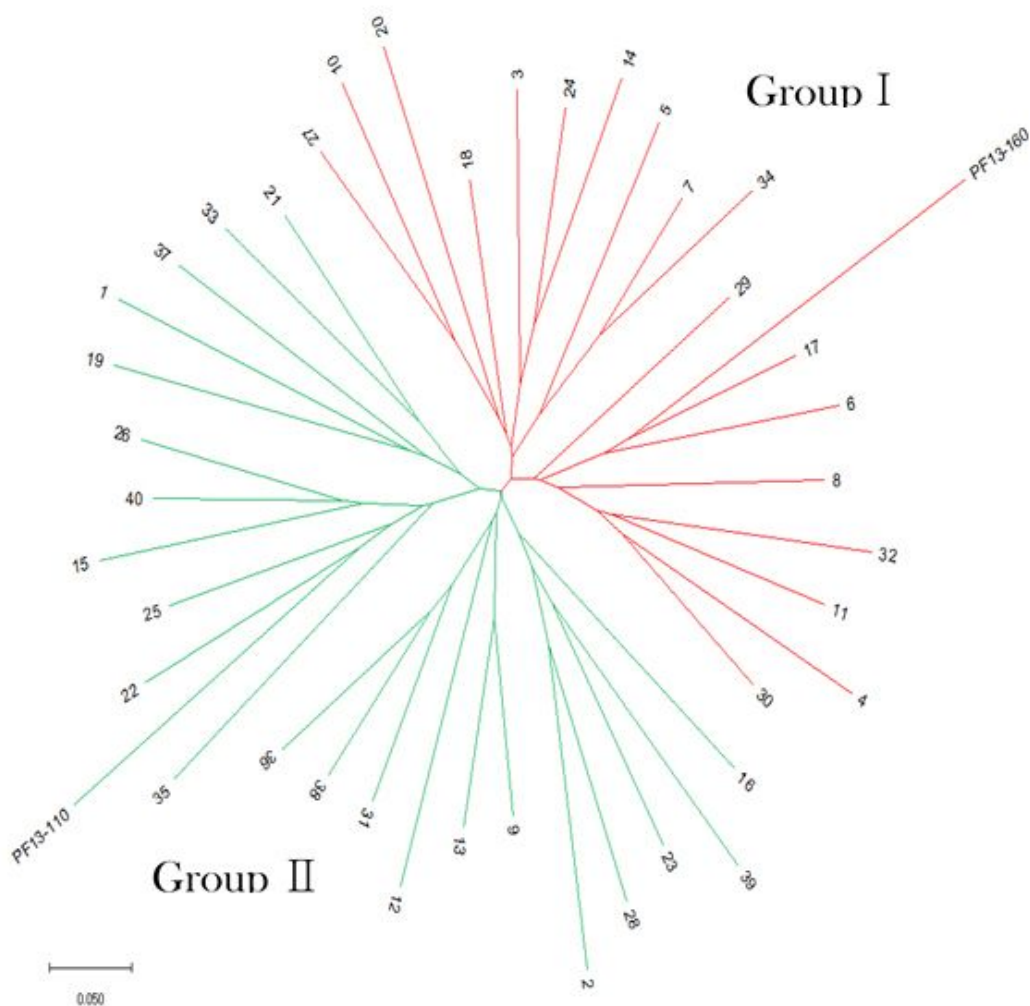
Figure 2

Neighbor-joining tree of all 40 individuals of the F2 population and two parental lines inferred from 39 SSR markers; Group I (red) represents F2 individuals, including Parent B (PF13-160, ▢), and Group II (green) represents F2 individuals, including Parent A (PF13-110, ▢).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementFig.1.docx
- SupplementTable1.docx
- SupplementTable2.docx