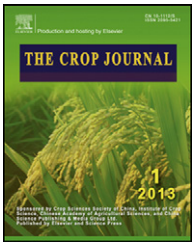
Available online at www.sciencedirect.com

ScienceDirect



Genome-wide association of 10 horticultural traits with expressed sequence tag-derived SNP markers in a collection of lettuce lines

Soonjae Kwon[a], *Ivan Simko*[b], *Barbara Hellier*[a], *Beiquan Mou*[b], *Jinguo Hu*[a,*]

[a]*US Department of Agriculture–Agricultural Research Service, Western Regional Plant Introduction Station, 59 Johnson Hall, Washington State University, Pullman, WA 99164, USA*

[b]*US Department of Agriculture–Agricultural Research Service, Crop Improvement Protection Research Unit, 1636 East Alisal Street, Salinas, CA 93905, USA*

ARTICLE INFO

Article history:
Received 24 April 2013
Received in revised form
27 June 2013
Accepted 30 July 2013
Available online 6 August 2013

Keywords:
LSGermOPA
Genome wide association study
SMTA
Significant marker trait association
Genetic diversity
Population structure

ABSTRACT

Genetic diversity, population structure, and genome-wide marker-trait association analyses were conducted on a special collection of 298 homozygous lettuce (*Lactuca sativa* L.) lines. Each of these lines was derived from a single plant that had been genotyped with 384 SNP markers using LSGermOPA. They included 122 butterhead, 53 romaine, 63 crisphead, 53 leaf and 7 stem types. Genetic diversity among these plants was assessed by pairwise comparison based on 322 high-quality SNP markers selected from 384 SNPs. Only 258 unique genotypes were identified among the 298 lines because 26 pairs or small groups (a total of 66 lines) shared identical genotypes. The average genetic similarity coefficient (GS) among these unique genotypes was 63.9% with a range of 40.6% to 99.8%. A phylogenetic tree was constructed based on the genotypic data. The most likely number of populations was estimated to be two or six. Association analysis between the 322 SNP markers and 10 phenotypic traits using the 258 homozygous lines was performed by three different methods: single factor analysis, general linear model analysis, and mixed linear model analysis. Nine significant marker-trait associations (SMTAs) were detected at $P < 0.0001$ with all three methods and also when considering kinship and/or population structure for this collection, with five SMTAs for seed coat color, one for leaf undulation, two for leaf anthocyanin, and one for stem anthocyanin. These markers will be useful in marker-assisted selection after further validation with segregating populations.

Production and hosting by Elsevier B.V. on behalf of Crop Science Society of China and Institute of Crop Science, CAAS.

Abbreviations: SNP, single nucleotide polymorphism; GWAS, genome wide association studies; MAS, marker assisted selection; LD, linkage disequilibrium; WRPIS, western region plant introduction station; UPGMA, unweighted pair group method with arithmetic mean; SFA, single factor analysis; GLM, general linear model; MLM, mixed linear model; GRIN, germplasm resources information network; SMTA, significant marker-trait association

* *Corresponding author.*
E-mail address: jinguo.hu@ars.usda.gov (J. Hu).

Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

Production and hosting by Elsevier

2214-5141/$ – see front matter. Production and hosting by Elsevier B.V. on behalf of Crop Science Society of China and Institute of Crop Science, CAAS.
http://dx.doi.org/10.1016/j.cj.2013.07.014

1. Introduction

Marker-assisted selection (MAS) has proven to be an effective tool in crop improvement. A prerequisite for successful MAS is to identify markers in close proximity to the genetic factors or genes controlling simple qualitative and complex quantitative traits of interest. Two approaches have been developed and applied to mapping genes in numerous plant species [1]: linkage mapping approach, which uses segregating populations derived from two parental lines, and association mapping that exploits biodiversity observed in germplasm collections of landraces, cultivars, and breeding lines [2]. The linkage mapping approach is limited to the variation between the two parents. Also, development of segregating populations may take several years if recombined inbred line populations are used for mapping [3,4]. The association mapping approach, which is based on linkage disequilibrium (LD), uses a collection of germplasm with a wide range of phenotypic and genetic variation [1]. Association mapping was initially developed to identify genes associated with human diseases, but was later applied to mapping genes in animal and plant populations [5–10]. In plants, LD-based association mapping started with the model plant *Arabidopsis* and was later extended to various crops such as rice (*Oryza sativa* L.) [11], grapevine (*Vitis vinifera* L.) [12], wheat (*Triticum aestivum* L.) [13], soybean (*Glycine max* (L.) Merr.) [14] and maize (*Zea mays* L.) [9,15]. In cultivated lettuce, association mapping has been used for mapping disease resistance genes [16,17].

Single nucleotide polymorphisms (SNPs) are the most abundant type of genetic variation. Theoretically, SNPs can have four alleles, but in practice, they have been used as bi-allelic markers since in over 99% of cases only two alleles have been observed at a given locus [18]. SNPs were estimated to occur once every 500 bp to 1 kb in the human genome and once every 1 kb in the rice genome when *indica-japonica* types were compared [19,20]. Besides being abundant in genomes, additional advantages of SNP markers are their co-dominant nature and amenability to high-throughput automation that allows rapid and efficient genotyping of large numbers of samples [21]. Therefore, SNP markers are frequently used in genetic analyses, such as phylogenetic analysis, detection of population structure, construction of genetic linkage maps, and genome-wide association studies [22–24].

Lettuce, *Lactuca sativa* L., $2n = 2x = 18$, is an important vegetable crop in the Asteraceae (Compositae) family. It is almost exclusively used as a fresh vegetable in salads and as an ingredient of various foods in the western marketplace [25,26]. However, in the eastern world lettuce is grown for its delicious stem [27]. Lettuce is one of the most valuable vegetable crops in the U.S. with an annual farm gate value of over $2.1 billion in recent years [28]. Different systems have been used in classifying lettuce cultivars into horticultural types based on morphological characteristics and/or end-user properties. We adopted the five-type system, i.e., crisphead (iceberg), butterhead, romaine (cos), leaf, and stem [29] because most of the accessions are documented under these types in the National Plant Germplasm System's Genetic Resource Information Network (GRIN) database. For high-throughput genotyping of lettuce germplasm, we recently developed the LSGermOPA, a custom Oligo Pool Assay targeting 384 expressed sequence tag-derived SNP loci (255 with known mapped positions) using the Illumina's GoldenGate assay platform [30]. High quality genotypic data were obtained from 354 of the 384 SNPs (success rate = 92.2%) for 148 lettuce accessions. The phylogenetic relationships and population structure based upon the LSGermOPA-generated SNP data were consistent with previous results using other marker systems [27,31–33].

Assessing genetic diversity and population structure within germplasm collections provides an important resource to end users and a management tool for curators. In addition, germplasm collections that possess a full range of genetic diversity and phenotypic expressions have the potential to serve as platforms for association studies to identify statistically significant relationships between polymorphic markers and genes of economic and biological merit [34]. In the current study, we focused on distilling the molecular diversity and genetic structure of 298 homozygous lettuce lines and using this information to assess genome-wide marker-trait associations between SNP markers and 10 horticultural traits.

2. Materials and methods

2.1. Plant materials, genomic DNA extraction and SNP genotyping assay

Three hundred and eighty-four individual plants sampled from 356 accessions were used in this study. For some accessions, more than one plant per accession was sampled based on observed differences in morphology. These accessions were collected worldwide during 1930s–2010s and are maintained at the USDA-ARS Western Regional Plant Introduction Station (WRPIS) in Pullman, Washington. Genomic DNA was extracted from single plants using the DNeasy 96 Plant Kit (Qiagen, Valencia, CA, USA). Quality and quantity of extracted DNA samples were evaluated with Fluoroskan Ascent FL (Thermo Scientific, Hudson, NH, USA). The SNP genotyping assay was carried out at the UC Davis Genome Center using 250 ng of genomic DNA per sample and the LSGermOPA panel targeting 384 EST-derived SNP loci. A more detailed description of the genotyping procedure can be found in our previous study [30]. Seeds of the genotyped plants were harvested and planted in 2011 and 2012 at the WRPIS Central Ferry Research Farm, Central Ferry, WA, for confirming homozygosity within accessions and for phenotypic evaluation.

2.2. Phenotypic evaluation

The phenotypic traits surveyed in the field from June to November, 2011 and 2012, included horticultural type, leaf color, bolting date, flowering date, leaf anthocyanin, stem anthocyanin, stem fasciation, leaf margin undulation, leaf blistering, and seed coat color. Bolting and flowering dates were recorded when the plant rachis was 10 cm and the terminal flower of the main axis was fully open, respectively. Leaf color, anthocyanin, margin undulation and blistering and horticultural type were recorded before the bolting stage; stem anthocyanin, and fasciation were recorded after bolting. Seed coat color was observed after harvest.