Scientia Horticulturae 243 (2019) 498–505

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Distance analysis among northern and southern European legume accessions using next-generation sequencing reveal discrepancies between geographic and genetic origins

M. Martos-Fuentes[a,b], C. Egea-Gilabert[a,c], I. Mezaka[d], J.A. Fernández[a,b], M. Egea-Cortines[a,c], J. Weiss[a,c,*]

[a] Instituto de Biotecnología Vegetal, Universidad Politécnica de Cartagena, Spain
[b] Departamento de Producción Vegetal, Universidad Politécnica de Cartagena, Spain
[c] Departamento de Ciencia y Tecnología Agraria, Universidad Politécnica de Cartagena, Spain
[d] Institute of Agricultural Resources and Economics, Zinatnes str. 2, Priekuli, Priekuli distr., Latvia

ARTICLE INFO

Keywords:
DNA barcoding
Genotype by sequence
Legumes
SNP
Genetic diversity

ABSTRACT

Genotyping of accessions is a prerequisite for management and exploitation of natural genetic variation. In some cases, the costs of genotyping maybe an issue to obtain enough information. We have developed a simple protocol based on sample barcoding allowing simultaneous genotyping of plant material based on a minimum number of Next-Generation-Sequencing (NGS) reads. We used a two-step PCR whereby, in the first step, SNP locus-specific primers ligated to common sequences were applied. This allowed, in a second PCR step and using the common sequences as primers in combination with a barcode, to amplify all SNPs of an accession in one PCR reaction. As a proof of concept, we determined the genetic structure of 72 accessions of *Vicia faba* and 72 of *Pisum sativum* from Spain, Portugal, Estonia and Latvia. We used 43 primer sets for *P. sativum* and 59 primer sets for *V. faba*. After sequencing result scoring, a total of 40 SNPs were used for downstream analysis in *P. sativum* and 45 in *V. faba*. Ten *P. sativum* and fourteen *V. faba* accessions had over 20% of heterozygote SNPs indicating that many accessions are landraces and not pure. Distance matrices of *P. sativum* and *V. faba* showed that genotypes classify into groups that are not always consistent with their geographical origin, except for two clusters of Latvian accessions in case of pea and three separate clusters formed by Spanish accession in case of faba bean. Our results demonstrate the feasibility of NGS-based genotyping using multiple barcoded samples.

1. Introduction

Understanding the distribution of genetic structure and variation of accessions are important to manage genetic resources, but also to utilize appropriate germplasm in breeding programs. Advances in molecular techniques have enabled the study of genetic variability at the DNA level, which has increased the accuracy in assessing the genetic diversity and identifying cultivars. Germplasm analysis takes advantage of the advent of Next-Generation Sequencing (NGS), which has caused a major decrease in the price of sequencing and changed the way molecular markers can be obtained. The most advanced NGS analysis is performed in model organisms such as *Arabidopsis* or rice, where sequenced genomes allow the resequencing of thousands of accessions, creating a framework of Single Nucleotide Polymorphism (SNPs), indels and copy number variants linked to traits (Li et al., 2014; Kawakatsu et al., 2016). This is the so-called Genotyping by Sequencing or GBS (Deschamps et al., 2012; He et al., 2014). SNP identification by sequencing plays an important role both for species identification and modern linkage mapping (Ganal et al., 2009), as SNPs represent an important source of polymorphism among and between genotypes. Whether a SNP is valuable for genotyping depends on a series of informative characters, which include the minor allele frequency and lack of missing data across genotypes. The number of SNPs necessary for genotyping may be as low as 7 SNPs as reported for genotyping hop varieties (Henning et al., 2015).

While NGS-based approaches are very powerful, the economic aspect of NGS can become an issue when large germplasm sets have to be analysed. The current trend is to perform a very high number of reads in NGS projects. However, read coverage, defined as the number of reads of each individual SNP from a specific sample, of as little as 2x-30x have

* Corresponding author at: Instituto de Biotecnología Vegetal, Universidad Politécnica de Cartagena, Spain.
E-mail address: julia.weiss@upct.es (J. Weiss).

https://doi.org/10.1016/j.scienta.2018.09.007
Received 18 May 2018; Received in revised form 19 August 2018; Accepted 4 September 2018
0304-4238/ © 2018 Elsevier B.V. All rights reserved.

been shown to be valid for heterozygosity detection or rare alleles in large populations (Li et al., 2011; Lynch, 2009). Barcoding of sub-populations of samples and a multiplex sequencing application, the so-called "Bar-seq", may be a solution for efficient and affordable GBS (Smith et al., 2010). However, the performance of barcoded PCR amplicon sequencing may be hampered by problems such as those introduced during the preparation of DNA, a failure of amplification depending on marker (SNP)/sample combination and a low efficiency during amplification, which depends on the sequence composition (Mallona et al., 2011; Pawluczyk et al., 2015). As a result, the number of sequences representing specific SNP/sample combinations in the different amplicon libraries may vary with under- and over-representations of certain sequences. The novel approach of genotyping by sequencing still requires the development of routine protocols specially as they become increasingly used in non-model species (Nybom et al., 2014).

The role of legumes in the human diet and nutrition of animals and farming systems is increasingly important as they have a low environmental impact compared to other crops or protein rich sources (Tilman and Clark, 2014). They can be considered as multipurpose crops because they are used as vegetables, grains and forage (Ntatsi et al., 2018). Furthermore, integration of legumes in crop rotation contributes to an improvement in soil fertility through nitrogen fixation and sustainability. Nowadays, soybean is globally the main protein source used (FAO, 2004). One of the mayor reasons outlined by the European Union for increasing domestic production of grain legumes is the reduction of dependency on soybean imports from the Americas (Cernay et al., 2015) and there is a strong interest in maximizing the use of locally produced grain legumes, such as peas (*Pisum sativum* L.), cowpea or faba beans (*Vicia faba* L.) (Carvalho et al., 2017; Koivunen et al., 2016).

Constraints for the production of faba bean and the reason for its decrease during the last century in European countries are low yields, susceptibility to biotic and abiotic stresses as well as contents of non-nutritional factors and therefore a reduced demand of bean as fodder plant together with a lack of breeding programs (Cubero, 2011). Europe accounts for 50–75% of world pea production. Dry pea ranks third to common bean and chickpea as the most widely grown pulse in the world with more than 14 million tons produced in 2016 (Tayeh et al., 2015; FAOSTAT, 2018). Advances in sequencing technologies have led to the development of large collections of EST-derived SNPs for several legume species including faba bean (Kaur et al., 2014) and pea (Leonforte et al., 2013).

The aim of the current work was to develop a simple, yet robust protocol for genotyping of multiple accessions and/or species in single NGS runs. The present study presents a novel protocol for the sequencing of a barcoded amplicon library, designed for the simultaneous identification of SNPs in independent samples on a single sequencing run. We were able to develop a minimal set of 40 SNPs for P. sativum and 43 for V. faba and a total of 24 accessions, 12 per species, that could be resolved on single sequencing reactions, thus enabling a significant reduction of costs and making NGS-based genotyping affordable. The results of our analysis using germplasm from Latvia, Estonia, Portugal and Spain could explain how diversity is distributed in relation to the geographic origin of the germplasm.

2. Materials and methods

2.1. Plant material

The accessions of *V. faba* and *P. sativum* selected for genotyping originated from Latvia, Estonia, Portugal and Spain. In total, 72 accessions were selected for each species based on local availability and usage (Table S1). In case of *P. sativum*, 30 accessions originated from Spain, 33 from Latvia, 5 from Estonia and 4 from Portugal. In case of *V. faba*, 51 accessions derived from Spain, 15 from Latvia, 4 from Portugal and 2 from Estonia. The accessions included cultivars and local varieties as well as genetic stocks. Spanish accessions were supplied by Universidad Politécnica de Cartagena (UPCT), Red Murciana de Semillas (RMS) and National Plant Genetic Resources Centre of the Spanish National Institute for Agriculture and Food Research and Technology (CRF-INIA). Portuguese accessions were provided by the National Institute for Agricultural and Veterinary Research (INIAV). Latvian accessions were provided by the State Priekuli Plant Breeding Institute (SPPBI) and Pure Horticultural Research Centre (PHRC) and by Estonian Crop Research Institute (ECRI) for the Estonian accessions. For each accession of each species, three seeds were germinated in a mixture of substrate/perlite/vermiculite (1:1:1) and leaves were harvested shortly after germination when sufficient plant material was available, corresponding to the vegetative growth stage when the first true leaf is unfolded.

2.2. DNA extraction

Fresh leaf material was pooled from three young plants/accession. DNA samples were extracted using the commercial kit 'Plant NucleoSpin' (Machery and Nagel, Düren, Germany). All extracted samples were quantified with a Nanodrop 2000 and were diluted to 50 ng/µl in order to have identical concentrations, if necessary in combination with isopropanol-ethanol precipitation. PCR reactions were performed using 100 ng of genomic DNA.

2.3. SNP selection, primer design

SNP sequences were selected based on sequence information available for high-density genetic maps of *P. sativum* with seven linkage groups corresponding to seven chromosomes and *V. faba* with twelve linkage groups corresponding to six chromosomes (Kaur et al., 2014; Leonforte et al., 2013). SNP sequences representing the described linkage groups were chosen and 43 primer sets for *P. sativum* and 59 primer sets for *V. faba* were designed using the Custom Primers-Oligo Perfect™ Designer tool from Invitrogen for a primer size of 15 bases, a primer melting temperature (Tm) of 50 °C and a PCR product size around 100 bp (Table S2). The forward SNP primers were designed with a common starting sequence "P1 – CCTCTCTATGGGCAGTCGGT GAT" (Ion Amplicon Library Preparation (Fusion Method) Life Technologies) and the reverse SNP primers started with a common sequence "B – CAGTAACAGCGGACGTGCTCTGAT" (Fig. 1 PCRI). These primers were applied in a first round of PCR. For the second round of PCR, we used forward primer "P1" and, as reverse primer, 12 special barcoded primers consisting of adaptor "A- CCATCTCATCCCTGCGTGTCTCC GAC" sequence plus a specific barcode of 5 bases and partial sequence "B- CAGTAACAGCGGACGTGCTC" (Fig. 1 PCRII). These 12 barcoded primers were used for the amplification of all accessions in pools of 12 where every pool contained equal aliquots of PCRI products of one accession. Each of the 12 accession, had a unique barcode (i.e. CTCAG for accession 1, CTGAC for accession 2). The sequences "A" and "P1" were universal primer sequences for the Ion Amplicon Library Preparation.

2.4. Amplicon library generation

PCR I and II were performed using the OneTaq ® DNA polymerase Kit (New England Biolabs) according to the manufacturer´s instruction. For PCR I, 100 ng (2 µl) of genomic DNA were amplified separately with each SNP primer pair. PCR I conditions were as follows: 10 cycles of denaturation at 94 °C for 10 s., annealing at 50 °C for 15 s., extension at 68 °C for 20 s. and 30 cycles of denaturation at 94 °C for 10 s., annealing at 60 °C for 15 s., extension at 68 °C for 20 s. The PCR I products of all SNPs were pooled for each accession and 1 µl aliquots were used for PCR II, always 12 accessins at a time applying the 12 different barcode primers. PCR II conditions were as followed: 10 cycles of denaturation at 94 °C for 30 s., annealing at 60 °C for 30 s., extension at 68 °C for 30 s

Download English Version:

https://daneshyari.com/en/article/9489616

Download Persian Version:

https://daneshyari.com/article/9489616

Daneshyari.com