Massive Analysis of cDNA Ends (MACE) for transcript-based marker design in pea (*Pisum sativum* L.)

Aleksandr Zhernakov [a], Björn Rotter [b], Peter Winter [b], Alexey Borisov [a], Igor Tikhonovich [a,c], Vladimir Zhukov [a,*]

[a] All-Russia Research Institute for Agricultural Microbiology, Podbelsky chausse 3, 196608 Saint-Petersburg - Pushkin, Russia
[b] GenXPro GmbH, Altenhöferallee 3, 60438 Frankfurt am Main, Germany
[c] Saint-Petersburg State University, Universitetskaya embankment, 7/9, 199034 Saint-Petersburg, Russia

ARTICLE INFO

ABSTRACT

Aimed at gene-based markers design, we generated and analyzed transcriptome sequencing datasets for six pea (*Pisum sativum* L.) genetic lines that have not previously been massively genotyped. Five cDNA libraries obtained from nodules or nodulated roots of genetic lines Finale, Frisson, Sparkle, Sprint-2 and NGB1238 were sequenced using a versatile 3′-RNA-seq protocol called MACE (Massive Analysis of cDNA Ends). MACE delivers a single next-generation sequence from the 3′-end of each individual cDNA molecule that precisely quantifies the respective transcripts. Since the contig generated from the 3′-end of the cDNA by assembling all sequences encompasses the highly polymorphic 3′-untranslated region (3′-UTR), MACE efficiently detects single nucleotide variants (SNVs). Mapping MACE reads to the reference nodule transcriptome assembly of the pea line SGE (Transcriptome Shotgun Assembly GDTM00000000.1) resulted in characterization of over 34,000 polymorphic sites in more than 9700 contigs. Several of these SNVs were located within recognition sequences of restriction endonucleases which allowed the design of co-dominant CAPS markers for the particular transcript. Cleaned reads of sequenced libraries are available from European Nucleotide Archive (http://www.ebi.ac.uk/) under accessions PRJEB18101, PRJEB18102, PRJEB18103, PRJEB18104, PRJEB17691.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications

Organism/cell line/tissue	*Pisum sativum* L., nodules or nodulated roots of pea genetic lines Finale, Frisson, Sparkle, Sprint-2 and NGB1238
Sex	–
Sequencer or array type	Illumina HiSeq 2000
Data format	Raw and analyzed
Experimental factors	–
Experimental features	The Massive Analysis of cDNA Ends (MACE) protocol was used for preparation of sequencing libraries.
Consent	Allowed for reuse.
Sample source location	Lines from Collection of All-Russia Research Institute for Agricultural Microbiology, Saint-Petersburg, Russia

1. Direct link to deposited data

http://www.ebi.ac.uk/ena/data/view/PRJEB18101
http://www.ebi.ac.uk/ena/data/view/PRJEB18102
http://www.ebi.ac.uk/ena/data/view/PRJEB18103
http://www.ebi.ac.uk/ena/data/view/PRJEB18104
http://www.ebi.ac.uk/ena/data/view/PRJEB17691

2. Introduction

Garden pea (*Pisum sativum* L.) is one of the most agriculturally important legumes in the world and a versatile model plant for studying the genetic bases of beneficial plant-microbe interactions [1]. Hence, the development of genetic and genomic resources for pea such as single nucleotide variants (SNV) datasets is demanded for both basic and applied science. These SNVs may serve as a base for marker development for genotyping and/or genetic mapping. Considering the lack of a pea genomic sequence, transcriptome analysis by next generation sequencing (NGS) is an appropriate solution for SNV discovery. We here focused our efforts on such genetic lines which have been used in several mutagenesis programs aimed at identification of pea symbiotic genes involved in the interaction of the plant with nodule bacteria and arbuscular-mycorrhizal fungi [2–6]. We expect that the development of transcript-based molecular markers will facilitate genetic mapping of symbiotic genes with unknown genomic location.

* Corresponding author.
E-mail address: vladimir.zhukoff@gmail.com (V. Zhukov).

3. Experimental design, materials and methods

3.1. Biological materials

Transcriptomic analysis was performed on five pea (*Pisum sativum* L.) genetic lines: Finale = JI2678 [2], Frisson = JI2491 [3], NGB1238 = JI0073 (also known as WBH1238, WL1238), Sparkle = JI0427 [4], Sprint-2 = JI2612 [6] (JI - identifiers of JIC Pisum Collection, https://www.seedstor.ac.uk/search-infocollection.php?idCollection=6). Seeds were surface-sterilized with concentrated sulfuric acid (98%) (15 min on a shaker), washed 10 times with autoclaved distilled water, and germinated on Petri dishes containing sterile vermiculite for 3 days. The germinated seeds were then planted into 2 L pots containing quartz sand (5 seedlings per pot), watered with nitrogen-free mineral nutrition solution [7], and inoculated with an aqueous suspension of *Rhizobium leguminosarum* bv. *viciae* RCAM1026 [8] (1×10^6 CFU per pot). Samples (nodules or nodulated roots of all plants from one pot) were harvested according to peculiarities of pea lines: on day 14 post inoculation (dpi) for Sparkle, on 21 dpi for Sprint-2, on 28 dpi for Finale, Frisson and NGB1238. Harvested material (mature nodules of lines Finale, Frisson and Sprint-2, nodulated roots of lines NGB1238 and Sparkle) was placed in liquid nitrogen, ground into powder, and stored at −80 °C until needed.

3.2. Libraries preparation and sequencing

RNA isolation, NGS-library preparation and sequencing were performed at GenXPro GmbH, Frankfurt am Main, Germany. RNA was isolated using the Nucleospin miRNA Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) according to the protocol for isolation of total RNA from plant tissue. MACE libraries were constructed using the MACE kit [9] according to the manual provided with the kit and sequenced on an Illumina HiSeq 2000 with 100 cycles.

3.3. Bioinformatics

For SNVs discovery we used as a reference the pea nodules transcriptome assembly [10] constructed for the genetic line SGE = JI3023, which is deposited at NCBI Transcriptome shotgun assembly (TSA) under accession GDTM00000000.1. Trimmed and cleaned reads of each library were mapped to the assembly with the Bowtie2 program v. 2.2.5 [11]. During the mapping process, SM-tag designating the pea genetic line was added to each read. Compiled SAM-files were converted to BAM format and merged into the single BAM-file. SNV-calling followed by preliminary filtering of SNVs with mapping quality lower than 20 were executed with the BCFtools utilities [12]. Sites where the coverage with high-quality bases (DP) was less than 10 were not considered and were marked as 'unknown' for a particular genetic line. Sites where the DV/DP ratio of the high-quality non-reference bases number (DV) to the total number of high-quality bases (DP) exceeded 0.9 were considered as SNVs (Suppl. Table 1). For the detected SNVs using the original script we searched for recognition sequences of restriction enzymes that would cut either the canonical or the variant site and thus would generate a *co*-dominant Cleaved Amplified Polymorphic Sequence (CAPS) marker. Recognition sequences of restriction enzymes longer than 3 bp were retrieved from the New England Biolabs (NEB, UK) catalogue.

3.4. Approbation of CAPS markers

For ten contigs containing in total 13 SNVs we developed CAPS-markers distinguishing differences either between lines Finale and NGB1238 (six detected SNVs) or lines Sparkle and NGB1238 (seven detected SNVs). PCR primers were designed on the base of sequences of publically available pea transcriptome assemblies and ESTs with help of the online tool Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast) [13], taking into account the exon-intron structure of assumed orthologous genes of *Medicago truncatula* Gaertn. predicted by aligning the pea contigs with *M. truncatula* genome (ver. Mt4.0, www.phytozome.org). PCR resulted in specific amplification in nine cases out of ten, and digestion with proper restriction endonuclease led to predicted digestion pattern for 11 SNV sites (Suppl. Table 2).

4. Conclusion

As a result, 34,711 polymorphic sites were characterized in 9724 contigs of the pea nodule transcriptome assembly. For 28,494 SNVs it is potentially possible to design CAPS markers. For a total of 10 loci primers were designed, and of these 9 could be amplified neatly. 8 of them could be digested differently for distinct lines with the appropriate restriction enzymes and are thus markers. The generated dataset provides necessary information for gene-based markers design in pea, which is useful, in particular, for genetic mapping of the genes related to symbiotic interactions with nodule bacteria and arbuscular-mycorrhizal fungi, since over 90% of described pea symbiotic mutants are obtained on backgrounds Finale, Frisson, SGE, Sparkle and Sprint-2 [14].

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.12.004.

Acknowledgements

References

[1] O.Y. Shtark, A.Y. Borisov, V.A. Zhukov, N.A. Provorov, I.A. Tikhonovich, Intimate associations of beneficial soil microbes with host plants. Soil Microbiology and Sustainable Crop Production, Springer 2010, pp. 119–196.

[2] K. Engvild, Nodulation and nitrogen fixation mutants of pea, *Pisum sativum*. Theor. Appl. Genet. 74 (1987) 711–713, http://dx.doi.org/10.1007/BF00247546.

[3] G. Duc, A. Messager, Mutagenesis of pea (*Pisum sativum* L.) and the isolation of mutants for nodulation and nitrogen fixation. Plant Sci. 60 (1989) 207–213, http://dx.doi.org/10.1016/0168-9452(89)90168-4.

[4] B. Kneen, N. Weeden, T. LaRue, Non-nodulating mutants of *Pisum sativum* (L.) cv. Sparkle. J. Hered. 85 (1994) 129–133.

[5] O. Kosterin, S. Rozov, Mapping of the new mutation blb and the problem of integrity of linkage group I. Pisum. Genet. 25 (1993) 27–31.

[6] A. Borisov, E. Morzhina, O. Kulikova, S. Tchetkova, V. Lebsky, I. Tikhonovich, New symbiotic mutants of pea (*Pisum sativum* L.) affecting either nodule initiation or symbiosome development. Symbiosis (1993) 297–313.

[7] A.Y. Borisov, S. Rozov, V. Tsyganov, E. Morzhina, V. Lebsky, I. Tikhonovich, Sequential functioning of Sym-13 and Sym-31, two genes affecting symbiosome development in root nodules of pea (*Pisum sativum* L.). Mol. Gen. Genet. MGG 254 (1997) 592–598, http://dx.doi.org/10.1007/s004380050456.

[8] V. Safronova, N. Novikova, Comparison of two methods for root nodule bacteria preservation: lyophilization and liquid nitrogen freezing. J. Microbiol. Methods 24 (1996) 231–237, http://dx.doi.org/10.1016/0167-7012(95)00042-9.

[9] A.M. Zawada, K.S. Rogacev, S. Müller, B. Rotter, P. Winter, D. Fliser, et al., Massive analysis of cDNA ends (MACE) and miRNA expression profiling identifies proatherogenic pathways in chronic kidney disease. Epigenetics 9 (2014) 161–172, http://dx.doi.org/10.4161/epi.26931.

[10] V.A. Zhukov, A.I. Zhernakov, O.A. Kulaeva, N.I. Ershov, A.Y. Borisov, I.A. Tikhonovich, De novo assembly of the pea (*Pisum sativum* L.) nodule transcriptome. Int. J. Genomics 2015 (2015), 695947. http://dx.doi.org/10.1155/2015/695947.

[11] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (2012) 357–359, http://dx.doi.org/10.1038/nmeth.1923.

[12] P. Danecek, A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, et al., The variant call format and VCFtools. Bioinformatics 27 (2011) 2156–2158, http://dx.doi.org/10.1093/bioinformatics/btr330.

[13] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13 (2012) 1, http://dx.doi.org/10.1186/1471-2105-13-134.

[14] A.Y. Borisov, T.N. Danilova, T.A. Koroleva, T.S. Naumkina, Z.B. Pavlova, A.G. Pinaev, et al., Pea (*Pisum sativum* L.) regulatory genes controlling development of nitrogen-fixing nodule and arbuscular mycorrhiza: fundamentals and application. Biologia 59 (2004) 137–144.