

# Preparation of DNA libraries for sequencing of *SAD* and *FAD* genes in flax

P. Kezimana<sup>1,2\*</sup>, T.A. Rozhmina<sup>1,3</sup>, G.S. Krasnov<sup>1</sup>, R.O. Novakovskiy<sup>1</sup>, L.V. Povkhova<sup>1,4</sup>, E.N. Pushkova<sup>1</sup>, M. Pavelek<sup>5</sup>, E.V. Romanova<sup>2</sup>, A.A. Dmitriev<sup>1</sup>, N.V. Melnikova<sup>1</sup>

<sup>1</sup> Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

<sup>2</sup> Peoples' Friendship University of Russia (RUDN University), Moscow, Russia

<sup>3</sup> Federal Research Center for Bast Fiber Crops, Torzhok, Russia

<sup>4</sup> Moscow Institute of Physics and Technology, Dolgoprudny, Russia

<sup>5</sup> Agritec, Plant Research LTD, Sumperk, Czech Republik

DOI 10.18699/ICG-PlantGen2019-49

© Autors, 2019

\* e-mail: k1par@mail.ru

**Abstract:** Sequencing technologies such as targeted amplicon sequencing are highly effective approaches for studying genetic variation in specific genomic regions as they allow efficient variant identification and characterization. They are useful in studying specific biosynthetic pathways in plants that will allow the breeding of new cultivars with targeted composition. Given that flax seed is a major source of omega-3 fatty acids, we designed a study of the polymorphisms in genes involved in fatty acid biosynthesis. In this study, we report on the library preparation for the targeted amplicon sequencing of these genes.

**Key words:** targeted amplicon sequencing; *SAD*; *FAD*; DNA library preparation; fatty acids; flax.

## 1. Introduction

Sequencing technologies have become an ideal approach to study genetic variation in key genes in several metabolic pathways in plants. DNA polymorphisms are important markers in genetic analyses and can be detected by using genome resequencing. However, using polymorphisms in modern genetic applications requires low-error high-throughput analytical strategies that can be limited by the presence of repetitive sequences and structural variants that lead to false positive results in the identification of polymorphic alleles (Qi et al., 2014). Moreover, high-throughput sequencing technologies have become a powerful and informative tool for single nucleotide polymorphism (SNP) discovery that leads to the identification of genetic variations in particular genes. However, although the sequencing of individual cultivar may be cost-effective, the sequencing of many cultivars for the assessment of genetic variations is both expensive and difficult to achieve. To circumvent these issues, the targeted deep sequencing of amplicons can be used as a powerful approach for the rapid phylogenetic classification and characterization of several hundreds to thousands of plant cultivars. Here, we describe an analysis strategy that allows assessment of the polymorphism in key genes of the metabolic pathway that can be used in the development of new cultivars. For this study, we analyzed the biosynthetic pathway of fatty acids (FAs) in flax, as flax seed is a very rich source of omega-3 FAs, to which several health benefits have been attributed. The key enzymes involved in FA biosynthesis are fatty acid desaturases and elongases (Warude et al., 2006). The desaturation of FAs is an important step as it determines the saturated to unsaturated fatty acid ratio and, ultimately, the end-use of the oil as a food source or for industrial applications (Lee et al., 2016). Genes involved in the desaturation of FAs in flax have been identified (Green, 1986; Fofana et al., 2004, 2006; Vrinten

et al., 2005; Banik et al., 2011). Stearoyl-ACP desaturases (*SAD*) introduce the first double bond at the  $\Delta 9$  position (Ohlrogge and Browse, 1995; Post-Beittenmiller et al., 1992) and the fatty acid desaturases (*FAD*) introduce an additional double bond in the FAs. Two paralogous of *SAD* loci, *SAD1* and *SAD2*, have been identified in flax (Singh et al., 1994; Fofana et al., 2004). Two closely related *FAD2* genes, *FAD2a* and *FAD2b*, encoding the proteins responsible for desaturation of oleic acid into linoleic acid by addition of a double bond at the  $\Delta 12$  position have been identified in flax (Krasowska et al., 2007; Khadake et al., 2009). Other *FAD* genes (*FAD3* genes) that encode proteins responsible for the desaturation of linoleic acid into linolenic acid by adding a double bond at the  $\Delta 15$  position have also been identified. More specifically, three *FAD3* genes were identified in the flax genome: *FAD3a*, *FAD3b* (Vrinten et al., 2005), and *FAD3c* (Banik et al., 2011). Given the advantages of amplicon sequencing, in this study, we describe the workflow for preparation of DNA libraries for targeted amplicon sequencing of *SAD*, *FAD2*, and *FAD3*.

## 2. Materials and methods

A total of 288 flax cultivars with different FA compositions, obtained from the Institute for Flax (Torzhok, Russia), were grown in a Petri dish for 5–7 days. DNA was extracted from the pools of seedlings (about 50 plants from each cultivar) with a modified CTAB method, as cetyltrimethylammonium bromide (CTAB) is a surfactant useful for isolation of DNA from tissues containing high amounts of polysaccharides (Clarke, 2009). The quality and quantity of the extracted DNA were assessed using gel electrophoresis, fluorometric analysis using Qubit 2.0 (Life Technologies, USA), and PCR analysis. The extracted DNA was purified using CleanMag DNA (Evrogen, Russia). For the preparation of *SAD*, *FAD2*,

and *FAD3* libraries, we modified the protocol for targeted amplicon sequencing on the Illumina platform ([https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry\\_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf)).

### 3. Results and discussion

For sample preparation, we used two-stage PCR, which allows adding the nucleotide sequences necessary for subsequent sequencing and sample barcoding to the target amplicons. This approach does not require the use of special kits but requires only specific primers and high-precision polymerase. Such sample preparation can significantly reduce the cost per sample. As recommended in the Illumina protocol, we defined forward and reverse primers that are complementary upstream and downstream of the region of interest, designed with overhang adapters, and used them to amplify templates from genomic DNA. For the regions of interest in this study (*SAD*, *FAD2*, *FAD3* genes), we analyzed the sequences from GenBank, found conservative domains, and designed primers for overlapping amplicons (400–520 nt in length) covering the gene region and additional 500–1000 bp in the promoter region and 100–200 bp at the end of the gene using NCBI Primer-BLAST and the MEGA software. As a result, we designed 12 primer pairs for *SAD* genes, 10 for *FAD2* genes, and 16 for *FAD3* genes. The primers were purchased from the Evrogen company.

The first-stage PCRs were carried out in a final volume of 10  $\mu$ l containing 5 ng of total genomic DNA, 0.5  $\mu$ M of each primer (forward and reverse), 1x Tersus PCR buffer (Evrogen), 0.2 mM dNTPs (Evrogen), and 1x Tersus polymerase (Evrogen). To increase the sensitivity, specificity, and yield, touchdown PCR was used, so the PCRs were performed on a GeneAmp PCR System 9700 (Applied Biosystems, USA) using the following conditions: an initial denaturation for 3 min at 95 °C followed by 10 cycles at 95 °C for 15 s, 65 °C minus 1 °C after each step for 30 s, 72 °C for 30 s and a subsequent run of 30 cycles at 95 °C for 15 s, 55 °C for 30 s, 72 °C for 30 s, and a final extension for 3 min at 72 °C. The PCR products were purified using paramagnetic beads from a CleanMag DNA kit (Evrogen), which allowed us to remove primer dimers and other contaminations.

In the second-stage PCR, we used a unique combination of forward and reverse Nextera XT v2 index primers for each sample that allowed double indexing, thus reducing the cost of sample preparation of the various amplicons, as the index primers allow amplification of any amplicons containing the universal adapters. Thus, the set of index primers is the same for different applications. For each gene studied, the amplicons were combined for each flax cultivar by taking 4  $\mu$ l per sample before the second-stage PCR. The PCRs were carried out in a final volume of 30  $\mu$ l containing 3  $\mu$ l of the purified first-stage PCR product, 0.3  $\mu$ M of each primer, 1x Tersus PCR buffer (Evrogen), 0.2 mM dNTPs (Evrogen), and 2x Tersus polymerase (Evrogen). The following conditions were used: an initial denaturation for 3 min at 95 °C followed by 10 cycles at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, and a final extension for 3 min at 72 °C.

The second-stage PCR products were purified two times with CleanMag DNA (Evrogen) to remove short DNA fragments and any contaminations before sequencing. The quantity and quality were assessed by Qubit 2.0 (Life Technologies) and an Agilent 2100 bioanalyzer (Agilent, USA). The mean length of DNA libraries was about 550–600 bp, corresponding to the expected length. The DNA libraries were normalized to 4 nM and pooled before sequencing on MiSeq Illumina (USA).

Briefly, the method used to prepare the DNA libraries for targeted amplicon sequencing makes the construction of DNA libraries possible with the introduction of adapter and index sequences to the locus-specific sequence being studied, thus allowing the sequencing of many samples simultaneously.

### 4. Conclusions

A simple and flexible method of library preparation is reported in this study that can accelerate genetic variation studies in target key genes involved in metabolic pathways. In this study, the genes examined are involved in FA biosynthesis in flaxseed, which is a major source of beneficial omega-3 FAs. The approaches reported in this study not only allow one to reduce the cost by sample with the two-step PCR with Nextera Illumina indexes but also allow flexibility for a wide range of experimental designs.

### References

- Banik M., Duguid S., Cloutier S. Transcript profiling and gene characterization of three fatty acid desaturase genes in high, moderate, and low linolenic acid genotypes of flax (*Linum usitatissimum* L.) and their role in linolenic acid accumulation. *Genome*. 2011;54:471–483. DOI 10.1139/g11-013.
- Clarke J.D. Cetyltrimethyl ammonium bromide (CTAB) DNA mini-prep for plant DNA isolation. *Cold Spring Harb. Protoc.* 2009, pdb.prot5177. DOI 10.1101/pdb.prot5177.
- Fofana B., Cloutier S., Duguid S., Ching J., Rampitsch C. Gene expression of stearoyl-ACP desaturase and  $\Delta$ 12 fatty acid desaturase 2 is modulated during seed development of flax (*Linum usitatissimum*). *Lipids*. 2006;41:705–712. DOI 10.1007/s11745-006-5021-x.
- Fofana B., Duguid S., Cloutier S. Cloning of fatty acid biosynthetic genes  $\beta$ -ketoacyl CoA synthase, fatty acid elongase, stearoyl-ACP desaturase, and fatty acid desaturase and analysis of expression in the early developmental stages of flax (*Linum usitatissimum* L.) seeds. *Plant Sci*. 2004. DOI 10.1016/j.plantsci.2004.01.025.
- Green A.G. Genetic control of polyunsaturated fatty acid biosynthesis in flax (*Linum usitatissimum*) seed oil. *Theor Appl Genet*. 1986;72:654–661. Available at: <https://link.springer.com/content/pdf/10.1007%2FBF00289004.pdf> [Accessed June 5, 2018].
- Khadake R.M., Ranjekar P.K., Harsulkar A.M. Cloning of a novel omega-6 desaturase from flax (*Linum usitatissimum* L.) and its functional analysis in *Saccharomyces cerevisiae*. *Mol. Biotechnol.* 2009; 42:168–174. DOI 10.1007/s12033-009-9150-3.
- Krasowska A., Dziadkowiec D., Polinceusz A., Plonka A., Łukasiewicz M. Cloning of flax oleic fatty acid desaturase and its expression in yeast. *J. Am. Oil Chem. Soc.* 2007;84:809–816. DOI 10.1007/s11746-007-1106-9.
- Lee J.M., Lee H., Kang S.B., Park W.J. Fatty acid desaturases, polyunsaturated fatty acid regulation, and biotechnological advances. *Nutrients*. 2016. DOI 10.3390/nu8010023.
- Ohlrogge J., Browse J. Lipid biosynthesis. *Plant Cell*. 1995;7:957–970. DOI 10.1105/tpc.7.7.957.

- Post-Beittenmiller D., Roughan G., Ohlrogge J.B. Regulation of plant Fatty Acid biosynthesis : analysis of acyl-coenzyme a and acyl-acyl carrier protein substrate pools in spinach and pea chloroplasts. *Plant Physiol.* 1992;100:923–930. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16653077> [Accessed April 26, 2019].
- Qi J., Chen Y., Copenhaver G.P., Ma H. Detection of genomic variations and DNA polymorphisms and impact on analysis of meiotic recombination and genetic mapping. *Proc. Natl. Acad. Sci. U.S.A.* 2014;111:10007–10012. DOI 10.1073/pnas.1321897111.
- Singh S., McKinney S., Green A. Sequence of a cDNA from *Linum usitatissimum* encoding the stearyl-acyl carrier protein desaturase. *Plant Physiol.* 1994;104:1075. DOI 10.1104/PP.104.3.1075.
- Vrinten P., Hu Z., Munchinsky M.-A., Rowland G., Qiu X. Two FAD3 desaturase genes control the level of linolenic acid in flax seed. *Plant Physiol.* 2005;139:79–87. DOI 10.1104/pp.105.064451.
- Warude D., Joshi K., Harsulkar A. Polyunsaturated fatty acids: Biotechnology. *Crit. Rev. Biotechnol.* 2006;26:83–93. DOI 10.1080/07388550600697479.

**Acknowledgements.** This work was performed within the framework of the Program of fundamental research for state academies for 2013-2020 years (№ 01201363824) and was funded by RFBR according to the research project 17-29-08036.

**Conflict of interest.** The authors declare no conflict of interest.