Investigation of Pin-II family genes variability in *Solanum* species by long amplicons sequencing on GS Junior (454)

Maxim S. Belenikin

Research Institute of Physico-Chemical Medicine, 119435 Moscow, Russia, camd@niifhm.ru

A.A. Krinitsina

Biology Department, Moscow State University, 119234, Moscow, Russia

A.F. Sadritdinova, A.V. Snezhkina, A.A. Dmitriev, N.Yu. Oparina, A.V. Kudryavtseva, N.V. Melnikova

Enghelhardt Institute of Molecular Biology, RAS, 119991, Moscow, Russia

Anna S. Speranskaya

Biology Department, Moscow State University, 119234, Moscow, Russia, hannadt@mail.ru

454 sequencing is a high throughput sequencing technique what can been applied to parallel investigation of plant multiple alleles of genes. Multiplexing using barcoded primers permits a large number of independent samples for analysis simultaneously.

Most *Solanaceae* contain the multi-gene family encoding members of the potato proteinase inhibitor II superfamily (Pin-II) which contribute to protection of plants from pathogens and pests. Members of Pin-II possesses of 150 b.p. repeat domain, which number can vary from 1 to 8 and inter-connecting linker. Considerable sequence diversity of Pin-II resulting from peptides variations in tandem sequence repeats, domain duplications and circularly permuted domain organizations [1].

We used 454 amplicon sequencing protocol to investigate the variation of two-domain Pin-II in *Solanum* species in particular in members of different populations of *S. nigrum* and *S. dulcamara*. The average read length of 454 sequencers is 400 b.p. what does not permits coverage of the 750 b.p. amplicons of target Pin-II genes therefore Roche GS Amplicon Variant Analyzer cannot be usable and obtained sequence data were processed by lab proprietary software.

For the 25 *Solanum* samples sequenced in this study 4293 usable 454 reads were obtained, for single samples number of reads ranged from 2 to 486.

It is known that PCR amplification can introduce recombination between templates studies. Previously studies have reported that PCR amplification resulted in the formation of

recombinant DNA sequences in from 5.4% to 20% - 37%. Pyrosequencing of standard PCR amplification products from a sample with a 50:50 mixture of two type DNA revealed that 14% of all sequencing reads were recombinants [2].

In our studies an 750 bp PCR product from plant genomes was amplified in two stage: (1) with using of designed for Pin-II non barcoded primers which additionally contained universal adaptors and (2) with barcoded sequencing primers that potentially increased probability of recombinant sequences. Nevertheless data analysis revealed recombinant reads only in 10% approximately.

After recombinant reads removing the residuary reads were processed and resulted to up to six different alleles for each of investigated *Solanum* samples.

Conclusion: 454 standart protocols and reagents (Lib-A) can be used for sequencing of amplicons of 1,5 length exceeding of recommended and data obtained may be used for analysis of alleles polymorphism of target genes.

The work was financially supported by program “Live Nature: Current State and Developmental Problems” (subprogram “Gene Pool Dynamics and Preservation”, 2012) of the Presidium of the Russian Academy of Sciences.

1. M.Mishra et al. (2012) Stress inducible proteinase inhibitor diversity in *Capsicum annuum*, *BMC Plant Biol.*, **12**: 217.

2. W.Shao et al. (2013) Analysis of 454 sequencing error rate, error sources, and artifact recombination for detection of Low-frequency drug resistance mutations in HIV-1 DNA, *Retrovirology*, **10**:18