Using DNA Metabarcoding to Identify the Composition of Herbal Teas: Comparative Analysis of Illumina and Ion Semiconductor High-throughput Sequencing Platforms

Khafizov K.F.[1], Speranskaya A.S. [1,2], Krinitsina A.A.[2], Samokhina E.N. [1], Logacheva M.D. [2], Shipulin G.A.[1]

[1] *Central Research Institute of Epidemiology, 111123, Moscow, Russia.*
[2]*M.V. Lomonosov Moscow State University, 119991 Moscow, Russia.*

Mislabeling appears to be a widespread problem for the herbal products. Advances in the DNA barcoding coupled with the next-generation sequencing offer a promising approach for the species-level identification of the eukaryote mixtures (the so-called "metabarcoding"). The usage of this approach for the species detection has the potential to monitor the contamination, misidentification, as well as the food fraud (see for example, Mishra et al., 2016; Prosser et al., 2017).

In this work, we identified the plant composition of commercially available herbal teas with a metabarcoding approach using the *ITS1-5.8SrRNA-ITS2* DNA barcode marker region and compared applicability of two high-throughput sequencing platforms: Illumina (MiSeq) *и* Ion Semiconductor (Ion S5).

The internal transcribed spacer (ITS) region (*ITS1*, *ITS2* and *5.8S* rDNA) of the nuclear ribosomal DNA (nrDNA) is traditionally used for the phylogenetic analysis, as well as plant and fungal species identification. For the analysis we have chosen multicomponent herbal teas (all from the same manufacturer) that are commercially available at the Moscow stores. The number of plant components that make up the tea varied from 11 to 12. As indicated in the manufacturer labels herbal tea product contained: Peppermint or mint leaves, Calendula flowers, Blooming sally, Chamomile flowers, Thyme grass, Dandelion roots, Juniper fruit, Sage leaves, Echinacea herb, Valeriana roots, Dog-rose fruits, Motherwort, Saint-John's-wort herbal, Plantain leaves, Oregano herbal, Stevia leaves, Hawthorn fruits, Linden flowers, Hop fruits, Air rhizomes, Horsetail, Shepherd's Purse herbal, Knot-grass, Yarrow grass, wild Strawberry leaves (i.e. genera of *Mentha*, *Calendula*, *Chamerion*, *Matricaria*, *Thymus*, *Taraxacum*, *Urtica*, *Juniperus*, *Salvia*, *Echinacea*, *Valeriana*, *Rosa*, *Leonurus*, *Hypericum*, *Plantago*, *Origanum*, *Stevia*, *Crataegus*, *Tilia*, *Humulus*, *Acorus*, *Equisetum*, *Capsella*, *Polygonum*, *Achillea* and *Fragaria*). Total DNA was extracted from ~2.5 gr. of tea (one teabag). The samples were homogenized manually using sterile mortars and liquid nitrogen. Subsequently, the DNA was extracted from ~50 ng of each sample with three different methods: (1) Doyle & Doyle, 1987 modified CTAB protocol (Krinitsina et al., 2015), (2) Dimond kit (DimondDNA) and (3) DNA-Sorb-C (AmpliSens); the last two were used according to the manufacturer's instructions.

For Ion S5 libraries construction the samples were first amplified by deploying the universal CBOL primers *ITS5/ITS2* and *ITS3/ITS4* (for *ITS1* и *ITS2* markers), after that the library preparation was conducted using standard protocol for the Ion Plus Fragment Library Kit. Sequencing was carried out on the Ion S5 employing Ion 530 Chip (Thermo Fisher Scientific Inc.) For Illumina amplicon library preparation the two-step PCR method was used. First-stage PCR products were flanked by the fusion primers containing CBOL and Illumina adaptor tail sequences. The second-stage PCR was performed with Nextera index primers. Libraries were sequenced on the MiSeq (Illumina) with MiSeq Reagent Kits V2 (500 cycles).

For the NGS data analysis, by employing the information from NCBI, we created a local database that contained sequences of *ITS1* and *ITS2* markers of the various plant and fungal species. Sequences then underwent semi-automatic filtration based on the various criteria in order to remove at least some of the false-positives. The data processing included: (a) filtration of the reads by corresponding quality scores, (b) primers trimming, (c) taxonomy identification of all the reads by aligning them against the local database using BLAST program, (d) clustering the reads based on their genera.

From the results obtained we can conclude that:

1) both sequencing platforms can be used for the genus-level identification of plant components of the herbal teas;
2) in order to develop a more accurate quantitative analysis, adaptation of the technique for the DNA extraction from "complex" plants (containing a large number of secondary metabolites represented by lignified parts), as well as the creation of local bases containing reference patterns of individual taxonomic groups of plants, is definitely needed;
3) all the samples contained impurities represented by weeds and mold fungus, although at different level;
4) at least one sample was falsified (one of the components was substituted by another one).

Prosser SW, Hebert PD (2017) Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding // Food Chem. 214: 183-191.

Mishra P, Kumar A, Nagireddy A, Mani DN, Shukla AK, Tiwari R, Sundaresan V (2016) DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. Plant Biotechnol J. 14(1): 8-21.

Krinitsina AA, Sizova TV, Zaika MA, Speranskaya AS, Sukhorukov AP. (2015) A Rapid and Cost-Effective Method for DNA Extraction from Archival Herbarium Specimens. Biochemistry (Mosc). 80(11):1478-84.