Food Control 93 (2018) 315–324

Contents lists available at ScienceDirect

Food Control

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

Comparative analysis of Illumina and Ion Torrent high-throughput sequencing platforms for identification of plant components in herbal teas

Anna S. Speranskaya [a, b, c], Kamil Khafizov [a, b, d], Andrey A. Ayginin [a, b, d], Anastasia A. Krinitsina [a, c], Denis O. Omelchenko [a], Maya V. Nilova [c], Elena E. Severova [c], Evgeniya N. Samokhina [b], German A. Shipulin [b], Maria D. Logacheva [a, c, e, *]

[a] Skolkovo Institute of Science and Technology, Moscow Region, Nobel St. 3, 143026, Russia
[b] Central Research Institute of Epidemiology, Moscow, Novogireevskaya St. 3a, 111123, Russia
[c] Lomonosov Moscow State University, Moscow, Leninskie Gory, GSP-1, 119991, Russia
[d] Moscow Institute of Physics and Technology, Moscow Region, Dolgoprudny, Institutskiy Ln. 9, 141701, Russia
[e] Extreme Biology Laboratory, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia

ARTICLE INFO

Article history:
Received 9 March 2018
Received in revised form
19 April 2018
Accepted 20 April 2018
Available online 23 April 2018

Keywords:
Metabarcoding
High-throughput sequencing
Food safety
Food quality control
Herbal tea
Lythrum

ABSTRACT

The rapid development of high-throughput sequencing (HTS) methods offers new opportunities for food quality control and identification of food components using the DNA barcoding approach (metabarcoding in cases of complex mixes). However, the protocols of DNA barcoding applied to food analysis are not yet fully established; testing and optimization are required to achieve the highest accuracy and cost efficiency. We report here a comparative study of the two most widely used sequencing platforms - Illumina and Ion Torrent - for composition analysis of herbal teas, and show that both technologies yield congruent results, both qualitatively and quantitatively. They have revealed the substitution of fireweed (*Epilobium angustifolium* L.) by *Lythrum sp.* in one of the samples. It was confirmed by classic methods of botanical analysis (anatomy and palynology). In most samples, undeclared components have been detected, such as bindweed (*Convolvulus*) and ragweed (*Ambrosia*), which are known toxic and allergy-causing plants.

1. Introduction

Food quality control is the essential part of public health care. One of the problems that quality control is facing is the incongruence between declared and actual composition. This incongruence is caused by contamination and accidental mistakes of raw materials or economically motivated adulteration during processing, which poses a threat to the health of the consumer and may reduce potential health benefits of the product (e.g., lack or substitution of a medicinal component), even causing allergic or toxic reactions (Cupp, 2000).

Tea is one of the most widely consumed drinks in the world. Most commonly called “tea”, it is a beverage made from the leaves of the *Camellia sinensis* (L.) Kuntze plant. Additionally, there are various so-called “herbal” or “phyto-” teas, which include mixes of different plant species apart from *C. sinensis*, often with medicinal or aromatic characteristics. More than a third of herbal teas sold by stores contain ingredients not indicated on the labels (Newmaster, Grguric, Shanmughanandhan, Ramalingam, & Ragupathy, 2013; Stoeckle et al., 2011). Due to flaws in the quality control of plant raw material collection and manufacture processing, herbal teas are often subject to simple mistakes or substitution and adulteration, especially if they consist of rare or expensive medicinal plants, e.g., (Boer et al., 2017; Vassou, Kusuma, & Parani, 2015). With the development of technologies for processing raw ingredients for the food industry and the globalization of food markets, it is becoming

* Corresponding author. Skolkovo Institute of Science and Technology, Moscow Region, Nobel St. 3, 143026, Russia.
E-mail addresses: hanna.s.939@gmail.com (A.S. Speranskaya), kkhafizov@gmail.com (K. Khafizov), ayginin75@gmail.com (A.A. Ayginin), ankrina@gmail.com (A.A. Krinitsina), omdeno@gmail.com (D.O. Omelchenko), nilova_maja@mail.ru (M.V. Nilova), elena.severova@mail.ru (E.E. Severova), ev.samokhina@gmail.com (E.N. Samokhina), shipgerman@gmail.com (G.A. Shipulin), maria.log@gmail.com (M.D. Logacheva).

https://doi.org/10.1016/j.foodcont.2018.04.040
0956-7135/

increasingly important to have precise instruments for the quality and safety analysis of food products.

Methods for the analysis and control of food content are constantly being developed. Traditionally, methods of analytical chemistry, organoleptic tests, macro- and microscopy are used for analysis. However, the problem of accurate and sensitive qualitative identification of food composition with plant components still exist due to numerous secondary metabolites, variations of individual chemistry profiles and the lack of distinct morphological features after food processing. Currently, molecular diagnostic methods based on the amplification or sequencing of the marker DNA regions look the most promising (Lo & Shaw, 2018).

High-throughput DNA sequencing (HTS) methods have undergone extremely rapid development over recent years. HTS allows determining the sequences of many thousands of fragments without their physical separation (e.g. by cloning) prior to sequencing and is thus very promising for the analysis of products, which represent complex mixes of components. The approach to the identification of the composition of complex mixes is called metabarcoding. Metabarcoding has a high potential to monitor contamination, misidentification and fraud in food, e.g., (Mishra et al., 2016; Prosser & Hebert, 2017). Though HTS has a great advantage in the identification of food component sources, contaminants and species composition, it has its limitations as well. It greatly depends on the data integrity of the reference sequences' database and some species are hard to clearly distinguish by means of HTS and DNA barcoding, and it is advised to use the integrative approach involving both chemical profiling and DNA-based barcoding methods for identification (Xu et al., 2018).

Currently, there are two widely used HTS approaches: the Ion Torrent and Illumina platforms (see Table 1 for characteristics). Despite high accuracy of both technologies, they display some platform-specific error profiles. For Illumina sequencers, inverted repeats and GC-rich sequence motifs, such as *GGC* and *GGT,* have increased error (substitution) frequencies (Meacham et al., 2011; Nakamura et al., 2011; Schirmer et al., 2015). For Ion Torrent sequencing, errors in homopolymer repeats resulting in false indels are the most common context specific errors (Bragg, Stone, Butler, Hugenholtz, & Tyson, 2013; Loman et al., 2012; Yeo et al., 2012). Also, the errors can occur not only during sequencing but also during sample preparation. Library construction for both Illumina and Ion Torrent sequencing includes two principal stages: (1) linking of specific adapter sequences to target DNA fragments usually followed by PCR with primers complimentary to the adapters and (2) clonal amplification of prepared libraries by solid-phase bridge PCR on the surface of the flow cell (Illumina) or emulsion PCR on the surface of microbeads (Ion Torrent). Different types of PCR introduce their own errors and thus it is important to test the performance of different HTS platforms for each specific task to estimate their applicability.

Performance of the Illumina and Ion Torrent has been compared before for 16S rRNA-based bacterial community profiling (Salipante et al., 2014), whole chloroplast barcoding (Brozynska, Furtado, & Henry, 2014), and differential gene expression (Lahens et al., 2017), but was never compared for food fraud analysis. To the best of our knowledge, this is a first direct side by side comparison of the two most widely used HTS platforms, Illumina (MiSeq) and Ion Torrent (Ion S5), applied to the problem of food quality control.

In this work, we identified the plant composition of commercially available herbal teas with metabarcoding using the internal transcribed spacers of ribosomal RNA gene operons - nrITS1 and nrITS2. They were chosen as markers because these two regions are found in multiple copies in the plant genome, can be amplified with universal primers and have lengths that are perfect for most sequencers. Despite a few drawbacks of nrITS regions (such as intragenomic polymorphism), they have been successfully used for the identification of various plant species in food analysis (Raclariu et al., 2017; Xin et al., 2013) and are recommended as universal barcodes for fungi taxon identification (Schoch et al., 2012).

2. Materials and methods

2.1. Samples and DNA extraction

Six herbal tea samples named *Mix1-6* (all from the same manufacturer) were examined. The samples were obtained from pharmacy retail outlets in Russia. The plant sources declared by the manufacturer represented the 32 species commonly used in traditional herbal medicine of Europe (see Supplementary Table 1). Common names on the product labels were translated from Russian and matched with scientific names using The PLANTS Database (National Plant Data Team, Greensboro, NC 27401-4901 USA., 2018).

Total DNA was extracted from ~2 g of tea (one teabag). The samples were homogenized in liquid nitrogen using sterile mortar and pestle. Subsequently, the DNA was extracted from ~50 mg of each sample with DiamondDNA Plant kit (ABT, Russia). After extraction DNA samples were additionally purified by the following method: Sera-Mag Magnetic Speed-beads (Dia.: 1 μm; 3 EDAC/PA5, GE Healthcare Biosciences) were prepared as described in (Molly, 2017) with the following modification: 200 μL of magnetic particles were washed twice with 500 μL of TE buffer, then diluted in 9750 μL of buffer solution (18% PEG-8000 (w/v), 1 M NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and stored at 4 °C as a stock solution. Immediately before DNA purification, the stock solution of the magnetic particles was diluted and thoroughly mixed with the buffer in the ratio of 1–3 (v/v). Then, DNA samples were purified with a suspension of magnetic beads in the ratio of 1–1.4 (v/v).

Table 1
Comparison of specifications of Illumina and Ion Torrent platforms.

Platform	Illumina MiSeq	Ion S5/Ion S5 Plus/Ion S5 Prime
Sequence yield per run	7.5–8.5 Gb on reagents v.2 12.5–15 Gb on reagents v.3	1,2-2 Gb on 520 chip 6-8 Gb on 530 chip
Accuracy	70% > Q30 at 600 cycles, 85% > Q30 at 500 cycles	85% > Q20
Systematic error	substitutions in GGC and GGT context	indels in homopolymer regions
Read length	500 (250 + 250) bp for reagents v.2 600 (300 + 300) bp for reagents v.3	~400 bp for double chip up to ~600 bp for single chip
Run Time	39 h for 500 cycles 56 h for 600 cycles (includes cluster generation, sequencing and base calling)	19,5 h for 400 bp (includes presequencing chip processing, initialization and sequencing)
Paired reads	Yes	No
Insert size	up to 550 bp	400 bp (up to 600 bp)

The concentration and purity of the DNA samples were assessed by A260/280 and A260/230 ratios on the NanoPhotometer N60-Touch (Implen, Germany) and by fluorescence intensity on the Qubit 3.0 fluorometer (Thermo Fisher Scientific, USA). DNA samples were normalized to 10 ng/μL before sample preparation. Then, 5 μL (50 ng) of each normalized sample was used for library preparation.

Sanger sequencing of nrITS of *Lythrum* species, obtained from the collection of Lomonosov Moscow State University Herbarium (MW; scanned samples available at https://plant.depo.msu.ru/), - *L. salicaria* L. (MW0202156), *L. virgatum* L. (MW0446120), *L. hyssopifolia* L. (MW0619919), *L. thymifolia* L. (MW0632153), *L. tribracteatum* Salzm. ex Ten. (MW1003497) and *L. junceum* Banks & Solander (MW0743064), - was performed on an Applied Biosystems DNA Analyzer using ABI PRISM BigDye Terminator v. 3.1 reagents.

2.2. DNA libraries preparation

For Ion S5 library preparation, the ITS1-5.8SrRNA-ITS2 DNA-barcode region of each sample was amplified by ITS5-F/ITS2-R and ITS3-F/ITS4-R primers from (Baldwin, 1992; White, Bruns, Lee, & Taylor, 1990) for nrITS1 and nrITS2 regions, respectively. Barcodes were amplified using "Hot-start" with a chemically modified DNA polymerase TaqF (AmpliSens, Russia) in the presence of the intercalating dye EvaGreen (Biotium, USA) on the QuantStudio 5 Real-Time PCR system (Thermo Fisher Scientific, USA). Subsequently, the products were purified with Sera-Mag magnetic beads and mixed in equimolar proportion. Then, all the subsequent procedures were conducted for the Ion Plus Fragment Library Kit according to the manufacturer's protocol. Procedures include DNA end blunting, ligation of the adapters to the blunt ends followed by 11 cycles of library amplification with the polymerase included in the Platinum PCR High-Fidelity SuperMix from the Ion Plus Fragment Library kit (Thermo Fisher Scientific, USA). Sequencing was carried out on the Ion S5 platform using Ion 520/530 Kit Chef reagent sample preparation kits with an Ion Chef instrument and Ion 530 chips (Thermo Fisher Scientific, USA).

For Illumina, the two-step PCR method was used for library preparation: first-stage PCR by the fusion primers containing primer sequences from (Baldwin, 1992; White et al., 1990) and Illumina adaptor tails (ITS5-Illu-F/ITS2-Illu-R and ITS3-Illu-F/ITS4-Illu-R). After that, the PCR products corresponding to nrITS1 and nrITS2 were purified by Sera-Mag beads and mixed in equimolar proportion for each mix. Second-stage PCR was performed with 50 ng of mixed products and Nextera index primers (Illumina, USA), according to the manufacturer's instructions. Both PCR stages were conducted using DNA polymerase TaqF (AmpliSens, Russia) in the presence of the intercalating dye EvaGreen (Biotium, USA) on the LightCycler®96 System (Roche, Switzerland). Illumina libraries were sequenced on a MiSeq with the MiSeq Reagent Kit v2 for 500 cycles (replicate 2) and on a HiSeq2500 with the Hiseq Rapid v. kit for 500 cycles (Illumina, USA) with a 251 + 251 cycles setting.

Primer sequences are presented in Supplementary Table 2. The composition of the PCR reaction mixture and the amplification program for both library preparation methods are presented in Supplementary Table 3.

2.3. Reference database

We have created a local reference database of the nrITS1 and nrITS2 markers based on the information available in NCBI. Over 700,000 nrITS1 and nrITS2 sequences of fungi and plants were downloaded from the NCBI and pre-filtered by length (200–10,000 bp). Poorly annotated and questionable records (Environmental, Unverified, Uncultured) were removed. The database was clustered by the CD-HIT program (Fu, Niu, Zhu, Wu, & Li, 2012) at 100% identity to discard duplicate sequences. Because NCBI contains a considerable number of erroneously annotated sequences, we performed an additional filtering step. The database was clustered again, but with a lower identity threshold (97%), to identify groups of records of the same genus. The taxonomic composition of the entries in the newly created clusters was analyzed and filtering was performed. If more than 80% of the sequences belong to the same taxon, and less than 20% to the other, the latter was removed from the database as likely erroneous. Such filtering was carried out at several taxonomic levels (from kingdom to family) and more than 5 thousand erroneous records were discarded. After filtering approximately 400,000 sequences of nrITS1, a similar procedure was used for the nrITS2 sequences remaining in the database. To ensure higher accuracy of the identification, the conservative sequences (rRNA gene fragments) that may erroneously align with reads were removed from the database. To do this, multiple alignment was performed, and conserved areas were identified and removed. This operation separated nrITS1 and nrITS2 sequences from rRNA genes. Final database includes more than 700,000 sequences, in total, of highly variable nrITS1 and nrITS2 markers of plants and fungi.

2.4. Data analysis

The pipeline for data analysis consisted of 4 modules and was developed based on the open source software: Blastn (Altschul, Gish, Miller, Myers, & Lipman, 1990) for alignment of reads, Prinseq-lite (Schmieder & Edwards, 2011) for reads filtering, cutadapt (Martin, 2011) for trimming of primers and conserved regions, and Biopython package (Cock et al., 2009) for fasta file processing. First, the sequences were filtered according to the quality scores and their lengths. Second, the primers and conserved regions of the *18s*, *5.8s*, and *28s* rRNA genes were removed from the reads. Third, filtered sequences were aligned by BLAST to the reference database, and alignments with an E-value higher than 1e-30 and sequence identity less than 90% were discarded. Fourth, the percentage composition of the sample was identified up to the genus level by summing up the alignments of the same species, considering only those species to which 100 or more reads were aligned in total. If it was impossible to reliably determine the genus, a higher taxonomic level was chosen.

To estimate the similarity between replicates and the results from different technologies, a Pearson correlation coefficient was calculated using data on OTU abundances. Abundances lower than 1% in any of the replicates were discarded.

A plant genus was considered "found" if its abundance was higher than 1% in at least 3 out of 4 replicates of the two sequencing platforms.

2.5. Anatomo-morphological and palynological analysis

For macro- and microscopic analysis, the content of one randomly selected teabag (2 g) was used. The material was first examined using a stereomicroscope Olympus SZ61 (Olympus Corp., Japan), then prepared for microscopic investigation using a standard protocol: boiling in 5% sodium hydroxide and washing with water (The State Pharmacopoeia of the Russian Federation, 2015; Vandyshev, Babaeva, & Miroshnikova, 2017). Temporary slides in glycerol were prepared to reveal specific anatomic characters, which were observed using light microscopy Nikon Eclipse Ci (Nikon, Japan). The pictures were taken with a Nikon DS-Vi1 camera. Identification was carried out using the following sources: (Jakovlev, 2013; The State Pharmacopoeia of the Russian Federation, 2015; Upton, Graff, Jolliffe, Länger, & Williamson,

2016; Vandyshev et al., 2017). Additionally, the fragments of the plants under investigation were compared with reference fragments from the collection of the Department of Higher Plants, Biological Faculty, Lomonosov Moscow State University.

For palynological analysis, 1 g of the sample was dehydrated with acetic acid for 12 h and treated with acetolysis mixture (Erdtman, 1960). Temporary slides in glycerol were prepared, and the pollen count was performed on up to 500 pollen grains at 400x magnification using a Nikon Eclipse Ci microscope (Nikon, Japan). For identification of the pollen grains, atlases (Moore, Webb, & Collinson, 1995; Reille, 1999) and databases ("Information System on Plant Morphology and Anatomy," 2018; "PalDat," 2018) were used.

3. Results

3.1. Sequence analysis

Each sample was sequenced on two HTS platforms - Illumina and Ion Torrent - in two replicates. The number of reads per sample after filtering varied from 31,049 to 223,587, with 83% (Illumina) and 75% (Ion Torrent) of the samples having more than 100,000 total reads (see Table 2). Such sequencing depth is higher than in recent studies on plant metabarcoding, e.g., (Bell et al., 2017; Sickel et al., 2015). Among these reads, however, 3.6–62.7% were of fungal origin and were excluded from the analysis. Then, 62–99% of reads were classified to the genera using the selected threshold.

Each herbal tea contains from 11 to 12 botanical sources declared by the manufacturer (in total, 32 different plants belonging to 26 genera of 15 families). All negative controls showed no traces of DNA and, therefore, were not sequenced. Analysis of nrITS1 and nrITS2 barcodes of the samples was performed against a local refined reference database (see Methods) and each hit was assigned to a specific genus as a percentage of the total identified reads for each sample.

For all six teas, the results produced with different technologies are similar (Table 3). The discrepancies are confined to the low abundance species. The correlation between replicates ranged from 0.55 to 0.96 (median correlation coefficient 0.89).

In all analyzed tea mixes, we found most of the components declared by the manufacturer, though some components were completely missing or present at a level below the threshold (<1%). The missing components are shared between mixes - e.g., *Valeriana* is claimed to be present in 5 mixes out of 6 and *Crataegus* in 6 mixes out of 6, but they are absent from all of them; *Rosa* is claimed to be present in all mixes but was found only in half of them. Additionally, 4 out of the 6 mixes have components undeclared by the manufacturer, including *Convolvulus* and *Polygonum*, which appear most frequently.

The most profound and unexpected deviation from the declared composition is the presence of *Lythrum* instead of *Epilobium* in the first tea mix. Though it was undeclared by the manufacturer, *Lythrum* was found in significant amounts from 17.4% to 48.6% by both sequencing platforms. The most widespread species of *Lythrum* is *L. salicaria*. It is a medicinal plant and is a component of some herbal tea compositions. Consensus nrITS1 and nrITS2 sequences of *Lythrum* inferred from reads are not completely identical with either *L. salicaria* or any other *Lythrum* nrITS sequences available in the NCBI GenBank database, as nrITS1 and nrITS2 *L. salicaria* (AY035749, AY035750, AF334772) showed 7% and 3% of Hamming dissimilarity, respectively. Other *Lythrum* species showed 15–22% and 3–22% of dissimilarity (AY905428, AY910747, MF964073, AY910748). In order to perform the species-level assignment for the undeclared component, we have sequenced and submitted to GenBank nrITS sequences of several *Lythrum* species: *L. salicaria* (MG975396), *L. virgatum* (MG975397), *L. hyssopifolia* (MG975399), *L. thymifolia* (MG975400), *L. tribracteatum* (MG975398) and *L. junceum* (MG975401). The species closest to the *Mix1* undeclared component in this set is *L. salicaria* (0% nrITS1 and 3% nrITS2 dissimilarity). While the sequences of *L. salicaria* are available in the GenBank, they are of different origin (China) while our sample (MG975396) is from Russia. This emphasizes the need for the sampling of multiple specimens from the same species covering the distribution range of this species. Other *Lythrum* species are much more distant from *Mix1 Lythrum* (up to 15% and 6–9% divergence for nrITS1 and nrITS2, respectively). *E. angustifolium* (JF976297, JF976296, JF976295, JF976294, JF976293, EPL58SR) is 29% and 33% distant from our sample sequence for nrITS1 and nrITS2, respectively. Thus, the undeclared component is either *L. salicaria* or a species very close to *L. salicaria*.

3.2. Botanical analysis

To test our findings based on metabarcoding with an independent approach we performed palynological and anatomo-morphological analysis for sample *Mix1*. The information on the pollen found in sample *Mix1* is presented in Table 4. Overall, it is congruent with the results of DNA-based analysis. In particular, it

Table 2
Number and proportion of the reads.

Sample	Total number of reads		Proportion of plant reads		Proportion of aligned		Total number of reads		Proportion of plant reads		Proportion of aligned	
	ITS1	ITS2	ITS1	ITS2	ITS1	ITS2	ITS1	ITS2	ITS1	ITS2	ITS1	ITS2
	Illumina replicate 1						Illumina replicate 2					
Mix1	88099	45323	0.72	0.80	0.65	0.33	25679	5370	0.71	0.81	0.64	0.13
Mix2	99370	53177	0.86	0.87	0.64	0.34	70921	23741	0.97	0.96	0.59	0.2
Mix3	98526	47653	0.68	0.82	0.66	0.32	101263	33542	0.58	0.84	0.62	0.21
Mix4	89618	53327	0.77	0.84	0.62	0.37	108809	36896	0.72	0.91	0.65	0.22
Mix5	95169	59788	0.49	0.63	0.59	0.37	93095	35117	0.37	0.60	0.45	0.17
Mix6	106992	50573	0.62	0.78	0.67	0.32	57040	11283	0.88	0.88	0.68	0.13
	Ion Torrent replicate 1						**Ion Torrent replicate 2**					
Mix1	83450	34260	0.61	0.26	0.69	0.28	38715	13716	0.64	0.47	0.3	0.98
Mix2	103470	13777	0.79	0.26	0.87	0.12	84465	36832	0.75	0.67	0.36	0.98
Mix3	123961	14322	0.54	0.21	0.87	0.1	111406	64982	0.81	0.71	0.35	0.98
Mix4	46287	18201	0.67	0.27	0.7	0.28	80514	44379	0.74	0.71	0.43	0.98
Mix5	158988	64599	0.47	0.13	0.68	0.28	70411	55276	0.73	0.64	0.39	0.97
Mix6	60188	10413	0.58	0.28	0.84	0.14	82137	55082	0.58	0.45	0.39	0.98

Table 3
Abundance of components by genera found and those not found in samples with Ion Torrent and Illumina HTS.

Genus found	Illumina #1 (%)	Illumina #2 (%)	Ion Torrent #1 (%)	Ion Torrent #2 (%)	nrITS1 median GC-content (%)
Mix1					
Mentha	5.41	1.47	4.48	4.26	65.91
Thymus	8.02	4.06	6.83	5.63	58.26
Fragaria	2.87	0.45	2.32	2.08	66.54
Stevia	14.22	11.19	3.75	16.1	46.44
Hypericum	1.57	NA	1.46	1.56	59.18
Rosa	3.01	0.65	3.34	2.72	57.59
Calendula	7.41	1.21	2.14	4.01	59.92
Matricaria	20.36	13.37	30.18	17.51	46.04
declared but not found					
Epilobium	NA	NA	NA	NA	52
Crataegus	NA	NA	NA	0.19	66.39
Valeriana	NA	NA	NA	NA	64.55
Leonurus	2.2	0.32	0.84	1.58	65.82
non-declared but found					
Lythrum	17.35	48.56	29.57	20.52	52.72
Urtica	3.52	12.1	4.39	5.71	49.48
Polygonum	3.92	2.58	2.78	3.67	62.92
Convolvulus	1.11	0.48	1.17	1.66	51.64
Mix2					
Stevia	19.39	25.6	4.09	25.56	46.44
Rosa	4.61	2.05	3.68	5.15	57.59
Echinacea	4.37	0.91	3.47	4.55	47.13
Hypericum	5.08	0.99	4.73	5.53	59.18
Taraxacum	5.9	2.23	6	8.23	48.96
Salvia	13.94	7.02	8.47	9.39	65.38
Urtica	37.98	59.14	65.34	33.87	49.48
Leonurus	4.79	1.34	2.34	3.7	65.82
declared but not found					
Juniperus	NA	NA	NA	NA	62.45
Fragaria	1.57	0.18	0.73	1.16	66.54
Crataegus	0.5	NA	NA	0.09	66.39
Mix3					
Mentha	3.5	1.94	4.61	2.64	65.91
Stevia	18.6	30.95	4.35	24.77	46.44
Fragaria	3.24	1.05	2.69	2.33	66.54
Thymus	14.25	23.47	21.47	14.35	58.26
Hypericum	3.8	1.14	4.54	3.74	59.18
Rosa	3.79	2.05	4.25	3.74	57.59
Tilia	5.29	2.06	1.06	3.45	62.29
Plantago	1.2	0.55	2.09	1.37	52.56
Leonurus	9.72	7.35	8.27	6.98	65.82
declared but not found					
Crataegus	0.55	0.08	0.45	0.21	66.39
Valeriana	0.23	0.07	NA	0.1	64.55
non-declared but found					
Convolvulus	9.56	15.07	17.21	10.89	51.64
Polygonum	4.82	2.9	5.34	2.56	62.92
Ambrosia	6.32	3.62	6.55	6.17	50
Betula	2.48	1.51	3.18	2.35	60.45
Origanum	1.56	1.39	2.78	1.66	57.27
Urtica	1.36	1.41	2.2	1.29	49.48
Halimocnemis	1.05	0.18	1.42	1.3	49.38
Mix4					
Mentha	5.29	3.02	7.21	4.29	65.91
Origanum	7.38	6.93	12.46	7.14	57.27
Stevia	19.33	26.54	6.61	23.24	46.44
Thymus	7.83	8.4	11.04	7.44	58.26
Calendula	9.8	8.04	3.61	8.47	59.92
Salvia	13.58	15.54	12.78	10.99	65.38
Matricaria	20.85	23.51	36.59	18.5	46.04
Leonurus	3.43	1.17	2.26	3.04	65.82

(*continued on next page*)

Table 3 (*continued*)

Genus found	Illumina #1 (%)	Illumina #2 (%)	Ion Torrent #1 (%)	Ion Torrent #2 (%)	nrITS1 median GC-content (%)
declared but not found					
Rosa	0.38	0.06	NA	0.41	57.59
Tilia	3.15	0.82	0.92	3.04	62.29
Crataegus	0.3	NA	NA	0.09	66.39
Valeriana	NA	NA	NA	NA	64.55
non-declared but found					
Convolvulus	2.1	1.49	3.12	2.6	51.64
Viola	1.14	0.25	1.61	1.64	64.43
Mix5					
Humulus	3	2.91	3.59	3.8	57.29
Mentha	2.32	1.01	2.76	1.86	65.91
Thymus	1.66	0.65	1.84	1.39	58.26
Stevia	21.84	35.73	6.65	25.35	46.44
Matricaria	12.34	10.37	18.41	10.95	46.04
Leonurus	20.43	25.56	15.17	16.85	65.82
declared but not found					
Tilia	2.18	0.68	0.54	2.21	62.29
Equisetum	NA	NA	NA	0.18	67.08
Rosa	0.29	0.12	NA	0.24	57.59
Acorus	NA	NA	NA	NA	74.26
Crataegus	2.93	0.49	2.68	0.74	66.39
Valeriana	NA	NA	NA	0.1	64.55
non-declared but found					
Onobrychis	9.99	7.5	20.38	11.63	53.07
Polygonum	5.93	2.71	7.68	2.5	62.92
Ambrosia	4.85	2.94	5.02	5.32	50
Echinacea	2.5	1.24	2.39	2.81	47.13
Convolvulus	1.6	1.12	2.61	1.92	51.64
Lythrum	1.36	1.52	2.23	1.74	52.72
Mix6					
Achillea	10.52	2.79	9.26	9.57	45
Fragaria	4.06	0.53	2.14	3.34	66.54
Stevia	17.11	13.82	3.11	22.47	46.44
Capsella	4.13	2.82	4.13	6.97	52.75
Polygonum	24.1	35.41	24.25	17.34	62.92
Urtica	23.52	42.97	39.2	21.87	49.48
Leonurus	5.65	0.81	2.47	4.29	65.82
declared but not found					
Equisetum	NA	NA	NA	NA	67.08
Rosa	0.83	NA	NA	0.9	57.59
Crataegus	0.66	NA	NA	0.14	66.39
Valeriana	NA	NA	NA	NA	64.55

Table 4
Pollen/spores found in sample *Mix1* by palynological analysis.

Genus/family found	Number of pollen and spores found	Abundance in the sample (%)
Anthemis type (including *Matricaria*)	138	24.6
Hypericum type	160	28.5
Stachys sylvatica type (including *Leonurus*)	48	8.6
Lythrum salicaria type	51	9.1
Rosaceae	71	12.7
Calendula type	33	5.9
Mentha type (including *Thymus*, *Mentha*)	9	1.6
Pinus	2	0.4
Centaurea scabiosa type	2	0.4
Ambrosia type	3	0.5
Tilia	4	0.7
Chenopodiaceae	1	0.2
Umbelliferae	2	0.4
Asteroideae	4	0.7
Urtica	1	0.2
Betula	3	0.5
Alnus	1	0.2
Poaceae	1	0.2
Polypodiaceae	1	0.2
Fungal spores	5	0.9
Undetermined	21	3.7

also reveals the substitution of *Epilobium* by *Lythrum*. These species are easily distinguished by the morphology and size of their pollen grains (see Supplementary Fig. 1).

At the quantitative level, the congruence between palynological and metabarcoding results is not high. For example, pollen analysis shows that *Hypericum* is the second major component, while the metabarcoding abundance is low. Partly, this incongruence is due to the obvious limitation of palynological analysis. Only species collected in the flowering stage can be detected, and the quantity of the pollen grains is not always proportional to the amount of biomass of the component, but rather reflects the pollen productivity and differs between species, e.g., (Broström et al., 2008). This may be due to the insufficient resolution, i.e., the species of related genera cannot always be distinguished.

Anatomo-morphological analysis reveals the presence of all declared components except for *E. angustifolium*, and the presence of several undeclared components, including *Lythrum* (Table 5, Supplementary Fig. 2).

It should be noted, however, that anatomo-morphological analysis is confined to only the fraction of the sample composed of plant parts, which retain taxon-specific characters and can be reliably identified. This is not the case for the fine-ground, almost homogenous fraction, which makes up approximately 65–70% of the sample. As there are no accepted standards and criteria for quantitative analysis of the herbal mixes, the amount of each component was not determined.

4. Discussion

We have found that, despite the difference in sample preparation and sequencing principle, both technologies yield similar results, at least on qualitative level. Some sequence patterns may cause errors during the Illumina sequencing process and the probability of an error depends on the type of sequence pattern. *GGT* (or *GGGT*) is the strongest motif, where the frequent error is at the *T*, resulting in an incorrect *G* (Meacham et al., 2011). Other authors demonstrated that the *GGC* (or *GGGC*) motif also influences error frequencies, causing C/G substitution (Nakamura et al., 2011; Schirmer et al., 2015). We analyzed the plant nrITS1/ITS2 database used for the analysis of herbal teas in this work (approximately 340,000 sequences) and found that 240,000 sequences contain one or more (up to five) *GGGC* motifs (see Supplementary Fig. 3a). At least one *GGGT* motif was found in more than 100,000 queries and only half of the nrITS1/ITS2 database plant sequences were free from this motif (see Supplementary Fig. 3b). In the Ion Torrent technology the most common context specific error is the false insertion or deletion in the homopolymer regions (Bragg et al., 2013; Loman et al., 2012; Yeo et al., 2012). It occurs during base-calling (a process of conversion of the raw signal from the sequencer into a sequence of nucleotides). The indel error rates of Ion Torrent increase markedly with the increase of the homopolymer length (Laehnemann, Borkhardt, & McHardy, 2016). Analysis of our plant nrITS1/ITS2 database sequences demonstrated that ~13.5% of sequences contain homopolymers with N = 7 and another portion (~1%) of sequences contain homopolymers with N = 8 (see Supplementary Fig. 2c). Notably, nrITS1 and nrITS2 sequences contain much fewer homopolymer regions than non-coding plastid markers (Devey, Chase, & Clarkson, 2009; Fazekas, Steeves, & Newmaster, 2010), which make them a better choice for HTS-based metabarcoding, especially if semiconductor sequencing is used.

Despite these differences in error profiles we found that the intra-platform (for both Illumina and Ion Torrent) and cross-platform variation is similar. The underlying source of variation between different runs of the same platform is unclear; presumably it is caused by the stochastic dynamics of PCR and other reactions that occur during sample preparation.

The most prominent incongruence between the composition declared by the manufacturer and that found using metabarcoding (supported by pollen and anatomy analysis) is the substitution of *Epilobium angustifolium* with *Lythrum* in one of the mixes. *E. angustifolium* and *Lythrum* are taxonomically far from one another, belonging to different families. However, *E. angustifolium* and some species of *Lythrum* (*L. salicaria*, *L. virgatum*) are quite similar in their appearance (Fig. 1), and an inexperienced collector can easily mix them up. Such substitution can be harmful to the consumer's health. Young shoots of *E. angustifolium* are consumed raw mixed with other greens or cooked. Its leaves could also be dried, boiled and fermented to make ale or tea; it also has medical uses to treat asthma, yeast infections, prostate, hemorrhoids and diarrhea (Granica, Piwowarski, Czerwińska, & Kiss, 2014). This plant species is traditionally used in Russia as a substitute or additive for a common tea made of *C. sinensis*. This is even reflected in its Russian common name - “Ivan-chai”, where “chai” translates as tea. *Lythrum* (exactly *L. salicaria*) is a plant widely used in European and Chinese traditional medicine (Piwowarski, Granica, & Kiss, 2015) as a remedy against gastrointestinal tract ailments. Its extract is rich in polyphenolic compounds that display a wide range of biological activities (Tunalier, Koşar, Küpeli, Çaliş, & Başer, 2007), some of which, however, are adverse, e.g., (Eck-Varanka et al., 2015).

Other discrepancies between declared and inferred composition consist of the absence of several plants - *Crataegus*, *Valeriana*, *Equisetum*, *Juniperus*, *Acorus*, and *Rosa*. Most of them are commonly used medicinal plants, are widespread and easy to obtain. It is thus unlikely that they are absent due to adulteration for economic reasons. Regarding *Equisetum*, we suggest that it is explained by the non-optimality of the primers. Indeed, based on the single complete *18S* rRNA gene sequence available for *Equisetum* (X78890, *E. robustum* A. Braun ex Engelm.), only 14 nucleotides of ITS5 primers are complementary to the target. ITS2 and ITS3 primers (which have the same annealing site but on different DNA strands) have 3–5 mismatches and ITS4 has 2 mismatches. The inefficiency of ITS primers from (White et al., 1990) to amplify nrITS2 was already reported by (Saslis-Lagoudakis et al., 2015). This highlights the need for new metabarcoding primers that are optimized for plants. *Acorus* has a very high GC-content (from 69.6% to 75.7%,

Table 5
Plants found in sample *Mix1* by anatomo-morphological analysis.

Genus/species	Part
Declared components	
Calendula officinalis L.	flowers
Crataegus sp.	fruits
Fragaria vesca L.	leaves
Hypericum sp.	leaves, flowers
Leonurus sp.	stems, leaves, flowers
Matricaria chamomilla L.	flowers, seeds at various degree of maturity
Mentha x *piperita* L.	leaves
Rosa sp.	fruits
Stevia sp.	leaves
Thymus sp.	flowers, leaves (in question)
Valeriana officinalis L.	roots
Non-declared components	
Lythrum sp.	flowers, leaves
Plantago sp.	leaves (two fragments per sample)
Poaceae	stems, leaves, flowers
Polygonum sp.	leaves (few fragments)
Setaria sp.	seed (one)
Undetermined plants	stems, leaves
Mineral fragments	NA

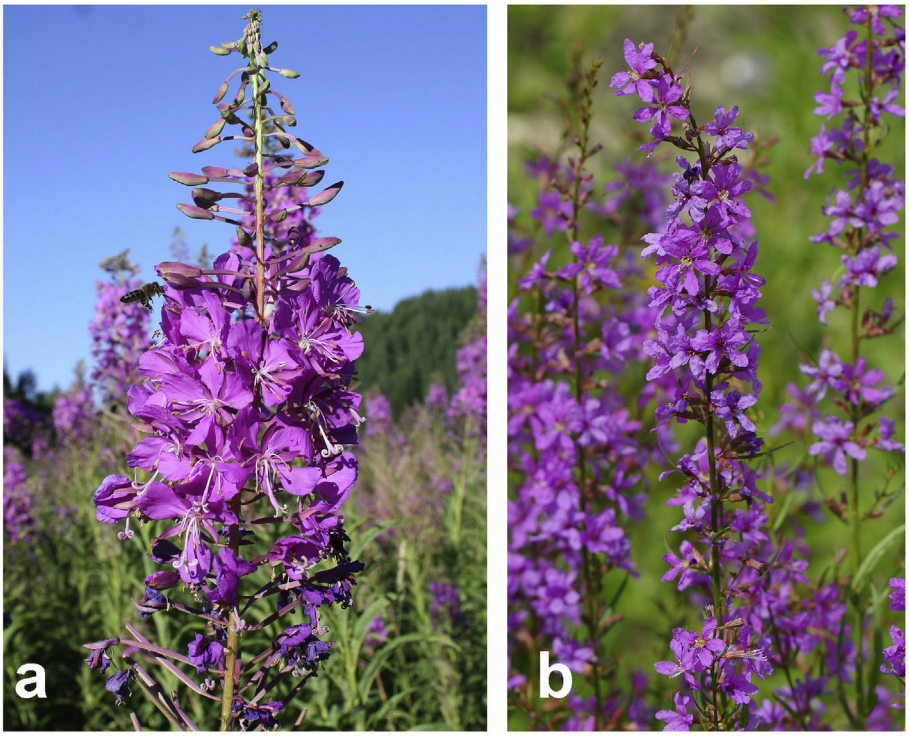

Fig. 1. a - Epilobium angustifolium, b - Lythrum virgatum.

which is among the top 10% highest in plants) in nrITS1 and nrITS2. High GC-content is known to adversely affect PCR (even completely blocking it); this requires special optimizations of the PCR protocol. Additionally, GC-rich regions lead to a pronounced bias effect in Illumina and Ion Torrent (PGM) sequencing technologies (Laehnemann et al., 2016; Nakamura et al., 2011). Such errors can obscure the taxonomic attribution of a read placing it in the undetermined category. Thus, we suggest that *Acorus* and *Equisetum* are absent due to the limitations of the experimental protocols. *Crataegus*, *Valeriana*, *Juniperus* and *Rosa* are claimed to be present in many mixes (*Valeriana* in 5 out of 6, *Crataegus* and *Rosa* in all of them) but are absent in all (*Crataegus* and *Valeriana*) or part (*Rosa*) of them. *Rosa* was found in 3 out of 6 mixes, which indicates that the issues with amplification, as described above, are not the case. The part of the plant that should be present in mixes for *Crataegus* and *Rosa* is fruit (and a fleshy berry-like cone for *Juniperus*). The fruits are quite large (0.7–2 cm in diameter) and there is no indication on the label if the fruits were ground by the manufacturer before packaging. If they were not ground sufficiently finely or not ground at all, they could be distributed unevenly during the packing of tea bags and absent in the teabag taken for the analysis. *Crataegus* and *Rosa* have been detected in the results of sequencing for several mixes, but under threshold level and thus classified as "not found".

Another issue revealed by metabarcoding is the presence of undeclared components. In addition to the *Lythrum* issue (discussed above), the most notable and most abundant of these components is *Convolvulus* (bindweed). Many species of this genus are invasive weeds. It was found in 4 mixes out of 6. Bindweed is a creeping or climbing plant that binds around the stems of other plants. The most common species, *C. arvensis* L., is toxic to animals (and presumably, to humans) due to the presence of tropane alkaloids (Schultheiss, Knight, Traub-Dargatz, Todd, & Stermitz, 1995). Its occurrence in pastures causes gastrointestinal disorders in horses (Todd, Stermitz, Schultheis, Knight, & Traub-Dargatz, 1995). Two of the analyzed herbal teas contain 4–6% of *Ambrosia* (ragweed), and this plant was detected in three additional teas, but at an abundance below the threshold. The most common species of this genus is *A. artemisiifolia* L., which is a widespread invasive weed; additionally, its pollen is highly allergenic (Taramarcaz, Lambelet, Clot, Keimer, & Hauser, 2005). The presence of *Ambrosia* pollen in *Mix1* was also verified by palynological analysis. The most likely reason for the *Convolvulus* and *Ambrosia* presence in herbal teas is the insufficient quality control of the collected raw plant material that missed plants tangled with bindweed and ragweed (or covered in its pollen). There are also less presented undeclared components, namely, *Polygonum*, *Viola*, *Urtica* and *Echinacea*, which are medicinal plants, including those used for other herbal tea products from the same manufacturer. Most likely their presence caused by cross-contamination during the production process.

Compared to other methods involving the analysis of complex plant mixes, in particular botanical analysis, which was used for one of the samples in this study, HTS offers a fast, high resolution method that is not limited by the type of the material or the degree of its homogenization and is easily scalable to an industrial scale. One of its major drawbacks is the price. It is still relatively high, but in view of the trend for the reduction of sequencing costs, e.g., ("DNA Sequencing Costs," 2017), this will not be the case in the near future.

5. Conclusion

HTS technologies offer a rapid and reliable method for the analysis of species composition in food and, although different in preparation protocols, specifications and known systematic biases, the two HTS platforms (Illumina and Ion Torrent) show similar results in this work, both qualitatively and quantitatively. In application to herbal teas, they show congruent results on the substitution of declared components and the presence of unlabeled

plants. However, there are certain challenges (see discussion on *Acorus* and *Equisetum*) that require further development of experimental and bioinformatic protocols to increase the resolution and quantitativeness of HTS-based solutions for food quality analysis.

Funding

This study was supported by Ministry of Education and Science of the Russian Federation, project # 14.609.21.0101 (unique identifier RFMEFI60917X0101).

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are grateful to Alexey Seregin (Lomonosov Moscow State University) for providing samples of *Lythrum* species for sequencing, Shirley A. Graham (Missouri Botanical Garden) for useful discussion on *Lythrum*, to Tamara Rib for the permission to use her photos and to American Journal Experts company for copyediting of the manuscript. This study was carried out using resources of the Skoltech Genomics Core Facility.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.foodcont.2018.04.040.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Baldwin, B. G. (1992). Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the compositae. *Molecular Phylogenetics and Evolution, 1*(1), 3–16. https://doi.org/10.1016/1055-7903(92)90030-K.

Bell, K. L., Fowler, J., Burgess, K. S., Dobbs, E. K., Gruenewald, D., Lawley, B., … Brosi, B. J. (2017). Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. *Applications in Plant Sciences, 5*(6). https://doi.org/10.3732/apps.1600124.

Boer, H. J. de, Ghorbani, A., Manzanilla, V., Raclariu, A.-C., Kreziou, A., Ounjai, S., … Gravendeel, B. (2017). DNA metabarcoding of orchid-derived products reveals widespread illegal orchid trade. *Proceedings of the Royal Society Series B, 284*(1863), 20171182. https://doi.org/10.1098/rspb.2017.1182.

Bragg, L. M., Stone, G., Butler, M. K., Hugenholtz, P., & Tyson, G. W. (2013). Shining a light on dark Sequencing: Characterising errors in Ion torrent PGM data. *PLoS Computational Biology, 9*(4). https://doi.org/10.1371/journal.pcbi.1003031.

Broström, A., Nielsen, A. B., Gaillard, M.-J., Hjelle, K., Mazier, F., Binney, H., … Sugita, S. (2008). Pollen productivity estimates of key european plant taxa for quantitative reconstruction of past vegetation: A review. *Vegetation History and Archaeobotany, 17*(5), 461–478. https://doi.org/10.1007/s00334-008-0148-8.

Brozynska, M., Furtado, A., & Henry, R. J. (2014). Direct chloroplast Sequencing: Comparison of sequencing platforms and analysis tools for whole chloroplast barcoding. *PLoS One, 9*(10), e110387. https://doi.org/10.1371/journal.pone.0110387.

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., … L, M. J. (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics, 25*(11), 1422–1423. https://doi.org/10.1093/bioinformatics/btp163.

Cupp, M. J. (2000). *Toxicology and clinical pharmacology of herbal products*. Springer Science & Business Media.

Devey, D. S., Chase, M. W., & Clarkson, J. J. (2009). A stuttering start to plant DNA barcoding: Microsatellites present a previously overlooked problem in non-coding plastid regions. *Taxon, 58*(1), 7–15.

DNA Sequencing Costs. (2017, October 31). *Data*. Retrieved March 4, 2018, from https://www.genome.gov/27541954/DNA-Sequencing-Costs-Data.

Eck-Varanka, B., Kováts, N., Hubai, K., Paulovits, G., Ferincz, Á., & Horváth, E. (2015). Genotoxic effect of Lythrum salicaria extract determined by the mussel micronucleus test. *Acta Biologica Hungarica, 66*(4), 460–463. https://doi.org/10.1556/018.66.2015.4.10.

Erdtman, G. (1960). The acetolysis method-a revised description. *Svensk Botanisk Tidskrift, 54*, 516–564. Retrieved from https://ci.nii.ac.jp/naid/10027358249/.

Fazekas, A., Steeves, R., & Newmaster, S. (2010). Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques, 48*(4), 277–285. https://doi.org/10.2144/000113369.

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics, 28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

Granica, S., Piwowarski, J. P., Czerwińska, M. E., & Kiss, A. K. (2014). Phytochemistry, pharmacology and traditional uses of different Epilobium species (Onagraceae): A review. *Journal of Ethnopharmacology, 156*, 316–346. https://doi.org/10.1016/j.jep.2014.08.036.

Information System on Plant Morphology and Anatomy. (2018). Retrieved March 4, 2018, from http://botany-collection.bio.msu.ru/.

Jakovlev, G. P. (2013). *Lekarstvennoe syr'e rastitel'nogo i zhivotnogo proishozhdenija: Uchebnoe posobie. Farmakognozija [Raw medicine materials of plant and animal origin: A workbook. Pharmacognosy]* (3rd ed.). Saint-Peterburg: SpecLit.

Laehnemann, D., Borkhardt, A., & McHardy, A. C. (2016). Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. *Briefings in Bioinformatics, 17*(1), 154–179. https://doi.org/10.1093/bib/bbv029.

Lahens, N. F., Ricciotti, E., Smirnova, O., Toorens, E., Kim, E. J., Baruzzo, G., … Grant, G. R. (2017). A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genomics, 18*(602). https://doi.org/10.1186/s12864-017-4011-0.

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., et al. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology, 30*(5), 434–439. https://doi.org/10.1038/nbt.2198.

Lo, Y.-T., & Shaw, P.-C. (2018). DNA-based techniques for authentication of processed food and food supplements. *Food Chemistry, 240*, 767–774. https://doi.org/10.1016/j.foodchem.2017.08.022.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal, 17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200.

Meacham, F., Boffelli, D., Dhahbi, J., Martin, D. I., Singer, M., & Pachter, L. (2011). Identification and correction of systematic error in high-throughput sequence data. *BMC Bioinformatics, 12*(451). https://doi.org/10.1186/1471-2105-12-451.

Mishra, P., Kumar, A., Nagireddy, A., Mani, D. N., Shukla, A. K., Tiwari, R., et al. (2016). DNA barcoding: An efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnology Journal, 14*(1), 8–21. https://doi.org/10.1111/pbi.12419.

Molly, M. (2017). *Preparation of 1.5 mg/mL Sera-mag carbolylate modified magnetic particles*. https://doi.org/10.17504/protocols.io.g2abyae. Retrieved from.

Moore, P. D., Webb, J. A., & Collinson, M. E. (1995). *Pollen analysis* (2nd ed.). Oxford: Blackwell Science.

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., … Kanaya, S. (2011). Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Research, 39*(13), e90. https://doi.org/10.1093/nar/gkr344.

National Plant Data Team, Greensboro, NC 27401–4901 USA. (2018). *USDA, NRCS. 2018. The plants database*. Retrieved March 4, 2018, from https://plants.usda.gov.

Newmaster, S. G., Grguric, M., Shanmughanandhan, D., Ramalingam, S., & Ragupathy, S. (2013). DNA barcoding detects contamination and substitution in North American herbal products. *BMC Medicine, 11*(222). https://doi.org/10.1186/1741-7015-11-222.

PalDat. (2018). Retrieved March 4, 2018, from https://www.paldat.org/.

Piwowarski, J. P., Granica, S., & Kiss, A. K. (2015). Lythrum salicaria L.—underestimated medicinal plant from European traditional medicine. A review. *Journal of Ethnopharmacology, 170*, 226–250. https://doi.org/10.1016/j.jep.2015.05.017.

Prosser, S. W. J., & Hebert, P. D. N. (2017). Rapid identification of the botanical and entomological sources of honey using DNA metabarcoding. *Food Chemistry, 214*, 183–191. https://doi.org/10.1016/j.foodchem.2016.07.077.

Raclariu, A. C., Paltinean, R., Vlase, L., Labarre, A., Manzanilla, V., Ichim, M. C., … Boer, H. de (2017). Comparative authentication of Hypericum perforatum herbal products using DNA metabarcoding, TLC and HPLC-MS. *Scientific Reports, 7*(1), 1291. https://doi.org/10.1038/s41598-017-01389-w.

Reille, M. (1999). *Pollen et spores d'Europe et d'Afrique du nord* (2nd ed.). Marseille: Laboratoire de botanique historique et palynologie.

Salipante, S. J., Kawashima, T., Rosenthal, C., Hoogestraat, D. R., Cummings, L. A., Sengupta, D. J., … Hoffman, N. G. (2014). Performance comparison of Illumina and Ion torrent next-generation sequencing platforms for 16S rRNA-Based bacterial community profiling. *Applied and Environmental Microbiology, 80*(24), 7583–7591. https://doi.org/10.1128/AEM.02206-14.

Saslis-Lagoudakis, C. H., Bruun-Lund, S., Iwanycki, N. E., Seberg, O., Petersen, G., Jäger, A. K., et al. (2015). Identification of common horsetail (*Equisetum arvense* L.; equisetaceae) using thin layer chromatography versus DNA barcoding. *Scientific Reports, 5*, 11942. https://doi.org/10.1038/srep11942.

Schirmer, M., Ijaz, U. Z., D'Amore, R., Hall, N., Sloan, W. T., & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research, 43*(6), e37. https://doi.org/10.1093/nar/gku1341.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics, 27*(6), 863–864. https://doi.org/10.1093/bioinformatics/btr026.

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., …,

Consortium, F. B. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences, 109*(16), 6241–6246. https://doi.org/10.1073/pnas.1117018109.

Schultheiss, P. C., Knight, A. P., Traub-Dargatz, J. L., Todd, F. G., & Stermitz, F. R. (1995). Toxicity of field bindweed (Convolvulus arvensis) to mice. *Veterinary & Human Toxicology, 37*(5), 452–454. Retrieved from http://europepmc.org/abstract/med/8592835.

Sickel, W., Ankenbrand, M. J., Grimmer, G., Holzschuh, A., Härtel, S., Lanzen, J., … Keller, A. (2015). Increased efficiency in identifying mixed pollen samples by meta-barcoding with a dual-indexing approach. *BMC Ecology, 15*(20). https://doi.org/10.1186/s12898-015-0051-y.

Stoeckle, M. Y., Gamble, C. C., Kirpekar, R., Young, G., Ahmed, S., & Little, D. P. (2011). Commercial teas highlight plant DNA barcode identification successes and obstacles. *Scientific Reports, 1*(42). https://doi.org/10.1038/srep00042.

Taramarcaz, P., Lambelet, C., Clot, B., Keimer, C., & Hauser, C. (2005). Ragweed (Ambrosia) progression and its health risks: Will Switzerland resist this invasion? *Swiss Medical Weekly, 135*(37–38), 538–548.

The State Pharmacopoeia of the Russian Federation. (2015) (XIII, Vols. 1–3). Retrieved from http://www.femb.ru/feml.

Todd, F. G., Stermitz, F. R., Schultheis, P., Knight, A. P., & Traub-Dargatz, J. (1995). Tropane alkaloids and toxicity of Convolvulus arvensis. *Phytochemistry, 39*(2), 301–303. https://doi.org/10.1016/0031-9422(94)00969-Z.

Tunalier, Z., Koşar, M., Küpeli, E., Çaliş, İ., & Başer, K. H. C. (2007). Antioxidant, anti-inflammatory, anti-nociceptive activities and composition of Lythrum salicaria L. extracts. *Journal of Ethnopharmacology, 110*(3), 539–547. https://doi.org/10.1016/j.jep.2006.10.024.

Upton, R., Graff, A., Jolliffe, G., Länger, R., & Williamson, E. (Eds.). (2016). *American herbal pharmacopoeia: Botanical pharmacognosy-microscopic characterization of botanical medicines*. CRC Press.

Vandyshev, V. V., Babaeva, E. Y., & Miroshnikova, E. A. (2017). *Standartizacija kachestva i metody ustanovlenija podlinnosti izmel'chennogo lekarstvennogo rastitel'nogo syr'ja i sborov. Farmakognozija [Standardization of quality and methods of authentication of grinded plant raw material and herbal collections. Pharmacognosy]* (4th ed.). Moscow: Peoples' Friendship University of Russia.

Vassou, S. L., Kusuma, G., & Parani, M. (2015). DNA barcoding for species identification from dried and powdered plant parts: A case study with authentication of the raw drug market samples of Sida cordifolia. *Gene, 559*(1), 86–93. https://doi.org/10.1016/j.gene.2015.01.025.

White, T. J., Bruns, T., Lee, S., & Taylor, J. L. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A Guide to Methods and Applications, 18*(1), 315–322.

Xin, T., Yao, H., Gao, H., Zhou, X., Ma, X., Xu, C., … Chen, S. (2013). Super food Lycium barbarum (Solanaceae) traceability via an internal transcribed spacer 2 barcode. *Food Research International, 54*(2), 1699–1704. https://doi.org/10.1016/j.foodres.2013.10.007.

Xu, M., Heidmarsson, S., Thorsteinsdottir, M., Kreuzer, M., Hawkins, J., Omarsdottir, S., et al. (2018). Authentication of Iceland Moss (Cetraria islandica) by UPLC-QToF-MS chemical profiling and DNA barcoding. *Food Chemistry, 245*, 989–996. https://doi.org/10.1016/j.foodchem.2017.11.073.

Yeo, Z. X., Chan, M., Yap, Y. S., Ang, P., Rozen, S., & Lee, A. S. G. (2012). Improving indel detection specificity of the Ion Torrent PGM benchtop sequencer. *PLoS One, 7*(9), e45798. https://doi.org/10.1371/journal.pone.0045798.