Heliyon 5 (2019) e01935

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.heliyon.com

Exploring DNA quantity and quality from raw materials to botanical extracts

Subramanyam Ragupathy [a,*], Adam C. Faller [a], Dhivya Shanmughanandhan [a], Prasad Kesanakurti [a], R. Uma Shaanker [b], Gudasalamani Ravikanth [c], Ramalingam Sathishkumar [d], Narayanasamy Mathivanan [e], Jingyuan Song [f], Jianping Han [f], Steven Newmaster [a,**]

[a] NHP Research Alliance, College of Biological Sciences, University of Guelph, Guelph, Ontario, N1G 2W1, Canada
[b] Department of Crop Physiology and School of Ecology and Conservation, University of Agricultural Sciences, GKVK, Bangalore, 560065, India
[c] Conservation Genetics Lab, Ashoka Trust for Research in Ecology and the Environment (ATREE), Royal Enclave, Srirampura, Jakkur PO, Bengaluru, 560064, India
[d] Plant Genetic Engineering Laboratory, Department of Biotechnology, Bharathiar University, Coimbatore, Tamil Nadu, 641046, India
[e] Centre for Advanced Studies in Botany, University of Madras Guindy Campus, Chennai, 600 025, Tamil Nadu, India
[f] Engineering Research Center of Traditional Chinese Medicine Resource, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, 100193, China

ARTICLE INFO

Keywords:
Food science
Molecular biology
DNA quality
DNA quantity
Botanicals

ABSTRACT

Objectives: The aim of this study was to explore the variability in DNA quality and quantity along a gradient of industrial processing of botanical ingredients from raw materials to extracts.
Methods: A data matrix was assembled for 1242 botanical ingredient samples along a gradient of industrial processing commonly used in the Natural Health Product (NHP) industry. Multivariate statistics was used to explore dependant variables for quality and quantity. The success of attaining a positive DNA test result along a gradient of industrial processing was compared among four biotechnologies: DNA barcoding, NGS, Sanger sequencing and qPCR.
Results: There was considerable variance in DNA quality and quantity among the samples, which could be interpreted along a gradient from raw materials with greater quantities (50–120 ng/μL) of DNA and longer DNA (400-500bp) sequences to extracts, which were characterized by lower quantities (0.1–10.0 ng/μL) and short fragments (50-150bp).
Conclusions: Targeted molecular diagnostic tests for species identity can be used in the NHP industry for raw and processed samples. Non-targeted tests or the use of NGS for any identity test needs considerable research and development and must be validated before it can be used in commercial operations as these methods are subject to considerable risk of false negative and positive results. Proper use of these tools can be used to ensure ingredient authenticity, and to avert adulteration, and contamination with plants that are a health concern. Lastly these tools can be used to prevent the exploitation of rare herbal species and the harvesting of native biodiversity for commercial purposes.

1. Introduction

Consumer confidence in herbal medicine has been challenged by reports of adulteration in the Natural Health Product (NHP) industry. These reports are based on scientific studies of commercial NHPs in the market place and have been published in peer-reviewed journals (Cheng et al., 2015; Gao et al., 2017; Han et al., 2016; Newmaster et al., 2013; Palhares et al., 2015; Raclariu et al., 2017; Shanmughanandhan et al., 2016). Although these studies are founded on good science, the focus in the scientific literature and uptake in the public media has taken a critical view of botanical industry; considerable adulteration was reported from marketplace testing (De Boer, Ichim and Newmaster, 2015). This is unfortunate because there exists traditional and scientific knowledge, including evidence-based health benefits for many herbal medicines, that offers the consumer an alternative to pharmaceuticals (Alvari et al., 2012; Chikezie and Ojiako, 2015; Ernst, 2005; Izzo et al., 2016).

In order to gain consumer confidence, the NHP industry must focus on advancing the development of herbal medicine by addressing the

* Corresponding author.
** Corresponding author.
E-mail address: ragu@uoguelph.ca (S. Ragupathy).

https://doi.org/10.1016/j.heliyon.2019.e01935
Received 2 February 2019; Received in revised form 23 May 2019; Accepted 6 June 2019

ingredient authentication issues facing the NHP industry. There are key uncertainties in quality control of the supply of botanical ingredients that could be addressed by science-based research and the development of innovative solutions for quality assurance (De Boer et al., 2015; Woolfe and Primrose, 2004). One of the key challenges in the NHP industry is validating the identity of species ingredients. There is considerable scientific evidence from published market place studies to suggest that the current quality control methods using morphological, microscopy and analytical chemistry are not sufficient for confirming the identity of many botanical ingredients (Ananingsih et al., 2013; Chen et al., 2014; Scora, 1966). Morphological methods are often not possible on starting materials that lack the presence of taxonomic characters needed to identify plant species. Microhistology based identification using microscopy is limited to a few species and the presence of anatomical characters, which must be interpreted by highly skilled technicians. Analytical chemistry is the most commonly used tool to validate botanical species ingredient identity in the NHP industry. Testing protocols such as HPLC, MS and TLC have been designed to identify chemical profiles and not necessarily species ingredients. It is very difficult to use analytical chemistry methods on processed materials due the lack of standard reference materials for processed botanicals and the lack of research on impact of processing on plant metabolites. The limited studies that exist suggest that targeted metabolites in plants are altered during the processing of NHPs, resulting in considerable variability in the test results or complete failure of test methods (Ananingsih et al., 2013). Correlations between metabolites and plant species are often assumed to be evidence for species identification, which is not an acceptable science-based approach because the presence of a phytochemical does not infer phylogenetic species concepts (Newmaster and Ragupathy, 2010; Urumarudappa et al., 2016). DNA is inherent and reflects the evolutionary processes that define species. This is not the case for plant chemicals of which most are common to more than one species or may be highly variable within species based on environmental mechanisms (Ananingsih et al., 2013). A recent solution is DNA-based biotechnology, which has been suggested as an additional tool for identity of botanical species ingredients in order to ensure quality assurance and control in the supply chain of NHPs (De Boer et al., 2015; Newmaster and Ragupathy, 2010; Orhan et al., 2016; Raclariu et al., 2017; Sucher and Carles, 2008).

Although DNA-based tools offer potential solutions for testing raw materials in the botanical supply chain, there are considerable research gaps that need to be addressed in order to test processed botanical ingredients. DNA methods are well established as forensic tools for species identity and as a standard diagnostic tool for probiotics and food borne pathogens in commercial sectors (Davis, 2014; Harris and Griffiths, 1992; Zhao et al., 2014). There are many tools to consider of which some are appropriately validated, and others are not fit for the purpose of species ingredient authentication. Although DNA barcoding has been suggested as a tool for authentication of botanical ingredients, this method is now quite antiquated and is limited by several issues related the processing of botanical ingredients; long sequence reads and limited to analysis of two regions. Next Generation Sequencing (NGS) also has great promise in the identification of multiple non-targeted species ingredients within one sample, but there are considerable issues that need to be addressed before it can be used as a legitimate commercial tool. For example, the lack of positive controls, inaccurate estimates of sequence reads (quantification may be biased and needs to be validated using dilution series experiments, which are lacking), considerable PCR bias and bioinformatic issues are causing a considerable number of false positives for contaminants in both raw and in processed materials (Burns et al., 2016; Coissac et al., 2012; Schrijver et al., 2012; Song et al., 2008). The key uncertainty in these methods highlights the lack of knowledge on the quantity and quality of DNA recovered from processed botanical ingredients. Raw materials have sufficient DNA that is of good quality for both DNA barcoding and in some cases NGS. This claim is supported by several studies (Cheng et al., 2015; Gao et al., 2017; Han et al., 2016; Lu et al., 2018; Newmaster et al., 2013; Palhares et al., 2015; Raclariu et al., 2017; Seethapathy et al., 2015; Shanmughanandhan et al., 2016) that have demonstrated the ability to recover high quality DNA sequences from raw materials. Some industry members have postulated that highly processed NHPs may not have any DNA present. Recent research by Lu et al. (2018) demonstrated the ability to visualize DNA in highly processed botanicals using adapter-ligation and PCR amplification methods. Faller et al. (2019) further demonstrated the identity of target species ingredients in highly processed extracts for green tea NHPs; he found short DNA fragments of high quality DNA that is detectable for identification. However, this research is limited to only a few botanicals leaving a gap in the breadth and depth of our knowledge of how mechanical processing such as grinding into powders or sanitization (e.g., steam treatment) affects the quality and quantity of DNA. This basic understanding is required in order to formulate proper hypotheses and statistical models required for assessing the probability of false negatives/positives in commercial test assays. A suite of these tools needs to be developed, optimized and tested in validation studies along a gradient of industrial processing in order to assess the applicability of commercialization of these tools and potential use in regulatory oversight.

The goal of this study is to explore the variability in DNA quality and quantity along a gradient of industrial processing of a considerable number of botanical ingredients from raw materials to botanical extracts. This seeks to provide breadth and depth in botanical taxa within an experimental model of increasing intensity of industrial processing. More specifically we deployed multivariate statistics to explore dependant variables for quality (fragment length = number of base pairs) and quantity (DNA content ng/μL) among 1242 botanical samples along a gradient of industrial processing. These samples were also analyzed using several DNA-based biotechnologies in order to assess the success of attaining a positive DNA test results from raw botanical materials to extracts. This exploratory research addressed the following questions 1) What is the variability in DNA quality and quantity along a gradient of industrial processing of botanical ingredients, and 2) how successful are different DNA-based tools (DNA Barcoding, NGS, Mini-sequence, Nucleotide Signature) in attaining positive identity results along a gradient of NHP industrial processing from the farm to finished products available to consumers.

2. Methods

A multivariate data matrix was assembled from 1242 botanical samples representing 275 medicinal plant species. Two dependant variables, DNA quantity (ng/μL) and quality (fragment size in bp) were assessed for each of the samples. Four independent variables represented a gradient of industrial treatments including 1) Raw material from the farm, 2) Powder (grinding of raw material), 3) Sanitization (steam treatment), and 4) Extraction (extractive distillation, supercritical fluid extraction). Each of the four levels of treatment contained variability among the class of treatment as defined by each manufacture.

2.1. DNA quantity and quality

2.1.1. DNA extraction

DNA was extracted using the Macherey-Nagel NucleoSpin® Plant II "Genomic DNA from Plant" Kit, from botanical materials (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Extractions of samples used 60 mg of dry starting material and were eluted in 60 μL of elution buffer. Protocol was followed as according to the manufacturer's instructions.

2.1.2. DNA quantity (volume; ng/μL)

DNA was quantified using a Qubit™ dsDNA High Sensitivity Assay Kit on the Qubit™ 3.0 Fluorometer according to the manufacturer's instructions. All ng/μL concentration readouts from the fluorometer were based on NHP DNA extract samples eluted in 60 μL of elution buffer.

2.1.3. DNA quality (fragment size; number of base pairs)

Fragment size was estimated using an Agilent 2100 Bioanalyzer to visualize size distribution in both smaller ranges of fragment sizes (specifically 25–1000 bp; using the DNA 1000 chip), as well as a larger 50–7000 bp range (using the High Sensitivity DNA chip) (Agilent Technologies, Santa Clara, CA). This estimate was further verified using a 4200 TapeStation instrument with a Genomic DNA ScreenTape (D1000) and reagents (Agilent, Santa Clara, CA, USA).

2.2. Assessing DNA based tools

The relative test results from four different molecular diagnostic tools were compared including the following: DNA Barcoding; Next Generation Sequencing; Mini-sequencing (Sanger sequencing); Nucleotide signatures (DNA Probes and qPCR). Each of these tools were tested on NHPs along a gradient of industrial treatments including: 1) raw material, 2) grinding into powder, 3) sanitization (steam treatment), and 4) extraction (extractive distillation, supercritical fluid extraction). The relative success was compared among the technologies, with "success" defined as producing a quality DNA sequence. These were all predefined tests designed to match a targeted botanical species ingredient and were named "positive test" if they matched the targeted botanical ingredient.

2.2.1. DNA barcoding

Two DNA regions (rbcL, and matK) were selected based on the recommendations by CBOL Plant Working Group (2009) as the official plant barcoding markers. We isolated total genomic DNA from approximately 10 mg of dried leaf material from each sample using the NucleoSpin® Plant II "Genomic DNA from Plant" kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Extracted DNA was stored in sterile microcentifuge tubes at –20 °C. The selected loci were amplified by polymerase chain reaction (PCR) on a PTC–100 thermocycler (Bio-Rad). DNA was amplified in 20 μL reaction mixtures containing 1 U AmpliTaq Gold Polymerase with GeneAmp 106 PCR buffer II (100 mm Tris-HCl pH 8.3, 500 mm KCl) and 2.5 mm MgCl2 (Applied Biosystems), 0.2mm dNTPs, 0.1mm of each primer (0.5mm for *matK*), and 20 ng template DNA. Amplified products were sequenced in both directions with the primers used for amplification, following the protocols of the University of Guelph Genomics facility (www.uoguelph.ca/~genomics). Products from each specimen were cleaned using Sephadex columns and run on an ABI 3730 sequencer (Applied Biosystems). Bidirectional sequence reads were obtained for all the PCR products. Sequences were assembled using Sequencher 4.5 (Gene Codes Corp), and aligned manually using Bioedit version 7.0.9.

2.2.2. Next generation sequencing

Library preparation was performed at the Genotypic Technology's Genomics facility. The PCR amplified product was checked on an agarose gel before proceeding to PCR indexing. The PCR index reaction was conducted using a Nextera® XT Index Kit v2 Set D kit from which Illumina sequencing adapters and dual indexing barcodes were added using limited cycle PCR (Initial Denaturation 95 °C for 5min, Followed by 10 cycles of 98 °C for 20s, 64 °C for 15s, 72 °C for 35s, a final extension of 72 °C for 5min). This provided a final product of ~520bp and ~570bp. The library was cleaned using HighPrep PCR magnetic beads (HighPrep PCR, Magbio, Switzerland) and was Qubit quantified using Qubit ds DNA HS kit (Invitrogen, USA). Qubit quantification resulted in concentration of 1.91 ng/μl with total yield of 19.1 ng. The prepared library was then validated for quality using High Sensitivity Bioanalyzer Kit (Agilent Technologies, USA) by running an aliquot on High Sensitivity Bioanalyzer Chip (Agilent Technologies, USA). The library was then pooled for Illumina MiSeq sequencing using the 250bp paired end read chemistry.

2.2.3. Mini-sequence analysis

Botanical species (275) as represented in the University of Guelph NHP Research Alliance Standard Biological Reference Material (SBRM) DNA library; Mini sequences were on average 100–120 bp and were sequenced using a tiered approach (*rbcL*; *ITS2*) (Newmaster et al., 2006). All triplicate samples of plant product (powder) from herbal products were sequenced. Total genomic DNA was isolated from 100 mg of the dry voucher specimen and 60 mg of the herbal product using the NucleoSpin_Plant II Mini DNA Extraction kit (Macherey-Nagel, Düren, Germany). Additionally, an extraction control was used to monitor the DNA extraction efficiency. The polymerase chain reaction (PCR) amplification was carried out using *rbcL* and *ITS2* universal primers in a 10 l μl reaction mixture that contained 0.001–20 ng of genomic DNA, 1.2 X of 10 X Pfu buffer, 3 mM of MgSO4 (Fermentas, Waltham, Massachusetts, USA), 250 l M of 2 mM dNTPs (Fermentas), 0.2 l M each of forward and reverse primers (10 l M), 0.5 U of 2.5 U Pfu DNA polymerase (Fermentas) and 0.5 % dimethyl sulfoxide (DMSO). A negative control was maintained for all PCR reactions. The primers and the reaction conditions for *rbcL* and *ITS2* were taken from Chen et al. (2010) and Fazekas et al. (2012). The PCR products obtained were outsourced for DNA sequencing. The amplicons were bi-directionally sequenced using ABI PRISM_377 sequencer (Applied Biosystems, Foster City, CA, USA). The chromatographic traces were aligned and codon read in the CodonCode Aligner version 3.0 (CodonCode, Centerville, MA, USA) and contigs were generated. Sequences were deposited into a local botanical reference material (RM) library database.

2.2.4. Nucleotide signature qPCR

The sequences with high sequence similarity to each candidate reference genes were aligned separately to highlight potential areas of polymorphism. The primer pairs were designed by PrimerSelect of Lasergene 8 (DNASTAR, Inc., Madison, WI, USA) following the recommendations by Udvardi et al. (2008), and primer pairs were selected from sequence region with the fewest polymorphisms. The amplification specificity of designed primers was confirmed via standard qPCR procedures under different annealing temperature (50, 53, 56 and 59 °C). PCR amplification efficiency was calculated using a two-fold serial dilution of the cocktail of DNA templates. All PCR products were examined by melting curve analysis and agarose gel electrophoresis.

2.3. Multivariate analysis

The relationship of the classification structure of the multivariate matrix of DNA quality and quantity along a gradient of industrial processing (details defined above) was analyzed with nonmetric multidimensional scaling (NMS) using the "R" software (Kruskal, 1964; Maechler et al., 2013; Newmaster et al., 2008; R Core Team, 2014). In NMS, the Bray-Curtis distance measure was used because of its robustness for both large and small scales on the axes. Data were standardized by variable maxima and two-dimensional solutions were appropriately selected based on plotting a measure of fit ("stress") to the number of dimensions. Stress represents distortion in the data, whereby a stress value over 0.15 indicates that the results are invalidated. One thousand iterations were used for each NMS run, using random start coordinates. The first 2 ordination axes were rotated to enhance interpretation with the different axes. Differences in the membership of clusters were analyzed using discriminant analysis and MRPP.

3. Results

3.1. DNA quantity and quality

There was considerable variation in the multivariate matrix of DNA quality and quantity from 1242 samples. The NMS ordination (Fig. 1) was constructed in a 2-dimensional structure with no additional variation on the third axes; relative eigenvalues did not change after the second axis (NMS axis 1 = 2.46; NMS axis 2 = 3.17; NMS axis 3 = 3.24; NMS stress value was 0.16). Variation in axes scores along NMS Axis one was

1.67 SD whereas variation in NMS axis 2 was 3.16 SD. The structure in the ordination resulted in several clusters of which sample results from raw materials were grouped in the top left corner of the ordination and those of extracts were on the bottom right hand corner. The sample results from ground samples (powder) and those subjected to manufacturing processing were oriented between the left and right sides of the ordination; sanitization tests overlapped with both powder samples and extracts. Two gradients were identified in the analysis including 1) DNA fragment size along axis 1 (Pearson Correlation 0.93; $p < 0.01$) with longer fragments in raw materials and shorter in extracts, and 2) DNA quantity along axis 2 in which more DNA was found in raw samples (highest 120 ng/μL) and less (0.01 ng/μL) in extracts (Fig. 2).

The distribution of DNA quantity was heavily skewed toward small amounts with increasing industrial processing (Fig. 2). Raw materials contained large amounts (120 ng/μL) of DNA with lesser amounts associated with specific species such as Aloe vera (1.8 ng/μl) or raw samples from roots/rhizomes (e.g., Ginseng). Most raw botanicals had over 50.0 ng/μL. Industrial grinding botanical samples into powder reduced the amount of DNA to <50.0 ng/μL. Sanitization such as steam treatment further reduced DNA quantity to <20.0 ng/μL. The extract process still resulted in the presence of DNA ranging from 0.1-10.0 ng/μL 3% of the samples failed to produce any measurable DNA (Fig. 3).

DNA quality as defined by fragment size is not evenly distributed along a gradient of industrial treatments. Large fragments (400-500bp) are associated with raw botanicals (Fig. 3). The first step in processing, which is grinding reduces fragment size followed by shorter fragments after product sanitization. The extract process results in many small fragments ranging from 50-150bp.

3.2. Assessing DNA based tools

There were considerable differences in the relative test results from different molecular diagnostic biotechnology. Although DNA barcoding was relatively successful on raw materials, it still failed to successfully identify 25% of the samples. DNA barcode failure increases with more intensive industrial processing with 90% failure in extracts. The results from next generation sequencing (NGS) were marginally better with considerable failure in extracts and the misidentification of incidental contaminant DNA as the dominant ingredient in ranging from 7% in raw materials to 41% of the extracts that were successful sequenced. False positives were validated by analytical chemistry. Mini-sequence methods produced results for all the raw materials and ground powder samples tested. The success of the mini-sequence method in extracts was twice that of NGS and six times greater than DNA barcoding. Targeted PCR methods using nucleotide signatures or DNA probes had the highest success with positive results for 76% of extracts. We were able to get positive results using targeted PCR in the samples that failed to produce any measurable DNA; this was not possible with the other three biotechnologies. Comparison of six commercial kits produced positive results ranging from 12-34% of extracts.

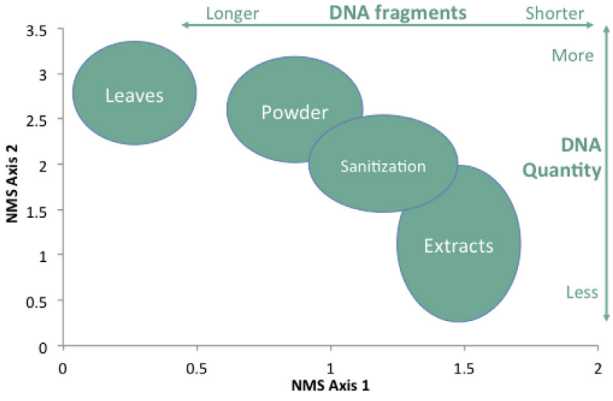

Fig. 1. NMS ordination of variation in DNA quantity and quality from 1242 samples along a gradient of industrial processing from raw botanicals to extracts.

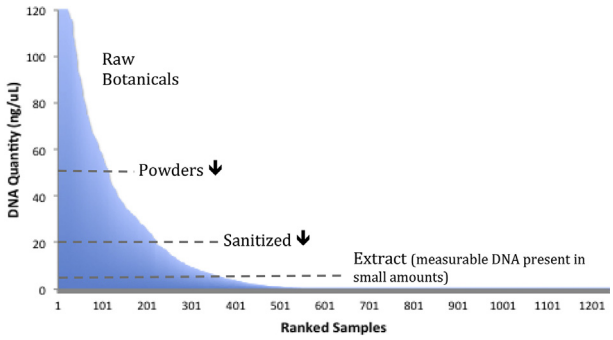

Fig. 2. Distribution of DNA Quantity for 1242 samples along a gradient of industrial processing from raw botanicals to extracts.

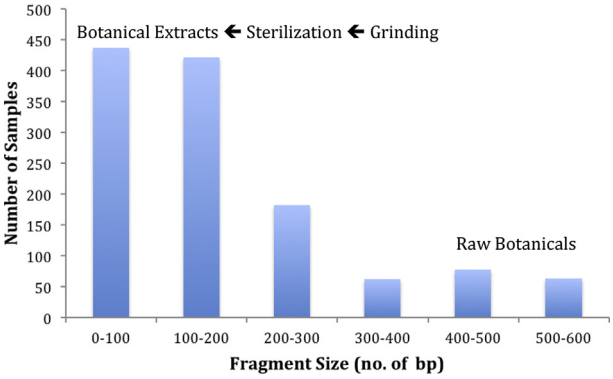

Fig. 3. DNA fragment size histogram for 1242 samples form a gradient of industrial processing from raw botanicals to extracts.

4. Discussion

Increasing demand for NHPs creates more pressure on supply and increases the probability of product adulteration and exploitation of native biodiversity. A considerable proportion (60%) of the world's population depends on traditional medicine, of which 80% of developing countries depend almost entirely on traditional medicine practices and herbal medicines as their primary health care (WHO, 2001). Rapid market growth at 7% per year has resulted in a global market value that is expected to reach $86–100 billion USD by 2022 (CRN, 2017; ZMR, 2017). Recent environmental concerns include the fact that some uncommon or rare species of plants are being harvested for commercial use as NHPs (Chen et al., 2016). There is immense pressure on the supply of botanical ingredients, which may explain why there is adulteration and product substitution levels of product ingredients (Cheng et al., 2015; Gao et al., 2017; Han et al., 2016; Newmaster et al., 2013; Palhares et al., 2015; Raclariu et al., 2017; Shanmughanandhan et al., 2016). The World Health Organization deemed the safety of NHPs a concern; with increasing use and reports of adverse reactions, regulation standards need to be reviewed to consider proper due diligence in quality assurance (WHO, 2004). There is an immediate need for more adequate tools to authenticate botanical supply chains for commercial use in the herbal industry.

Although analytical chemistry tools are already well-established tools for quality control in the NHP industry, the addition of molecular diagnostic tools to quality control systems should be considered in order to enhance efforts to mitigate adulteration. The discussions on whether to use analytical chemistry or DNA based methods for quality assurance should be based on a logical argument and relate to “fit for purpose”

testing within the quality control system. Analytical chemistry testing protocols have been designed to identify chemicals and not necessarily species ingredients. DNA is inherent and reflects the evolutionary processes that define species; this is not the case for chemicals, of which many are common to more than one species. We advocate the use of both methods for optimal quality control programs including the use of 1) chemical testing methods to validate the presence of chemical ingredients against validated standard reference materials, and 2) molecular diagnostic DNA-based methods to identify species ingredients based on validated methods with positive and negative controls. DNA methods are well established as forensic tools for species identity and as a standard diagnostic tool for probiotics and food borne pathogens (Davis, 2014; Law et al., 2015; Zhao et al., 2014). Molecular diagnostic tools such as species-specific nucleotide signatures using targeted methods on PCR platforms can be validated for commercial use. This molecular diagnostic biotechnology is poised to reinforce quality control systems against the risk of fraudulent product substitution, adulteration, contamination, and unlabelled fillers.

Industry claims that DNA is completely destroyed or eliminated by industrial processing, thus precluding DNA-based testing, are not supported by our research. Although we found a considerable gradient of lower quantity of DNA represented by short sequences with intensive industrial processing, we also demonstrated considerable success in obtaining positive DNA identity test results from herbal extracts; some industry stakeholders thought this was not possible. These results have a logical explanation founded on mechanistic hypotheses. Herbal ingredients are exposed to various processing procedures including techniques such as mechanical grinding, heat treatments, ultraviolent light exposure, filtration and fluid extraction. The mechanistic physical impact on the structure of the DNA following industrial processing is a plausible explanation for lower quantity and shorter fragments of DNA. For example, standard production protocols of NHPs typically include high heat (>100 °C) procedures for the purpose of drying, sanitization or sterilization (Woolfe and Primrose, 2004). Previous research has demonstrated that high temperatures such as those used in NHP processing are sufficient to break covalent bonds in dsDNA, and with prolonged exposure to temperatures above 100 °C, completely degrade DNA (Karni et al., 2013). Therefore, this would prohibit use of DNA based diagnostic techniques due to absence of DNA template for amplification and sequencing. Additionally, UV light, which can be used in NHP processing, is well known to be damaging to DNA via thymine dimer formation; structurally compromised DNA may also be unable to be effectively amplified or sequenced (Pehrson, 1989). Current research (Faller et al., 2019) has demonstrated that processing of green tea leaves during extract manufacturing results in shorter fragments of DNA (Faller et al., 2019). DNA extraction methods that yield longer strands such as precipitation methods (Malentacchi et al., 2016) may be used for optimal results when working with unprocessed plant materials. However, we have recently used new methods such as precipitation methods for barcoding and NGS, and do not see any improvement in the results. This is because the processing of NHP samples during manufacturing breaks up the DNA into small fragments; thus, there are no long fragments to recover. These mechanisms are useful in interpreting our results. We encourage further research based on experimental models that utilize industrial processing treatments to test these hypotheses.

There was considerable variance in the relative success of molecular diagnostic biotechnologies in testing NHP samples from a gradient of industrial processing from raw botanicals to extracts. DNA barcoding had the lowest success of all the molecular biotechnologies even when testing raw materials. This is likely due to the to several factors including but limited to the following 1) DNA barcoding by definition has limited success with only a couple (*rbcL* and *matK*) plastid regions (Fazekas et al., 2010; Percy et al., 2014), 2) one of the regions (*matK*) is known to have relatively low PCR success (Fazekas et al., 2008; Kress and Erickson, 2007; Newmaster et al., 2008), 3) both barcode regions are very long (>500bp) and therefore not found in most of the processed samples according to our research (Fazekas et al., 2012; Kress et al., 2005). We suggest that DNA barcoding *sensu stricto* is not used for authentication of botanical ingredients, as it will result in relatively low success in raw materials and very low success in processed materials.

Although the non-targeted approach of NGS makes it an attractive tool for potentially revealing any adulterant, NGS technology presents several serious issues that prevent its current use as a commercial tool. The lure of NGS is founded on the fact that it is an exciting research tool, as it is very sensitive and can be used to detect multiple sources of DNA in a single analysis. This makes NGS particularly useful for addressing certain research questions focused on recreating dietary diversity (Pompanon et al., 2012) or the study of traces of environmental DNA (eDNA) in water and soils (Rees et al., 2014). However, the published literature signifies there are considerable limitations and problems with NGS that present an immediate impediment to generating scientifically valid test results for commercial use (Bianchi et al., 2016; Coissac et al., 2012; Lighten et al., 2014; Waits and Paetkau, 2005). Notably, NGS results may indicate presence of species in a sample due only to detection of small amounts of small fragmented contaminant DNA of which it may over-estimate the number of sequences. We call these small DNA fragments “incidental DNA fragments”. It is reasonable to postulate that all food and NHP products have many small fragments of incidental DNA from a number of different species, particularly agricultural species. In fact, regulatory and pharmacopoeial botanical monographs specifically allow small amount (2–5%) of foreign organic matter, and inadvertent plant parts from co-mingled species in a harvested crop. Agricultural harvesting systems are not sterile environments and small amounts of DNA from surrounding vegetation will be present in all products. NGS resulted in a considerable number of false positives in our study. We suggest that this may be due to the fact that NGS is so sensitive it is detecting incidental DNA in commercial products and through mechanisms such as PCR bias overestimating the number of sequences resulting in false positive for adulterants. This type of test result is a very big problem for the industry, as they may have to reject a shipment, change a supplier or recall a product line. The relatively low success of NGS in processed products is due to the fact that the meta-barcoding NGS pipeline needs considerable research and development. A NGS pipeline standard operating protocol needs to be developed for a short DNA sequence library. Based on our research we suggest 80–100 bp. It would be difficult to go shorter than 80 bp because there needs to be space for primer design, and enough space to capture sequence variability to differentiate all the target and non-target species. This is a challenge considering there are over 1000 commoditized target botanical species and an undetermined number of adulterants and possible contaminants. This brings in a second very large issue with respect to variability in PCR bias among plant species (Morgante and Olivieri, 1993). Corrective algorithms for PCR bias can be created (Coissac et al., 2012; Reinecke et al., 2015) but need to be developed for a complete DNA sequence library. If one species is left out of the corrective bioinformatic model, then the results are dubious. These issues can be used to explain why NGS was relatively unsuccessful in our study and is not useful as a commercial test at this time. In fact, during the last year our research institute has reviewed over 100 commercial NGS tests from NHP products of which the results did not make any sense to the NHP industry member; further forensic investigation revealed that in every case there was either a false negative, false positive or in many instances both. Limits of detection (LOD) and quantification (LOQ) must be developed and validated to ensure that incidental DNA fragments present in every sample are below some threshold. This would allow an acceptable low level of contaminant DNA sequences in a quality assurance system and prevent false positives for adulterants that could incorrectly trigger a costly non-conformity process in a manufacturing facility. We suggest further research is needed to resolve these issues before NGS can be implemented as a tool in the quality control and identity of NHPs. A targeted molecular diagnostic tool should be properly validated (Newmaster et al., 2018) method including LOD/LOQ with defined acceptable thresholds for small

amounts of incidental DNA sequences commonly found in every substrate.

There are two approaches to developing a DNA-based assay for testing species identity of botanical ingredients. A targeted assay or defined test in which the test is designed to identify one species in a product at a time; multiple assays/tests can be utilized on one sample to confirm the identity of a mixture of species. A non-targeted test seeks to list all the species in a product including the main species ingredient, mixtures of species, adulterants, and contaminants. Although this test is highly desirable, it is only theoretically possible and has only been demonstrated within a controlled research projects in a lab. Currently it is not possible to design a non-targeted assay as it would need to be validated on a NGS platform, which is not possible for reasons previously discussed in this manuscript. It is also a considerable challenge to develop a reliable statistical model for a non-targeted assay for the NHP industry because there are over a 1000 species that need to be in the model. Currently we have a basic RM DNA library for the top 250 species; only 25% of the work has been completed for a foundational DNA library for the NGS SOP pipeline and bioinformatic model. We are working on the challenge of assembling such a library, NGS pipeline and statistical model, with an estimated completion in 3–5 years. Targeted tests have been validated (Newmaster et al., 2018) are available now and can be deployed on sanger sequencing, qPCR/ddPCR, or targeted NGS once the critical issue inherent to this method have been resolved. Targets can include the botanical ingredients, and known adulterants including closely related species and contaminants. We have developed this approach using a short sequence botanical RM DNA library using sanger sequencing and nucleotide signature qPCR, which was used in our study. The results indicate there is considerably more success using this approach rather than the non-targeted NGS approach. It is important to note that we were able to get positive results using targeted PCR in the samples that failed to produce any measurable DNA; this was not possible with the other three biotechnologies. We did not use a targeted NGS approach as it is not fully developed and suffers from the same issues defined in the previous paragraph. The NHP sector can learn from the food ingredient, food borne pathogen and probiotic sectors in which they have utilized targeted tests mainly on qPCR platforms. These industries have avoided non-targeted tests and they have not embraced NGS technology as a commercial tool for quality control (Davis, 2014; Law et al., 2015; Zhao et al., 2014; Newmaster et al., 2019).

We encourage researchers and industry stakeholders to collaborate in the assemblage of botanical reference materials (RM) and validated molecular diagnostic tools. The NHP Research Alliance is providing coordination as it is focused on working with industry leaders and other researchers to build and curate a comprehensive, vouchered RM library for botanical ingredients in the food and NHP industry based on genome scans and NMR analytical chemistry models for over 1000 plant species with widespread population sampling. This will be used to develop more advanced molecular diagnostic tools (e.g., micro arrays, microfluidic devices, oxford nanopore etc.) focused on species-specific nucleotide signatures for botanical ingredients and adulterants. It will be developed in two phases of which phase one is a targeted approach that follows intensive validation methods (Newmaster et al., 2018) that can be deployed identity tools for qualifying supply chains and as a targeted screening tool for adulterants. Phase two is non-targeted in research and development that builds on the validated assays and RM libraries. We feel that these tools will serve the botanical ingredient industry in multiple sectors (NHPs, food, beverage, alcohol, spices etc.) with reinforced quality control systems that will protect companies against the risk of fraudulent product substitution, adulteration, contamination, and unlabelled fillers.

Declarations

Author contribution statement

Subramanyam Ragupathy, Adam Faller, Dhivya Shanmughanandhan, Prasad Kesanakurti, R. Uma Shaanker, Gudasalamani Ravikanth, Ramalingam Sathishkumar, Narayanasamy Mathivanan, Jingyuan Song, Jianping Han, Steven Newmaster: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This work was supported by Natural Health Products Research Alliance, University of Guelph, grant #300014600053312000.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Alvari, A., Mehranz, S.O.R., Ahmad, F., Abdin, M., 2012. Contemporary overview on clinical trials and future prospects of hepato-protective herbal medicines. Rev. Recent Clin. Trials 7 (3), 214–223.

Ananingsih, V.K., Sharma, A., Zhou, W., 2013. Green tea catechins during food processing and storage: a review on stability and detection. Food Res. Int. 50 (2), 469–479.

Bianchi, V., Ceol, A., Ogier, A.G.E., de Pretis, S., Galeota, E., Kishore, K., et al., 2016. Integrated systems for NGS data management and analysis: open issues and available solutions. Front. Genet. 7, 75.

Burns, M., Wiseman, G., Knight, A., Bramley, P., Foster, L., Rollinson, S., et al., 2016. Measurement issues associated with quantitative molecular biology analysis of complex food matrices for the detection of food fraud. Analyst 141 (1), 45–61.

CBOL Plant Working Group, Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., et al., 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. Unit. States Am. 106 (31), 12794–12797.

Chen, S.-L., Yu, H., Luo, H.-M., Wu, Q., Li, C.-F., Steinmetz, A., 2016. Conservation and sustainable use of medicinal plants: problems, progress, and prospects. Chin. Med. 11 (1), 37.

Chen, S., Pang, X., Song, J., Shi, L., Yao, H., Han, J., Leon, C., 2014. A renaissance in herbal medicine identification: from morphology to DNA. Biotechnol. Adv. 32 (7), 1237–1244.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., et al., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 5 (1), e8613.

Cheng, X., Su, X., Chen, X., Zhao, H., Bo, C., Xu, J., et al., 2015. Biological ingredient analysis of traditional Chinese medicine preparation based on high-throughput sequencing: the story for Liuwei Dihuang Wan. Sci. Rep. 4 (1), 5147.

Chikezie, P.C., Ojiako, O.A., 2015. Herbal medicine: yesterday, today and tomorrow. Altern.; Integr. Med. 04 (03), 195.

Coissac, E., Riaz, T., Puillandre, N., 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. Mol. Ecol. 21 (8), 1834–1847.

CRN, 2017. Consumer Survey on Dietary Supplements, Was Commissioned by the Council for Responsible Nutrition (CRN) and Conducted by Ipsos Public Affairs. Retrieved from. https://www.crnusa.org/resources/crn-2017-annual-survey-dietary-supplements. (Accessed 10 December 2017).

Davis, C., 2014. Enumeration of probiotic strains: review of culture-dependent and alternative techniques to quantify viable bacteria. J. Microbiol. Methods 103, 9–17.

De Boer, H.J., Ichim, M.C., Newmaster, S.G., 2015. DNA barcoding and pharmacovigilance of herbal medicines. Drug Saf. 38 (7), 611–620.

Ernst, E., 2005, August. The Efficacy of Herbal Medicine - an Overview. Fundamental and Clinical Pharmacology.

Faller, A.C., Ragupathy, S., Shanmughandhan, D., Zhang, Y., Lu, Z., Chang, P., et al., 2019. DNA quality and quantity analysis of camellia sinensis through processing from fresh leaves to a green tea extract. J. AOAC Int. 102.

Fazekas, A., Burgess, K., Kesanakurti, P.M., Percy, D., Hajibabaei, M., W Graham, S., et al., 2008. Assessing the utility of coding and non-coding genomic regions for plant DNA barcoding. PLoS One 3.

Fazekas, A.J., Kuzmina, M.L., Newmaster, S.G., Hollingsworth, P.M., 2012. DNA barcoding methods for land plants. In: Kress, W.J., Erickson, D.L. (Eds.), DNA Barcodes: Methods and Protocols. Humana Press, Totowa, NJ, pp. 223–252.

Fazekas, A., Steeves, R., Newmaster, S., 2010. Improving sequencing quality from PCR products containing long mononucleotide repeats. Biotechniques 48 (4), 277–285.

Gao, Z., Liu, Y., Wang, X., Song, J., Chen, S., Ragupathy, S., et al., 2017. Derivative technology of DNA barcoding (nucleotide signature and SNP double peak methods) detects adulterants and substitution in Chinese patent medicines. Sci. Rep. 7 (1), 5858.

Han, J., Pang, X., Liao, B., Yao, H., Song, J., Chen, S., 2016. An authenticity survey of herbal medicines from markets in China using DNA barcoding. Sci. Rep. 6 (1), 18723.

Harris, L.J., Griffiths, M.W., 1992. The detection of foodborne pathogens by the polymerase chain reaction (PCR). Food Res. Int. 25 (6), 457–469.

Izzo, A.A., Hoon-Kim, S., Radhakrishnan, R., Williamson, E.M., 2016. A critical approach to evaluating clinical efficacy, adverse events and drug interactions of herbal remedies. Phytother Res. 30 (5), 691–700.

Karni, M., Zidon, D., Polak, P., Zalevsky, Z., Shefi, O., 2013. Thermal degradation of DNA. DNA Cell Biol. 32 (6), 298–301.

Kress, W.J., Erickson, D.L., 2007. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One 2 (6), e508.

Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H., 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. Unit. States Am. 102 (23), 8369–8374.

Kruskal, J.B., 1964. Nonmetric multidimensional scaling: a numerical method. Psychometrika 29 (2), 115–129.

Law, J.W.-F., Ab Mutalib, N.-S., Chan, K.-G., Lee, L.-H., 2015. Rapid methods for the detection of foodborne bacterial pathogens: principles, applications, advantages and limitations. Front. Microbiol. 5, 770.

Lighten, J., van Oosterhout, C., Bentzen, P., 2014. Critical review of NGS analyses for de novo genotyping multigene families. Mol. Ecol. 23 (16), 3957–3972.

Lu, Z., Rubinsky, M., Babajanian, S., Zhang, Y., Chang, P., Swanson, G., 2018. Visualization of DNA in highly processed botanical materials. Food Chem. 245, 1042–1051.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., 2013. Cluster: Cluster Analysis Basics and Extensions.

Malentacchi, F., Pizzamiglio, S., Ibrahim-Gawel, H., Pazzagli, M., Verderio, P., Ciniselli, C.M., Wyrich, R., Gelmini, S., 2016. Second SPIDIA-DNA External Quality Assessment (EQA): influence of pre-analytical phase of blood samples on genomic DNA quality. Clin. Chim. Acta 454, 10–14.

Morgante, M., Olivieri, A.M., 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J. 3 (1), 175–182.

Newmaster, S.G., Fazekas, A.J., Ragupathy, S., 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Can. J. Bot. 84 (3), 335–341.

Newmaster, S.G., Fazekas, A.J., Steeves, R.A.D., Janovec, J., 2008. Testing candidate plant barcode regions in the Myristicaceae. Mol. Ecol. Resour. 8 (3), 480–490.

Newmaster, S.G., Grguric, M., Shanmughanandhan, D., Ramalingam, S., Ragupathy, S., 2013. DNA barcoding detects contamination and substitution in North American herbal products. BMC Med. 11 (1), 222.

Newmaster, S.G., Ragupathy, S., 2010. Ethnobotany genomics - discovery and innovation in a new era of exploratory research. J. Ethnobiol. Ethnomed. 6 (1), 2.

Newmaster, S.G., Shanmughanandhan, D., Kesanakurti, P., Shehata, H., Ragupathy, S., 2018. Guidelines for validation of qualitative real-time PCR methods for molecular diagnostic identification of botanicals. J. AOAC Int. Communicated.

Newmaster, S.G., Shanmughanandhan, D., Kesanakurti, P., Shehata, H., Faller, A., Della Noce, I., Lee, J.Y., Rudzinski, P., Lu, Z., Zhang, Y., Swanson, G., Hanner, R., Ragupathy, S., 2019. Recommendations for validation of real-time PCR methods for molecular diagnostic identification of botanicals. J. AOAC Int.

Orhan, I.E., Senol, F.S., Skalicka-Wozniak, K., Georgiev, M., Sener, B., 2016. Adulteration and safety issues in nutraceuticals and dietary supplements: innocent or risky?. In: Nutraceuticals. Elsevier, pp. 153–182.

Palhares, R.M., Gonçalves Drummond, M., dos Santos Alves Figueiredo Brasil, B., Pereira Cosenza, G., das Graças Lins Brandão, M., Oliveira, G., 2015. Medicinal plants recommended by the world health organization: DNA barcode identification associated with chemical analyses guarantees their quality. PLoS One 10 (5), e0127866.

Pehrson, J.R., 1989. Thymine dimer formation as a probe of the path of DNA in and between nucleosomes in intact chromatin. Proc. Natl. Acad. Sci. Unit. States Am. 86 (23), 9149–9153.

Percy, D.M., Argus, G.W., Cronk, Q.C., Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., et al., 2014. Understanding the spectacular failure of DNA barcoding in willows (Salix): does this result from a trans-specific selective sweep? Mol. Ecol. 23 (19), 4737–4756.

Pompanon, F., Deagle, B.E., Symondson, W.O.C., Brown, D.S., Jarman, S.N., Taberlet, P., 2012. Who is eating what: diet assessment using next generation sequencing. Mol. Ecol. 21 (8), 1931–1950.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.

Raclariu, A.C., Mocan, A., Popa, M.O., Vlase, L., Ichim, M.C., Crisan, G., et al., 2017. Veronica officinalis product authentication using DNA metabarcoding and HPLC-MS reveals widespread adulteration with veronica chamaedrys. Front. Pharmacol. 8, 378.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M., Gough, K.C., 2014. REVIEW: the detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. J. Appl. Ecol. 51 (5), 1450–1459.

Reinecke, F., Satya, R.V., DiCarlo, J., 2015. Quantitative analysis of differences in copy numbers using read depth obtained from PCR-enriched samples and controls. BMC Bioinf. 16 (1), 17.

Schrijver, I., Aziz, N., Farkas, D.H., Furtado, M., Gonzalez, A.F., Greiner, T.C., et al., 2012. Opportunities and challenges associated with clinical diagnostic genome sequencing. J. Mol. Diagn. 14 (6), 525–540.

Scora, R.W., 1966. Problems in chemotaxonomy: the influence of varying soil conditions, of geographical and individual variants upon the distribution of certain substances in chromatographed extracts ofmonarda fistulosa. Plant Soil 24 (1), 145–152.

Seethapathy, G.S., Ganesh, D., Santhosh Kumar, J.U., Senthilkumar, U., Newmaster, S.G., Ragupathy, S., et al., 2015. Assessing product adulteration in natural health products for laxative yielding plants, Cassia, Senna, and Chamaecrista, in Southern India using DNA barcoding. Int. J. Leg. Med. 129 (4), 693–700.

Shanmughanandhan, D., Ragupathy, S., Newmaster, S.G., Mohanasundaram, S., Sathishkumar, R., 2016. Estimating herbal product authentication and adulteration in India using a vouchered, DNA-based biological reference material library. Drug Saf. 39 (12), 1211–1227.

Song, H., Buhay, J.E., Whiting, M.F., Crandall, K.A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc. Natl. Acad. Sci. Unit. States Am. 105 (36), 13486–13491.

Sucher, N., Carles, M., 2008. Genome-based approaches to the authentication of medicinal plants. Planta Med. 74 (6), 603–623.

Udvardi, M.K., Czechowski, T., Scheible, W.-R., 2008. Eleven golden rules of quantitative RT-PCR. Plant Cell 20 (7), 1736–1737.

Urumarudappa, S.K.J., Gogna, N., Newmaster, S.G., Venkatarangaiah, K., Subramanyam, R., Saroja, S.G., et al., 2016. DNA barcoding and NMR spectroscopy-based assessment of species adulteration in the raw herbal trade of Saraca asoca (Roxb.) Willd, an important medicinal plant. Int. J. Leg. Med. 130 (6), 1457–1470.

Waits, L.P., Paetkau, D., 2005. Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. J. Wildl. Manag. 69 (4), 1419–1433.

WHO, 2001. Legal Status of Traditional Medicines and Complimentary/alternative Medicine: Worldwide Review. WHO Publications. Retrieved from. http://apps.who.int/medicinedocs/en/d/Jh2943e.

WHO, 2004. World Health Organization Guidelines on Safety Monitoring of Herbal Medicines in Pharmacovigilance Systems. Retrieved from. http://apps.who.int/medicinedocs/documents/s7148e/s7148e. (Accessed 20 October 2017).

Woolfe, M., Primrose, S., 2004. Food forensics: using DNA technology to combat misdescription and fraud. Trends Biotechnol. 22 (5), 222–226.

Zhao, X., Lin, C.-W., Wang, J., Oh, D.H., 2014. Advances in rapid detection methods for foodborne pathogens. J. Microbiol. Biotechnol. 24 (3), 297–312.

ZMR, 2017. Herbal Supplement Market by Source (Leaves, Barks, Fruits & Vegetables, Roots); by Application (Pharmaceuticals, Food & Deverages, Personalcare); by Function (Medicinal, Aroma): Global Industry Perspective, Comprehensive Analysis and Forecast, 2017 - 2022. Retrieved from. https://www.zionmarketresearch.com/inquiry/herbal-supplement-market. (Accessed 10 December 2017).