IFSCC INTERNATIONAL FEDERATION OF SOCIETIES OF COSMETIC CHEMISTS



# DNA Metabarcoding As a Tool to Trace Plants of Interest in Ingredients or Cosmetics

*N. Dubrulle, N. Giraud*
DNA Gensee, 17 rue du Lac St André, BP 90342, 73377 Le Bourget du Lac Cedex, France

**Keywords:** Metabarcoding, DNA analysis, vegetal oil, traceability, plant species

This publication was presented as a podium presentation at the 28th IFSCC Congress, October 27-30, 2014, Paris, France

## INTRODUCTION

The cosmetics and fragrances industries are increasing the amount of natural raw materials in their products. Thus, plant authenticity control is more and more important for industrial users and for customers. As a matter of fact, the high cost of such natural materials could lead to adulterations. Today, this control is mainly provided by physiochemical analyses [1]. Although these analyses make it possible to know the composition of certain chemical molecules, they cannot always give the botanical or the geographical origin of the plants used. Moreover, some plants do not possess precise enough chemical signatures to be detected, for example, plants that do not have aromatic molecules that are detectable by gas chromatography.

Every plant has a unique genetic signature, and the current molecular biology techniques could be very efficient in completing the physiochemical results [2]. Until now, we worked with substrates containing a large quantity of non-degraded DNA and needed to have an *a priori* idea of which plants we were looking for [3, 4]. For such analysis, both nuclear and chloroplastic or mitochondrial DNA are used.

In the case of nuclear DNA, mostly ITS (Internal Transcribed Spacers) DNA fragments are used. These DNA fragments or markers have been used to identify plant species [5, 6]. However, they are too long (several hundred base pairs) to be used in the case of degraded DNA like that in processed products.

Mitochondrial and chloroplastic DNA are interesting because they have a higher number of copies per cell. However, mitochondrial DNA varies too little to be able to discriminate and have good taxonomic resolution. It does not differentiate plants at the species level [7,8]. For this reason, chloroplastic DNA is preferred. Several markers of this organelle have been widely used, such as ndhF [9], rbcL [10] or trnK [11]. In 2005, Shaw and coworkers [12] established a list of 21 potential noncoding chloroplast regions. Kress and coworkers in 2005 [13] proposed use of the trnH-psbA intergenic region to obtain a resolution at the species level. However, the amplicon (DNA sequence amplified by PCR), with an average length of 465 base pairs, is still too long to allow access to degraded DNA. Overall, this subject is the center of several research projects [4, 14, 15].

The use of the trnL intron has found various applications in ecology [16-26].

Next-generation sequencing techniques (NGS, high-throughput or massive sequencing) have recently made possible to work on very short DNA sequences. The DNA barcoding then becomes metabarcoding thanks to the power of these analyses that can now take into account the environment of the analyzed sample. For instance, metabarcoding was tested for honey composition analysis [27].

In this article we want to summarize the technological benefit of this approach. Using vegetable oils as examples, we show how powerful these tools are in identifying and certifying the plant composition of processed products.

## EXPERIMENTAL

**Samples: Oils**

In collaboration with Ales Groupe, we gathered two types of oils to test the ef-

### Abstract

*Authenticity of natural raw materials has become increasingly important for customers, authorities and the cosmetics and fragrances industries. DNA metabarcoding is a powerful tool to retrieve plant species contained in a product using the fact that every plant has a unique genetic signature. DNA metabarcoding combines this approach with next-generation sequencing technology. Using this innovative method is revolutionary for botanical authenticity control because of its accuracy and its reliability. In this article we show that plant DNA can be retrieved in processed products such as oils. The DNA sequences are assigned to the plant species from which the oil is made.*

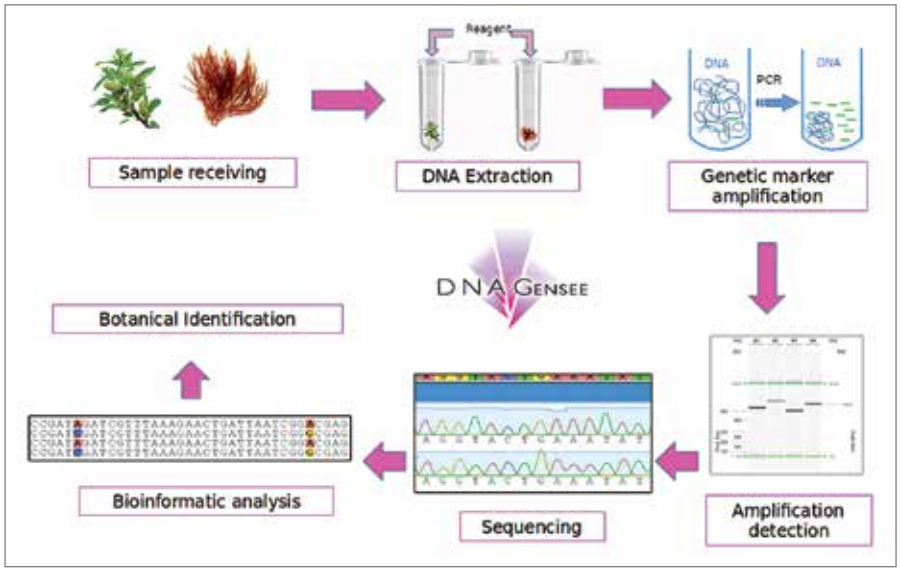


*Figure 1* Schematic representation of the major protocol steps to identify the plant species in products. The DNA barcoding and metabarcoding process illustrated is copyrighted by DNA Gensee.

ficiency of the metabarcoding technology to obtain their plant composition. These were poppy vegetable oil, which is a 100 % organic oil obtained from the first cold pressing of the seeds, and bitter orange essential oil, which is obtained by cold pressing the zests of bitter orange.

**DNA metabarcoding**

As shown in ***Figure 1***, several steps are necessary to access the DNA sequences coming from products. Extraction and PCR (polymerase chain reaction) amplification of DNA were carried out using universal primers and a well-defined protocol. The DNA concentration could be measured spectroscopically (SimpliNano 4285 V2.0.0, GE Health care) prior to sequencing.

DNA metabarcoding was then performed using a next-generation sequencing technique with a HiSeq Sequencing instrument (HiSeq, Illumina Inc.). Finally, a bioinformatics analysis was performed to identify species by comparing the sequences with the GenBank® database [28]. The results consisted of the DNA sequences found and their relative amount. These allowed calculation of the percentages of genetic sequences the plant species found and construction of abundance diagrams.

## RESULTS AND DISCUSSION

The results of next-generation sequencing (NGS) constitute the various genetic sequences found in the analyzed sample and their relative abundance.

These DNA sequences could come from plants from which the product was made and/or from plant tissues introduced during processes, transfers or samplings.

It is necessary to note that the quantity of DNA of the plant of interest is generally small because the industrial process degrades this DNA. The DNA of contaminant pollens introduced after the process will be not degraded and thus be present in great quantities.

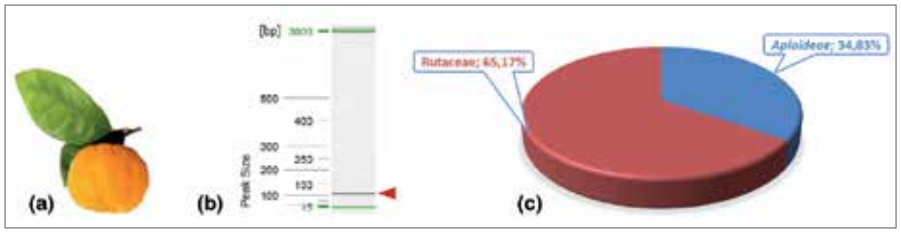


*Figure 2* Results of the next-generation sequencing (NGS) study on an essential oil. (a) Picture of the plant species *Citrus aurantium*; (b) electrophoresis gel of the DNA after the extraction and amplification steps; the arrow shows the amplified DNA ; (c) relative abundance diagram obtained after NGS analysis and comparison with the Genbank® database.

By comparing the genetic sequences with databases, we can assign a plant family, a genus or a species to a given sequence. The level of precision depends on the size of the databases.

This DNA analysis provides elements of traceability with respect to the plant DNA found. It reflects the life of raw materials and provides information on the impact of the process undergone by the plants of interest.

**Essential oil of bitter orange**

This essential oil is generally obtained by a long maceration process and filtration steps. This process disrupts the DNA content of the product. The sample of bitter orange essential oil used in this study was made by cold pressing of the zests of *Citrus aurantium* ***(Figure 2a)***. Extraction and amplification were successful ***(Figure 2b)***. As ***Figure 2c*** shows, 65.17 % of the genetic sequences were assigned to the Rutaceae family (the family of *Citrus aurantium*) and 34.83 % to the Apioideae subfamily. Although assignation gives only the plant family, traceability elements were also obtained showing that the plant of interest is recovered in the product. Using and analyzing a tissue of *Citrus aurantium* to obtain the genetic signature as a reference will refine the result of this study.

**Vegetable oil – poppy oil**

This vegetable oil was obtained from the first cold pressing of the seeds, resulting in a light oil in which the DNA is generally well preserved. The poppy oil was extracted from *Papaver somniferum* ***(Figure 3a)***. Extraction and amplification of DNA showed a good amount and good quality of DNA ***(Figure 3b)***, making it possible to perform a new-generation sequence (NGS) analysis. As can be seen in ***Figure 3c***, 98.46 % of the genetic sequences were assigned to *Papaver somniferum*. A small amount of non significant DNA sequences in the oil was

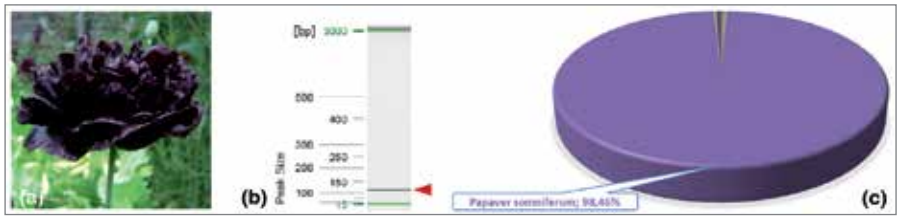


*Figure 3* Results of the next-generation sequencing (NGS) study on a vegetable oil. (a) Picture of the plant species *Papaver somniferum*; (b) electrophoresis gel of the DNA after the extraction and amplification steps, the arrow shows the amplified DNA; (c) relative abundance diagram obtained after NGS analysis and comparison with the Genbank® database.

assigned to other species and probably comes from contamination.

## CONCLUSION

Barcoding and metabarcoding technologies are evolving quickly these days, as can be seen from the numerous scientific articles on this subject [2, 6, 16, 27, 29]. In this study, we show the power of new-generation sequencing to retrieve plant DNA from processed products. Depending on the size of the databases and quality of the extracted DNA, the assignation brings relevant information concerning the plant species. Highly processed products are more difficult to work with but we were able to show that the product came from the good family or genus of the plant or plant species. If we have the genetic signature of the plant of interest, it is possible to compare it with the genetic sequences found in the product and then prove without doubt the traceability.

So far the plant databases are incomplete; many species have not been sequenced and referenced. Our aim is to enrich sequence databases and also develop new genetics tools for more precise assignation (up to species or even subspecies). Metabar-coding represents a considerable breakthrough in the analytics world. Moreover, it could be apply to very different domains, such as cosmetics, fragrances, nutrition, nutraceuticals and health, and for different types of complex substrates, including oils, creams, extracts and shampoos.

**Acknowledgements**

The authors would like to thank *Amélie Herlin-Bernal* and *Jean-Christophe Choulot* from Ales Groupe (89, rue Salvador Allende,CS 40028, 95871 Bezons Cedex) for their samples and help. Thanks also goes to DNA Gensee's technical team for all their hard work.

Furthermore, DNA Gensee is very grateful to have been given the opportunity to share these results with the IFSCC.

**References**

[1] *Betz J.M., Brown P.N., Roman M.C.,* Accuracy, precision, and reliability of chemical measurements in natural products research, Fitoterapia, **82** (2011) 44-52.

[2] *Coutinho Moraes D.F., Still D.W., Lum M.R., Hirsch A.M.,* DNA-Based Authentication of Botanicals and Plant-Derived Dietary Supplements: Where Have We Been and Where Are We Going?, Planta Med., **81** (2015) 687-695.

[3] *Mishra P., Kumar A., Nagireddy A., Mani D.N., Shukla A.K., Tiwari R. et al.,* DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market, Plant Biotechnol J., **14** (2016) 8-21.

[4] *Ajmal Ali M., Gyulai G., Hidvégi N., Kerti B., Al Hemaid F.M.A., Pandey A.K. et al.,* The changing epitome of species identification – DNA barcoding, Saudi J. Biol. Sci., **21** (2014) 204-231.

[5] *Zhang J.Q., Meng S.Y., Wen J., Rao G.Y.,* DNA barcoding of Rhodiola (crassulaceae): a case study on a group of recently diversified medicinal plants from the Qinghai-Tibetan Plateau, PloS One, **10** (2015) e0119921.

[6] *Cheng T., Xu C., Lei L., Li C., Zhang Y., Zhou S.,* Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity, Mol. Ecol. Resour., **16** (2016) 138-149.

[7] *Palmer J.D., Herbon L.A.,* Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence, J. Mol. Evol. **28** (1988) 87-97.

[8] *Gualberto J.M., Mileshina D., Wallet C., Niazi A.K., Weber-Lotfi F., Dietrich A.,* The plant mitochondrial genome: dynamics and maintenance, Biochimie, **100** (2014) 107-120.

[9] *Zhang W., Fan X., Zhu S., Zhao H., Fu L.,* Species-specific identification from incomplete sampling: applying DNA barcodes to monitoring invasive solanum plants, PloS One, **8** (2013) e55927.

[10] *MacIvor J.S.,* DNA barcoding to identify leaf preference of leafcutting bees, R. Soc. Open Sci., **3** (2016) 150623.

[11] *Clement W.L., Donoghue M.J.,* Barcoding success as a function of phylogenetic relatedness in Viburnum, a clade of woody angiosperms, BMC Evol. Biol., **12** (2012) 73.

[12] *Shaw J., Lickey E.B., Beck J.T., Farmer S.B., Liu W., Miller J., et al.,* The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis, Am. J. Bot., **92** (2005) 142-166.

[13] *Kress W.J., Wurdack K.J., Zimmer E.A., Weigt L.A., Janzen D.H.,* Use of DNA barcodes to identify flowering plants, Proc. Natl. Acad. Sci. U. S. A., **102** (2005) 8369-8374.

[14] *Dong W., Liu J., Yu J., Wang L., Zhou S.,* Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding, PloS One, **7** (2012) e35071.

[15] *Wynns J.T., Lange C.B.A.,* A comparison of 16 DNA regions for use as phylogenetic markers in the pleurocarpous moss genus Plagiothecium (Hypnales), Am. J. Bot., **101** (2014) 652-669.

[16] *Kress W.J., García-Robledo C., Uriarte M., Erickson D.L.,* DNA barcodes for ecology, evolution, and conservation, Trends Ecol. Evol, **30** (2015) 25-35.

[17] *Valentini A., Miquel C., Nawaz M.A., Bellemain E., Coissac E., Pompanon F. et al.,* New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach, Mol. Ecol. Resour., **9** (2009) 51-60.

[18] *Soininen E.M., Valentini A., Coissac E., Miquel C., Gielly L., Brochmann C. et al.,* Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures, Front. Zool., **6** (2009) 16.

[19] *Quéméré E., Hibert F., Miquel C., Lhuillier E., Rasolondraibe E., Champeau J., et al.* A DNA metabarcoding study of a primate dietary diversity and plasticity across its entire fragmented range, PloS One, **8** (2013) e58971.

[20] Czernik M., Taberlet P., Swisłocka M., Czajkowska M., Duda N., Ratkiewicz M., Fast and efficient DNA-based method for winter diet analysis from stools of three cervids: moose, red deer, and roe deer, Acta Theriol. (Warsz.)., **58** (2013) 379-386.

[21] *Hibert F., Taberlet P., Chave J., Scotti-Saintagne C., Sabatier D., Richard-Hansen C.,* Unveiling the diet of elusive rainforest herbivores in next generation sequencing era? The tapir as a case study, PloS One, **8** (2013) e60799.

[22] *Staudacher K., Wallinger C., Schallhart N., Traugott M.,* Detecting ingested plant DNA in soil-living insect larvae, Soil Biol. Biochem., **43** (2011) 346-350.

[23] *Liepelt, S., Sperisen, C., Deguilloux, M.F., Petit, R.J., Kissling, R., Spencer, M., et al.,* Authenticated DNA from ancient wood remains, Ann. Bot., **98** (2006) 1107-1111.

[24] *Sønstebø J.H., Gielly L., Brysting A.K., Elven R., Edwards M., Haile J., et al.,* Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate, Mol. Ecol. Resour., **10** (2010) 1009-1018.

[25] *Gould B.A., León B., Buffen A.M., Thompson L.G.,* Evidence of a high-Andean, mid-Holocene plant community: An ancient DNA analysis of glacially preserved remains, Am. J. Bot., **97** (2010) 1579-1584.

[26] *Pedersen J.S., Valen E., Velazquez A.M.V., Parker B.J., Rasmussen M., Lindgreen S. et al.,* Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome, Genome Res., **24** (2014) 454-466.

[27] *Hawkins J., de Vere N., Griffith A., Ford C.R., Allainguillaume J., Hegarty M.J. et al.,* Using DNA Metabarcoding to Identify the Floral Composition of Honey: A New Tool for Investigating Honey Bee Foraging Preferences, PloS One, **10** (2015) e0134735.

[28] *Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Sayers E.W.,* GenBank, Nucleic Acids Res., **37** (2009) D26-31.

[29] *Kress W.J., Erickson D.L., Jones F.A., Swenson N.G., Perez R., Sanjur O. et al.,* Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama, Proc. Natl. Acad. Sci. U. S. A., **106** (2009) 18621-18626.

Corresponding Author
*N. Dubrulle*
DNA Gensee
Le Bourget du Lac
France
nelly.dubrulle@dnagensee.com