

## RESEARCH ARTICLE

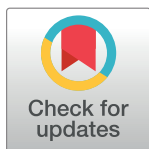
# What is in your cup of tea? *DNA Verity Test* to characterize black and green commercial teas

Olga De Castro<sup>1\*</sup>, Maria Comparone<sup>1</sup>, Antonietta Di Maio<sup>1</sup>, Emanuele Del Guacchio<sup>2</sup>, Bruno Menale<sup>1</sup>, Jacopo Troisi<sup>3</sup>, Francesco Aliberti<sup>4</sup>, Marco Trifuoggi<sup>5</sup>, Marco Guida<sup>4</sup>

**1** Department of Biology, Botanical Garden, University of Naples Federico II, Naples Italy, **2** Botanical Garden, University of Naples Federico II, Naples Italy, **3** Department of Medicine and Surgery and Dentistry, University of Salerno, Salerno, Italy, **4** Department of Biology, University of Naples Federico II, Naples Italy, **5** Department of Chemical Sciences, University of Naples Federico II, Naples, Italy

 These authors contributed equally to this work.

\* [odecastr@unina.it](mailto:odecastr@unina.it)



## Abstract

In this study, we used several molecular techniques to develop a fast and reliable protocol (*DNA Verity Test*, *DVT*) for the characterization and confirmation of the species or taxa present in herbal infusions. As a model plant for this protocol, *Camellia sinensis*, a traditional tea plant, was selected due to the following reasons: its historical popularity as a (healthy) beverage, its high selling value, the importation of barely recognizable raw product (i.e., crushed), and the scarcity of studies concerning adulterants or contamination. The *DNA Verity Test* includes both the sequencing of DNA barcoding markers and genotyping of labeled-PCR DNA barcoding fragments for each sample analyzed. This protocol (*DVT*) was successively applied to verify the authenticity of 32 commercial teas (simple or admixture), and the main results can be summarized as follows: (1) the *DVT* protocol is suitable to detect adulteration in tea matrices (contaminations or absence of certified ingredients), and the method can be exported for the study of other similar systems; (2) based on the BLAST analysis of the sequences of *rbcl+matK±rps7-trnV<sup>(GAC)</sup>* chloroplast markers, *C. sinensis* can be taxonomically characterized; (3) *rps7-trnV<sup>(GAC)</sup>* can be employed to discriminate *C. sinensis* from *C. pubicosta*; (4) ITS2 is not an ideal DNA barcode for tea samples, reflecting potential incomplete lineage sorting and hybridization/introgression phenomena in *C. sinensis* taxa; (5) the genotyping approach is an easy, inexpensive and rapid pre-screening method to detect anomalies in the tea templates using the *trnH<sup>(GUG)</sup>-psbA* barcoding marker; (6) two herbal companies provided no authentic products with a contaminant or without some of the listed ingredients; and (7) the leaf matrices present in some teabags could be constituted using an admixture of different *C. sinensis* haplotypes and/or allied species (*C. pubicosta*).

## OPEN ACCESS

**Citation:** De Castro O, Comparone M, Di Maio A, Del Guacchio E, Menale B, Troisi J, et al. (2017) What is in your cup of tea? *DNA Verity Test* to characterize black and green commercial teas. PLoS ONE 12(5): e0178262. <https://doi.org/10.1371/journal.pone.0178262>

**Editor:** Massimo Labra, Università degli Studi di Milano-Bicocca, ITALY

**Received:** February 13, 2017

**Accepted:** May 4, 2017

**Published:** May 23, 2017

**Copyright:** © 2017 De Castro et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper.

**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

An ancient English proverb states “*The path to heaven passes through a teapot*”, and this proverb is sadly true in recent years. In fact, considering the latest dramatic news concerning food

fraud and soil contamination (e.g., soil contamination in Indian tea estates soil; [1, 2]), it is likely that “*the path to heaven*” can be accelerated by what you drink or eat (e.g., the tragic case of the “*soil of fire*” in southern Italy; [3, 4]). Currently, the adulteration of food is frequent, and the scientific community has used been advanced technologies to protect humans from incorrect practices in food or drink production (e.g., [5, 6, 7, 8, 9, 10]). Indeed, some products can represent real and dangerous “mirror for larks” in terms of health, as the presumed natural or biological aliments or supplements purchased by consumers to protect their health lack valid quality controls [11]. Moreover, these controls are occasionally difficult to perform, reflecting the intrinsic nature of the product (e.g., food supplements or herbal medicine [12, 13, 14]). The international trade of herbal products is in continuous development for both alimentary and pharmaceutical purposes [15]; in fact, many plants are used daily in the preparation of foods and herbal teas (e.g., [16, 17]). Specifically, tea infusions are widely used as both pleasant drinks and for their many beneficial properties [18], and an accurate definition of the compounds present in these teas is important not only for the consumer but also for the producers and control authorities [19]. Because of the fragmentation or pulverizing of the vegetal material, it is often difficult to identify the species among the ingredients using traditional analyses (i.e., macroscopic and microscopic morphology) [20, 21]. Thus, a DNA barcoding technique has recently been used to address this problem, resulting in a useful system for the detection of the plants employed in these teas and the characterization of the possible adulteration and/or contamination in a wide range of plant-based foods (e.g., [22, 23]), becoming a universal adopted approach in the last few years [24, 25].

In this study, we used a multi-faceted DNA barcoding approach to develop a fast and reliable protocol (*DNA Verity Test*, *DVT*) for the taxonomic characterization and confirmation of herbal infusions. As a model plant for this protocol, *Camellia sinensis* (L.) Kuntze, the traditional tea plant, was selected for (1) its historical popularity as a (healthy) beverage [26, 27, 28], (2) its high selling value (cfr. [27]), (3) the importation of barely recognizable raw product (i.e., crushed) [29], and (4) the scarcity of studies concerning adulterants or contaminations [30, 31].

The *DNA Verity Test* was successively applied to verify the authenticity of 32 commercial tea packages (simple or admixture). Briefly, the *DNA Verity Test* verified both the sequencing of DNA barcoding markers and the genotyping of labeled-PCR DNA barcoding fragments for each sample analyzed.

The methodology of *DNA Verity Test* presented here is very detailed and is a promising tool for checking the authenticity of tea samples and it could be also suitable for application in different study systems.

In addition, the development of *DNA Verity Test* is part of a wider project comprising a multi-faceted pilot study with the purpose of analyzing 32 brands of European and Italian commercially available teas (16 black and 16 green teas) using different analytical approaches evaluate the presence of mycotoxins, microbial contaminants, heavy metals and phthalates (O. De Castro, unpublished data).

## Materials and methods

### Tea sampling design

Currently, tea products are available at a variety of mainstream outlets, such as supermarkets, health food stores, drug stores and online from herbal supply companies. In the present study, a total of 32 tea packages (*C. sinensis*) were purchased from markets in Naples (southern Italy) and from online-shops (Table 1) and subsequently tested using a blind experiment for the analyses reported below. All products were also available to consumers through online-shops, representing 17 Italian or internationally famous brands (seven and ten, respectively). Within

**Table 1. List of Italian commercialized black and green tea packages analyzed in the present study (N and V samples, respectively).** Information for each accession about the marketing quality (high, medium and low), sales network (D = discount supermarket; H = herbalist shop; S = supermarket; P = drugstore), price (€) {(A), < 1 €; (B), 1 < € < 2; (C), 2 < € < 4; (D), 4 < € < 6; (E), > 6 €} and molecular results for *rbcl* and *rps7-trnV<sup>(GAC)</sup>* sequences (presence of a SNP in the 68 bp coding region of *rbcl*, A = adenine, C = cytosine; *rps7-trnV<sup>(GAC)</sup>*, 239 bp = *Camellia sinensis*, 226 bp = *C. pubicosta*; in smaller font, the nucleotide/fragment less represented).

Code	Marketing quality*	Sales network	Price	Labelling information	<i>rbcl</i> haplotype (68 bp SNP)	<i>rps7-trnV<sup>(GAC)</sup></i> haplotype (bp)
<b>Black teas</b>						
N1	Medium	H	D	Organic black tea leaves of <i>Camellia sinensis</i> (L.) Kuntze (100%)	A/c	239
N2	Medium-good	P, H	D	Tea leaves [ <i>Camellia sinensis</i> (L.) Kuntze]	C	239
N3	Good	S	B	Black tea, aromas	C	239
N4	Low-medium	S	C	Not reported	C	239
N5	Good	P, H	C	Black certified tea (100%)	C/a	239
N6	Good	S	B	Tea	C	239/226
N7	Good	S	C	Decaffeinated tea	C	239
N8	Good	S	B	Not reported	C	239/226
N9	Good	S	C	Black tea (94%), lemon aroma (6%)	C/a	239/226
N10	Low	S	B	Black tea (100%)	C	239/226
N11 <sup>†</sup>	Good	S	D	Decaffeinated black tea, caffeine < 0.1%	C/a	239/226
N12	Low	S	B	Black tea	C	239/226
N13	Low	S	B	Black tea (95%), natural lemon aroma (5%)	C/a	239/226
N14	Low	S	B	Black tea leaves	C	239/226
N15	Medium	H	C	Decaffeinated organic black tea, caffeine < 0.1%	C	239
N16	Good	S	C	Sugar, acidifier (citric acid), decaffeinated tea extract, aromas, lemon juice powdered (0.5%); gluten free	N.d.	N.d.
<b>Green teas</b>						
V1	Good	P, H	E	Pure leaves of green tea ( <i>Camellia sinensis</i> Kuntze)	A	239
V2	Medium	P, H	E	Green tea leaves ( <i>Camellia sinensis</i> ) (99%), bergamot essential oil (1%)	A	239
V3	Medium	H	D	Organic green tea leaves [ <i>Camellia sinensis</i> (L.) Kuntze] (100%)	A	239
V4	Medium	P, H	D	Green tea leave	A	239
V5	Medium-good	P, H	D	Unfermented organic tea leaves ( <i>Camellia sinensis</i> L.)	A/c	239
V6	Medium	H	C	Green tea certified by Fairtrade	C	239/226
V7	Medium	H	D	Decaffeinated; ingredient not listed	A	239
V8	Good	S	C	Green tea (100%)	C	239/226
V9	Good	S	C	Green tea, aromas, and citrus peels (2.1%: grapefruit, lemon, lime, and orange)	C	239/226
V10 <sup>†</sup>	Good	S	C	Green tea, aromas, and spices (2%) (anise, cinnamon, and licorice)	C	239/226
V11	Good	S	C	Green tea, aromas, and mint (7.9%)	C	239/226
V12	Good	S	C	Green tea	A/c	239
V13	Low	S	B	Green tea	A	239
V14	Low	S	B	Organic green tea	A/C	239
V15	Low	D	A	Green tea	A/C	239
V16	Medium	S	A	Sugar, acidifier (citric acid), green tea soluble extract, aromas, and ginseng soluble extract	N.d.	N.d.

\* , value determined by a questionnaire conducted on 25 people (13 females and 12 males)

† , adulterated product; N.d., no datum.

<https://doi.org/10.1371/journal.pone.0178262.t001>

these 17 brands, 16 samples were fermented tea (only black tea) and 16 samples were raw tea (only green tea). Decaffeinated, soluble, admixture (i.e., the presence of other plant material) and flavored samples were also considered (Table 1). The tea samples were selected considering the trade network (supermarket, drug store and herbalist shop), the price (cheap and expensive), the marketing quality (packaging, publicity and brand) and the presence of filters in the packages (except for one type of soluble green tea). This information (except the brand name) is reported in Table 1. The tea samples were stored at room temperature prior to the molecular analyses, and duplicates for each tea package-lot were analyzed. As a preliminary molecular analysis, specimens of *C. sinensis* (CS-DNAs) were obtained from the Botanical Garden of Naples (BGN). The tissues of other species indicated in the admixture tea samples were obtained from BGN as reference samples [i.e., *Mentha* L. sp. (peppermint and corn mint), *Citrus* L. sp. (lemon, orange, grapefruit, and lime), *Pimpinella anisum* L. (anise), *Cinnamomum verum* J. Presl or *C. cassia* (L.) J. Presl (cinnamon), and *Glycyrrhiza glabra* L. (licorice)].

## DNA Verity Test (DVT) procedure

**Barcode markers.** Candidate DNA barcoding markers were selected based on previous tests, which had been conducted using universal primers [25, 31, 32, 33, 34], GenBank data and reference *C. sinensis* DNA from BGN. The following four criteria for selecting an ideal nucleotide sequences barcode were considered: (1) highly efficient amplification, (2) high quality sequences (e.g. no unambiguous sequences caused by double peaks or stuttering effect), (3) an exhaustive sequence database publicly available, and (4) high species discrimination capability.

The following two molecular techniques were used for the *DNA Verity Test (DVT)*: sequencing via Sanger chemistry and genotyping of fluoresced amplified fragment via capillary electrophoresis. After in silico and laboratory tests, the candidate markers for the sequencing approach included the following: genes and plastid intergenic spacers {*rbcL*, *matK* and *trnH*<sup>(GUG)</sup>-*psbA*, *rps7-trnV*<sup>(GAC)</sup>, respectively} and a nuclear intergenic spacer (ITS2). For the genotyping analyses, the considered markers were: an intron {P6 loop of the *trnL*<sup>(UAA)</sup>}, plastid intergenic spacers {*trnH*<sup>(GUG)</sup>-*psbA* and *rps7-trnV*<sup>(GAC)</sup>}, and a nuclear intergenic spacer (ITS2). The universal barcode primers and the specific barcoding primers designed in the present study are reported in Table 2.

**Tea DNA extraction.** A total of 30 mg of shredded material present in tea bags from each package was used for DNA extraction. Aiming at selecting the best DNA isolation procedure to obtain high yield and quality of extracted DNA [42], a preliminary analysis on three samples (green, black and admixture tea) was performed using four commercial kits {PowerPlant Pro DNA Isolation Kit (Mo Bio), Plant Genomic DNA Extract Mini Kit (Fisher Molecular Biology), ZR Plant/Seed DNA MicroPrep (Zymo Research), GeneAll Exgene Plant SV kit (GeneAll Biotechnology)} and two detergent protocols [43, 44]. The isolated DNA was analyzed using both a spectrophotometer Nanodrop 2000 (Thermo Fisher Scientific) to quantify its purity grade (260/280 and 260/230) and a Qubit 3 Fluorometer to determine the precise DNA concentration (Life Technologies, Thermo Fisher Scientific). In addition, a visual estimate was obtained using 1% agarose electrophoresis with Gel Red strained (Biotium) band intensities and GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific). After preliminary analysis, all extracted genomic DNA were estimated using both fluorometric and electrophoresis analyses.

For fresh plant references from BGN, 100 mg of leaf tissue was used for the DNA extraction using GeneAll Exgene Plant SV kit (GeneAll Biotechnology) according to the manufacturer's instructions.

**PCR amplification.** Molecular markers were amplified using a high-fidelity DNA polymerase, and the primers are listed in Table 2. The PCRs were performed using 10 ng of

Table 2. Primers used in the present study for DNA barcoding analyses.

Locus	Primer name	Sequence (5'-3')	Ta	Primer note	Reference
<b>rbcl</b>	rbclLa_F	ATG TCA CCA CAA ACA GAG ACT AAA GC	55	Universal	[35]
	rbclLajf634R	GAA ACG GTC TCT CCA ACG CAT			[36]
	CS_rbcl-300rev	TAA AGG ATA CGC TAC ATA AGC	55	Specific for <i>Camellia sinensis</i> ; used to check the SNP-M (68 bp) in sequencing PCR	In the present study
<b>matK</b>	1R_KIM-f	ACC CAG TCC ATC TGG AAA TCT TGG TTC	55	Universal	Ki-Joong Kim, pers. comm.
	3F_KIM-r	CGT ACA GTA CTT TTG TGT TTA CGA G			
<b>trnH<sup>(GUG)</sup>, psbA</b>	psbA3'f	GTT ATG CAT GAA CGT AAT GCT C	55	Universal	[37]
	trnHf*	GTT ATG CAT GAA CGT AAT GCT C			[38]
	Cinnamon_2psb-for	ACG AGT CGT TGA AGG ATC AAT	56	Specific for <i>Cinnamomum verum</i> and <i>C. cassia</i> (cinnamon)	In this study
	Cinnamon_2trnH-rev	TGC AGG TTG GTA CAG AAG AA			
	Citrus_psbA-in	GTA TTG ATC CGT TAT TTA GTC G	57	Specific for <i>Citrus</i> sp. (lemon, orange, grapefruit, and lime)	In this study
	Citrus_trnH-in	ATC TTA TCT TAC TTA TGA AGA ACC			
	Glycyrrhiza_psbA-in	GTT TTA AAG AAG GAT ACG AGG	55	Specific for <i>Glycyrrhiza glabra</i> (licorice)	In this study
	Glycyrrhiza_trnH-inA	TAC ATT CGC CCT TCT TAT AC			
	Mentha_2psb_for	TTC CAG GCA AGT CCA ATA CT	56	Specific for <i>Mentha</i> sp. ( <i>M. aquatica</i> , <i>M. arvensis</i> , <i>M. canadensis</i> , <i>M. longifolia</i> , <i>P. pulegium</i> , <i>M. suaveolens</i> , <i>M. pulegium</i> , and <i>M. x piperita</i> , <i>M. spicata</i> )	In this study
	Mentha_2trnH_rev	TGT GTA GGA GTT TTT GAA AAT AGA C			
	Pimpinella_psbA-in	ACC TAG CTG CTG TTG AAG CTC	60	Specific for <i>Pimpinella anisum</i> (anise)	In this study
	Pimpinella_trnH-in	GGA GCA ATA TCG CTT TCT TGA TAG A			
<b>rps7- trnV<sup>(GAC)</sup></b>	CS_rps7_1	GTG CTA TGG CTC GAA TCC GT	67	Specific for <i>Camellia</i> sp.; used to discriminate <i>C. sinensis</i> from <i>C. pubicosta</i>	In this study
	CS_trnV <sup>(GAC)</sup> _1*	ACC ACG TCA AGG TGA CAC TC			
<b>P6 loop trnL<sup>(UAA)</sup></b>	g	GGG CAA TCC TGA GCC AA	55	Universal	[39]
	h*	CCA TTG AGT CTC TGC ACC TAT C			
<b>ITS2</b>	S2F	ATG CGA TAC TTG GTG TGA AT	55	Universal	[40]
	S3R*	GAC GCT TCT CCA GAC TAC AAT			
	BELL-1	GGD GCG GAK AHT GGC CYC CCG TGC	55	Universal	[41]
	BELL-3	GAC GCT TCT CCA GAC TAC AAT			

\*, labeled

<https://doi.org/10.1371/journal.pone.0178262.t002>

genomic DNA, 5 X Phusion HF buffer and 0.5 uM of each primer with Phusion Hot Start II DNA Polymerase (Thermo Fisher Scientific) according to the manufacturer's instructions. The amplification of the recalcitrant DNA templates was performed using AmpONE Fast Pfu DNA polymerase (GeneAll Biotechnology).

**Sequencing approach.** According to the sequencing approach, the amplified fragments longer than 350 bp (e.g., *rbcL/matK*) were purified using PEG8000 precipitation (PEG 20%, 2.5 M NaCl). In contrast, polymorphic PCR fragments shorter than 350 bp [i.e., *trnH<sup>(GUG)</sup>-psbA/rps7-trnV<sup>(GAC)</sup>/ITS2*] were purified using Monarch PCR and the DNA Cleanup kit (New England BioLabs). Approximately 10 ng/100 bp of the purified templates were sequenced according to Di Maio and De Castro [45] using a fluorescent dye (Bright Dye Terminator Cycle Sequencing Kit, ICloning). The reactions were purified using BigDye X Terminator Purification Kit (Applied Biosystems, Thermo Fisher Scientific) and read using an automated sequencer (3130 Genetic Analyzer, Life Technologies, Thermo Fisher Scientific). The sequences were analyzed using AB DNA Sequencing Analysis version 5.2 software (Applied Biosystems, Thermo Fisher Scientific Inc.), edited in Chromas lite ver. 2.1.1. software ([http://technelysium.com.au/?page\\_id=13](http://technelysium.com.au/?page_id=13)), and assembled and aligned in BioEdit ver. 7.2.5 software [46]. PCR fragments with multiple peaks within the sequence were cloned using the CloneJET PCR Cloning Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. Transformation was performed using StrataClone SoloPack Competent Cells (Agilent Technologies). The bacteria were cultured in LB medium at 37°C for 30 min and subsequently transferred to LB agar plates containing 100 ug/ml ampicillin. Ninety-six randomly selected clones from each transformation were amplified using the corresponding DNA barcoding primers. For each PCR fragment polymorphic from the *C. sinensis* reference (e.g., CS<sub>trnH-psbA</sub> ref = 510 bp), five PCR samples were sequenced.

The identification of unknown sequence barcodes from the tea samples was conducted using the Basic Local Alignment Search Tool (BLAST; [47]) implemented in two barcode libraries {GenBank and The Barcode Of Life Data system (BOLD); [48, 49]}. To optimize correct identifications, the closest match for each molecular marker was defined as the target with the highest percentage identity using an arbitrary cut-off of 95% and an E-value < 1e<sup>-4</sup> or greater in terms of overlap with the query sequence.

**Genotyping approach.** Regarding the genotyping approach, the amplification procedure for the P6 loop, *trnH<sup>(GUG)</sup>-psbA*, *rps7-trnV<sup>(GAC)</sup>* and ITS2 markers was the same as that reported above, except for the use of 0.5 uM of fluorescently labeled reverse primer (Table 2). Five microliters {ca. 20 ng (P6 loop), 40 ng (*rps7-trnV<sup>(GAC)</sup>*) and 100 ng (*trnH<sup>(GUG)</sup>-psbA* and ITS2)} of amplified labeled fragments were purified using 2 uL of CleanSweep PCR Purification Reagent (Applied Biosystems, Thermo Fisher Scientific); 0.5 uL {ca. 3 and 7.5 ng, respectively; Relative Fluorescence Units (RFU) > 6000} and 1 uL of 1:5 dilution (ca. 1 and 3 ng, respectively; RFU ≥ 3000) for each purification were loaded onto a 3130 Genetic Analyzer with 0.4 uL of fluorescently labeled internal size standard (GeneScan 1000 ROX or 600 LIZ dye Size Standards, Applied Biosystems, Life Technologies). *Camellia sinensis* PCR-labeled standards were also generated for each marker used (CS-ref) and for the other plant species present in the admixture tea infusions (i.e., mint, lemon, orange, grapefruit, lime, anise, cinnamon and licorice, Table 1). Raw data were scored with an internal size standard using Peak Scanner version 1.0 software (Applied Biosystems, Life Technologies).

## Results and discussion

### DNA Verity Test (DVT) protocol

**DNA extraction.** First, we obtained a standard and rapid method to achieve good quality genomic DNA free from inhibitors (e.g., polysaccharides, polyphenol and phenolic compounds)



that could interfere with the activity of DNA polymerase in the PCR amplification, as previously reported [50]. Indeed, according to Graham [51], some compounds are present at high concentrations in the leaves of *C. sinensis* (e.g., polysaccharides > 12%, secondary metabolites = 40% on dry leaf weight) and can be co-precipitated with DNA determining failure in PCR amplification reactions. The genomic DNA from the tea samples was isolated using the Exgene Plant SV Kit, which produced a better quantity ( $70.3 \text{ ng/uL} \pm 4.3 \text{ SE}$ ) and quality of DNA amplification ( $A_{260}/A_{280} = 1.82 \pm 0.005 \text{ SE}$ ;  $A_{260}/A_{230} = 1.47 \pm 0.02 \text{ SE}$ ) compared with other kits. The DNA was eluted in 100  $\mu\text{L}$  of nuclease-free molecular biology grade water (Ambion, Thermo Fisher Scientific).

**PCR amplification.** In *C. sinensis* reference DNAs (CS-DNAs), *rbcL* oligos produce a fragment of 654 bp; *matK* oligos produce a fragment of 889 bp; *trnH*<sup>(GUG)</sup>-*psbA* oligos produce a fragment of 510 bp; *rps7-trnV*<sup>(GAC)</sup> produce a fragment of 239 bp; P6 loop oligos produce a fragment of 90 bp and ITS2 oligos produce fragments of 475 bp (ITS2-Chen oligos) and 345 bp (ITS2-Chiou oligos). High efficiency DNA amplification was obtained for all the markers, except for the *matK* gene region (i.e., low-medium), as also reported in the literature (e.g., [31, 33, 34]). Polymerase stuttering effects were eliminated using high-fidelity DNA Polymerase (Phusion Hot Start II), particularly in *trnH*<sup>(GUG)</sup>-*psbA*, where a fixed *polA*<sub>13</sub> was present at the end of sequence {5'-*trnH*<sup>(GUG)</sup>-*psbA*-3'} [52].

**PCR sequencing.** The PCR products from CS-DNAs corresponding to the above DNA markers were sequenced and high-quality sequences were obtained for all markers (i.e., no double peaks caused by more sequences), except for the nuclear markers ITS2-Chen and ITS2-Chiou [40, 41]. Comparing the raw data of electropherograms obtained using the different ITS2 oligos (Chen vs. Chiou), the ITS2-Chious sequences present less noise, but double peaks were consistently present. It is highly likely that the double peaks reflect the different design of the forward primer (BELL1) into a conserved part of ITS2, which amplified a partial ITS2, determining less noise compared with the total ITS2 of Chen *et al.* [40], where the forward oligo is designed before the ITS1 (i.e., 5.8S rRNA). The presence of different ITS2 sequences was also demonstrated through the cloning of the PCR fragment from the reference DNA of two different accessions of *C. sinensis*; ten positive colonies for each sample were sequenced and blasted using the GenBank database, confirming the absence of DNA as a contaminant (GenBank accessions: from KY928288 to KY928307). The presence of a multi-family of ITS sequences in the CS-DNAs was also observed in an additional assessment of the tea samples, which did not show contamination phenomena or adulteration detected with the other molecular markers employed. These data (i.e., intra-variability of ITS) were consistent with the literature [42, 53, 54], where a probable incomplete lineage sorting together to a high anthropic manipulation of the species (e.g., domestication and cultivation) would amplify the genetic variability of this nuclear molecular marker. Thus, ITS2 has been discarded from the analyses of the tea samples in the present study, reflecting potential double peaks and noise in the raw data (i.e., ambiguous sequences) and thus reflecting its intrinsic nature and not the presence of actual contaminations or adulterants. Indeed, for the simple characterization of the species, the consensus sequence could be used as reported in Lee *et al.* [55], even if careful manual editing must be performed to obtain the sequences.

Another molecular marker discarded as an ideal sequencing barcode was *trnH*<sup>(GUG)</sup>-*psbA*. Even if this marker is easily amplified, its sequence was not discriminant among congeneric *Camellia* species, in contrast to *rbcL*+*matK*. Indeed, BLAST *trnH*<sup>(GUG)</sup>-*psbA* sequence comparison (excluding polyA<sub>n</sub>) showed 100% identity in addition to *C. sinensis*, with more than ten *Camellia* taxa [e.g., *C. cuspidata* (Kochs) H. J. Veitch, *C. danzaiensis* H.T.Chang & K.M. Lan, *C. elongata* (Rehder & E.H.Wilson) Rehder, *C. euryoides* Lindley, *C. fraterna* Hance, *C. forrestii* (Diels) Cohen-Stuart, *C. grandibracteata* Hung T. Chang & F. L. Yu, *C. leptophylla* S.

Ye Liang ex Hung T. Chang, *C. oleifera* Abel, *C. pitardii* Cohen-Stuart, *C. pubicosta* Merr., *C. reticulata* Lindley and *C. yunnanensis* (Pitard ex Diels)]. Some of these taxa resulted in incomplete sequences, resulting in incorrect comparison outputs (e.g., *C. cuspidata*, *C. elongata*, *C. euryoides*, *C. fraterna*, *C. forrestii*, *C. oleifera*, *C. reticulata*, and *C. yunnanensis*). Notably, a polyA<sub>n</sub> sequence was detected in *trnH*<sup>(GUG)</sup>-*psbA*, representing a potentially diagnostic marker to discriminate species similar to *C. sinensis* (e.g., *C. sinensis* A<sub>13</sub> vs. *C. pubicosta* A<sub>14</sub>) but showing the intrinsic problematic nature of a polyN (e.g., stuttering, homoplasy). Thus, this marker was not used for taxonomical discrimination in the present study.

In contrast, the *rbcL* marker was used, reflecting both good discrimination power and universal primer amplification [31] that can also detect the macro-contamination of other species [56]. This marker resulted in two haplotypes for a single nucleotide polymorphism {SNP-A and SNP-C (M)} in the 68 bp coding region of *rbcL* (5'-AAATTGA-M-TTATTA-3', GenBank reference: KJ806281, 57209 bp; *C. sinensis* var. *sinensis*, complete chloroplast genome), consistent with the report of Stoeckle *et al.* [31]. Indeed, according to these authors, when geography, tea type and taxonomical characterizations were available for *C. sinensis* reference samples, this nucleotide variation (SNP-C vs. SNP-A) was strongly associated with products from India in comparison with China, with black vs. green tea and var. *assamica* vs. var. *sinensis*, respectively. Notably, based on the increase of new accessions in GenBank (e.g., new *Camellia* chloroplast genomes [57, 58, 59], currently the SNP-C shows a lower taxonomical discrimination level compared with SNP-A. In subsequent BLAST searches, haplotype SNP-A discriminates for *C. leptophylla*, *C. oleifera* and *C. sinensis*; in contrast, haplotype SNP-C is also associated with *C. sinensis*, *C. crapnelliana* Tutcher, *C. cuspidata*, *C. grandibracteata*, *C. granthamiana* Sealy (= *C. albogigas* Hu), *C. impressinervis* H.T.Chang & S.Y.Liang, *C. japonica* L., *C. kissii* Wall, *C. oleifera*, *C. petelotii* (Merr.) Sealy, *C. pubicosta*, *C. sasanqua* Thunb., *C. taliensis* (W. W. Sm.) Melch., *C. tonkinensis* (Pit.) Cohen-Stuart and *Tutcheria hirta* (Hand.-Mazz.) H. L. Li (= *Pyrenaria hirta* (Hand.-Mazz.) H. Keng).

According to the *matK* marker, even if its primers did not show good universality [31, 33, 34], this marker was used to confirm the presence of *C. sinensis* when *rbcL* results in SNP-A or to narrow the taxonomical field in the teabags when the SNP-C *rbcL* haplotype was present in the sample. BLAST analyses suggest that this barcoding molecular marker was able to discriminate the four *Camellia* species (*C. grandibracteata*, *C. leptophylla*, *C. pubicosta* and *C. sinensis*) with 100% identity.

Based on these results, an additional and exclusive marker has been demonstrated to discriminate *C. pubicosta* from *C. sinensis*, when the templates represent SNP-C in *rbcL*, as suggested above. Through the alignment of the chloroplast genomes of these two species, the intergenic region between *rps7* and *trnV*<sup>(GAC)</sup> was selected for the presence of two exclusive deletions in *C. pubicosta* (7 and 6 bp; 102651 and 102718 bp, respectively; GenBank reference: KJ806277, *C. pubicosta*, complete chloroplast genome). The oligos were designed between these two deletions, amplifying a 226 bp fragment in *C. pubicosta* and a 239 bp fragment in *C. sinensis* for both sequencing and genotyping analyses (Table 2).

**PCR capillary genotyping.** According to capillary genotyping electropherograms, ITS2 was excluded due to the presence of multiple peaks, as confirmed by previous sequencing results. Even if the *trnH*<sup>(GUG)</sup>-*psbA* intergenic spacer showed a nucleotide sequence shared with other allied *Camellia* species, this molecular marker was demonstrated as a better molecular marker than the P6 loop. In fact, *trnH*<sup>(GUG)</sup>-*psbA*, when compared with the P6 loop, showed a better discriminant power in terms of length (e.g., in eudicotyledons, the *trnH*<sup>(GUG)</sup>-*psbA* range is from 152 bp to 851 bp [60]; and the P6 loop range varied from 22 bp to 122 bp, [61]), and nucleotide variability [39, 60, 61]. Surely, the P6 loop can be useful with a higher



concentration of degraded DNA (i.e., ancient DNA), where it is difficult to obtain fragments longer than 150 bp [39, 62].

In conclusion, *trnH*<sup>(GUG)</sup>-*psbA* was employed as an important pre-screening marker to detect potential adulterants via genotyping, as this marker can individuate the polymorphisms of PCR as fragments of different lengths (observable as anomalous peaks) compared to the reference standard of *C. sinensis* and/or allied BLAST species, which result in the same length in bp (CS<sub>TP-ref</sub> = 510 bp). In parallel, *rps7-trnV*<sup>(GAC)</sup> was also used to discriminate *C. pubicosta* (226 bp) from *C. sinensis* (239 bp) for the difference in length of the amplified fragment (see above in paragraph “PCR sequencing”).

In the tea templates with only *C. sinensis* certified on the label (Table 1), the anomalous peaks with an RFU < 5%, (compared to the peak of *C. sinensis*) were interpreted as casual micro-contaminations and therefore excluded. Any deviation from the standard (i.e., reference profile of the listed ingredients) has been further investigated through cloning and sequencing the anomalous fragments.

## DNA Verity Test (DVT) applied to commercial teabags

**DNA extraction (Exgene Plant SV kit).** All DNA extracted from tea samples showed a degraded pattern on electrophoretic analysis; lower concentrations were observed in black tea and in decaffeinated tea samples (N11, N15 and V7). In these templates, electrophoretic analysis showed a more degraded DNA genomic pattern than the others. No DNA was detected in soluble tea samples (N16 and V16).

**PCR amplification {*trnH*<sup>(GUG)</sup>-*psbA* and *rps7-trnV*<sup>(GAC)</sup> fluorescently labeled; *rbcL* and *matK*}.** No amplification of any tested marker was obtained from the two soluble tea samples (N16 and V16). Cases of recalcitrant PCR amplifications were observed for all decaffeinated samples (V7, N7, N11 and N15) and some black tea samples (N1, N4, N7, N11, N15). These recalcitrant DNA samples were amplified using AmpONE Fast Pfu DNA Polymerase, which was more robust than Phusion DNA Polymerase.

**PCR genotyping {*trnH*<sup>(GUG)</sup>-*psbA* and *rps7-trnV*<sup>(GAC)</sup> fluorescently labeled}.** Using the genotyping of *trnH*<sup>(GUG)</sup>-*psbA* for pre-screening, all the samples showed a peak at 510 bp, corresponding to the *C. sinensis* reference (CS<sub>TP-ref</sub>). However, several samples were characterized based on anomalous peaks. For example, among the teabags that listed only *C. sinensis*, a black decaffeinated sample (N11) resulted in the presence of an additional, highly visible peak at 365 bp {RFU(365bp:510bp) = ca. 1:2}; in contrast, in the tea samples with additional material from other plants (admixture teas, V9-V11), the capillary genotyping data did not correspond with the number and intensity of peaks deduced by ingredients certified on the label, except for the confirmed presence of *C. sinensis* (peak at 510 bp). For example, V10 presents only an additional peak, which could be due to anise {ca. 317 bp; RFU (317 bp:510 bp) = ca. 1:3}, while the peaks of other certified ingredients, namely cinnamon (523 bp) and licorice (433 bp), were not detected. Low peaks (0.5% < RFU < 1.2%), assume to be from *Citrus* species (lemon, orange, grapefruit, and lime, ca. 552–577 bp) were individuated in the V9 green tea sample. A similar pattern was observed in the V11 sample (green tea), where a peak corresponding to mint (*Mentha* sp.) (ca. 436 bp) showed an RFU of approximately 1.7%, as referenced to the *C. sinensis* peak [i.e., RFU (436 bp:510 bp) = ca. 1:57].

As a consequence of these unexpected results, two further procedures were used: (1) the *trnH*<sup>(GUG)</sup>-*psbA* of samples with anomalous peaks numbers was sequenced in addition and successively cloned; and (2) the admixture samples (V9, V10 and V11) with a lower number or less marked peaks compared with the reference profiles of the listed ingredients were amplified using the primers for the *trnH*<sup>(GUG)</sup>-*psbA* specifically designed in the present study (Table 2)

(see paragraph, “Anomalous tea template”). The *trnH*<sup>(GUG)</sup>-*psbA* marker was also sequenced in the samples with normal capillary genotyping data (one peak at 510 bp) to generate reference Sanger chromatograms for comparison and no variability was observed among the accessions (GenBank accession: KY989996).

In addition, the genotyping of *rps7-trnV*<sup>(GAC)</sup> can be a suitable additional pre-screening short marker for eventual contaminants, even if this molecular region is not yet well represented in GenBank; in contrast, it has been useful to discriminate between *C. sinensis* and *C. pubicosta*. The capillary genotyping patterns confirmed the results obtained by the sequencing approach (see [results](#) in the next paragraph). Surely, the genotyping approach was faster and more efficient compared with Sanger sequencing.

**PCR sequencing {*rbcl+matK+rps7-trnV*<sup>(GAC)</sup>}**. According to the chromatograms of the *rbcl* sequences, the tea samples showed the presence of the two haplotypes (SNP-A and SNP-C; GenBank accession: KY989997 and KY989998, respectively) ([Table 1](#)), as also reported by Stoeckle *et al.* [31]. In the templates analyzed in the present study, among the green teas, six samples showed the presence of SNP-A, one sample indicated the presence of SNP-C and four showed heterozygous peaks (SNP-M), with adenine more represented than cytosine (V5 and V12, SNP-A/c) or similarly represented (V14 and V15, SNP-A/C). All the black tea samples showed the presence of SNP-C, except for five accessions, characterized by a heterozygous peak (SNP-M) ([Table 1](#)), with cytosine more represented than adenine (N5, N9, N11 and N13, SNP-C/a) and vice versa (N1, SNP-A/c). An additional oligo (CS\_*rbcl*-300rev) was designed as an additional assessment of this SNP ([Table 2](#)).

In the *rbcl* sequences of the samples with additional peaks in the *trnH*<sup>(GUG)</sup>-*psbA* genotyping profile, the corresponding Sanger *rbcl* sequences resulted in multiple peaks (N11, V10) or decreased background noise (V9, V11).

Regarding the *matK* sequences, no variability was observed among the tea accessions (GenBank accession: KY989999), and no multiple peaks were evident for the anomalous and/or admixture templates as in the corresponding *rbcl* sequences, confirming the low universality of these primers to detect potential DNA contaminations or different DNA in the admixture tea templates (e.g., [34]).

*Rps7-trnV*<sup>(GAC)</sup> was amplified and sequenced to discriminate *C. sinensis* (haplo-CS 239 bp; GenBank accession: KY990000) from *C. pubicosta* (haplo-CP 226 bp; GenBank accession: KY990001) in tea templates where the SNP-C haplotypes of *rbcl* were present. This marker was also sequenced in samples with SNP-A haplotypes of *rbcl* to generate reference Sanger chromatograms for comparison.

Seven black templates (N1-N5, N7 and N15) revealed the presence of *rps7-trnV*<sup>(GAC)</sup> belonging to *C. sinensis*; in contrast, several tea templates showed a sequence with multiple peaks, initiated by the first indel present in *C. sinensis* vs. *C. pubicosta* alignment. In some templates, the sequences were unreadable on account of the same height of the peaks (V9, V10, V11 and N9), but in the other ones, the sequence most represented can be deduced (V6, V8, N6, N8, N10-N14) ([Table 1](#)). In the latter case, the most represented sequence was easily analyzed and corresponded to that of *C. sinensis*. For the templates with peaks of the same intensity, a 2% agarose gel electrophoresis showed clear separation of the two bands; in contrast, in the other templates, the PCR was cloned and sequenced. The analyses of Sanger electropherograms applied to these fragments confirmed the presence of both *C. pubicosta* and *C. sinensis* in the same amplified templates ([Table 1](#)). As anticipated in the previous paragraph, these data were confirmed using a genotyping approach.

**Anomalous tea template.** According to the previous capillary genotyping data, the following two typologies of anomalous tea templates occur in our sampling: (I) (simple) sample with additional peak(s), even if *C. sinensis* was certified as the only ingredient (N11) and (II)

admixture samples with the number or height of the peaks lower than expected in terms of species listed as ingredients (V9, V10 and V11).

In the simple tea templates (i.e., point I), N11 (a black decaffeinated sample) showed the presence of *trnH*<sup>(GUG)</sup>-*psbA* Sanger-sequences with additional peaks. The sequences of *trnH*<sup>(GUG)</sup>-*psbA* clones of the additional peak were identical (GenBank accession: KY928308). The best identity (query cover) and E-value obtained were, respectively, only 84% (100%) and  $8e^{-86}$ , which was discriminatory for *Pipturus argenteus* (G. Forst.) Wedd. (Urticaceae, Boehmerieae) (GenBank accession: KU564710). These low values reflected the absence of an exhaustive data bank for this molecular marker. Thus, we also cloned the *rbcL* sequence, where multiple peaks were evident (see above). The obtained *rbcL* sequence of the contaminant (GenBank accession: KY928310) best matches with *Pouzolzia* sp. [585/585 bp (100%), query cover = 96%, E-value = 0, GenBank accession: KF138239] and namely with two accessions of *P. zeylanica* (L.) Benn. [583/585 bp (99.5%), query cover = 96%, E-value = 0, GenBank accession: KF138241; and 552/552 bp (100%), query cover = 91%, E-value = 0, GenBank accession: KF496389]. According to these results, we hypothesized the contaminant was a tropical taxon belonging to Urticaceae family and was likely a *Pouzolzia* taxon. Briefly, *Pouzolzia* Gaudich. (Urticaceae, Boehmerieae) is a pantropical genus of shrubs or herbs, [63, 64] represented by approximately 40 species [65, 66, 67, 68, 69] (only 32 species were assessed using The Plant List [70]). Currently, only the following six species are present in GenBank based on *rbcL* sequences: *P. australis* (Endl.) Friis & Wilmot-Dear, recently transferred from *Boehmeria* Jacq. [71], *P. calophylla* W.T. Wang & C.J. Chen (by the synonym *P. argenteonitida* W.T. Wang), *P. guineensis* Benth., *P. sanguinea* (Blume) Merr. (within var. *sanguinea* and var. *elegans* [Wedd.] Friis & Wilmot-Dear), *P. mixta* Solms, and *P. zeylanica*. Moreover, *matK* sequences are listed for only the following four species: *P. guineensis*, *P. mixta*, *P. sanguinea* (both for var. *sanguinea* and var. *elegans*), and *P. zeylanica*, indicating incomplete DNA barcoding characterization. *Pouzolzia* species are well known in the Old Tropical World as sources of fiber [63] and for their medicinal properties. For example, *P. zeylanica* (= *P. indica* Linn.) is also commonly used as an officinal plant (root, stem and leaves) for some medical remedies (e.g., mastitis, pyogenic infection, diarrhea, indigestion, and vulnerary) [72, 73, 74]. Moreover, *P. hirta* (L.) Hassk. var. *hirta* (= *P. bennettiana* Wight) is largely employed as a medicinal brew, referred to as “Oyik” in India [75]. *Pouzolzia zeylanica* or allied taxa are frequent in the area where *C. sinensis* is cultivated [63], explaining the remarkable presence of this plant (approximately 50%) as a filler in the black tea (N11) examined in this study.

In the admixture templates (i.e., point II), V10 had an obvious contaminated sequence, but only one certified additional ingredient was identified as anise {*trnH*<sup>(GUG)</sup>-*psbA* clones = *Pimpinella anisum/aromatic/flabellifolia* 100% identity; GenBank accession: KY928309}, while the other two are lacking (cinnamon and licorice). In contrast, V9 and V11 showed lower additional peaks at ca. 430 bp for *Mentha* sp. and ca. 552–577 bp for *Citrus* sp., respectively. These admixture templates were amplified using specific primers for the plant species identified in the ingredient list but were not clearly assessed in the genotyping and sequencing/cloning data.

According to the PCR results, in the template V10, no amplification fragments were obtained using the primers specific for cinnamon and licorice, confirming the absence of these two certified ingredients. According to the amplification and cloning results, anise was well represented in this tea green template and did not correspond with the quantity reported on the labels (2%). Considering the two other green templates, V9 and V11, positive amplifications were obtained using the primers specific for *Citrus* and *Mentha* sp., respectively, confirming the presence of these plants but not in the quantity listed on these green tea templates (i.e., *Citrus* 2% and *Mentha* 7.9% in V9 and V11, respectively).

**Taxonomical characterization and systematic considerations.** The level of specific discrimination was optimal, confirming both the presence of *C. sinensis* in the tea templates through the joint use of *rbcL*+*matK*+*rps7-trnV*<sup>(GAC)</sup> molecular markers. *Rps7-trnV*<sup>(GAC)</sup> is an additional and exclusive marker, which has been fundamental to discriminate *C. sinensis* from *C. pubicosta*, where the *rbcL* sequence has the nucleotide cytosine at 68 bp (SNP-C) (Table 1).

In summary, *rbcL*(SNP-A)+*matK* sequences can identify *C. sinensis* and *C. leptophylla*; in contrast, *rbcL*(SNP-C)+*matK*+*rps7-trnV*<sup>(GAC)</sup> matches *C. sinensis* and *C. grandibracteata*. This species and *C. leptophylla* are critical taxa, as these species are also regarded as mere varieties of *C. sinensis* or as not distinct from it and, thus, as potentially only two wild crop relatives of the cultivated tea [76, 77]. Moreover, both *C. leptophylla* and *C. grandibracteata* were characterized at an obscure or limited native range. Indeed, *C. leptophylla* has been identified only at the top of Daqing Mountain (Guangxi, western China) [78], while pertinent herbarium specimens from two further localities in Guangxi were subsequently identified as *C. sinensis* [77]. In addition, *C. grandibracteata* is exclusively identified from tea gardens along the Mekong River in Yunxian (south-west China) [76], and it would not seem to exist in the wild. Therefore, this species might result from spontaneous hybridization or introgression among cultivated plants (Shi-Xiong Yang and Dong-Wei Zhao, personal communication in [76]).

Based on the presence of the heterozygosity peaks in the *rbcL* sequence (SNP-A/C) and both fragments of the *rps7-trnV*<sup>(GAC)</sup> sequence (haplo-CS 239 bp/haplo-CP 226 bp) (see Table 1), the leaf matrices present in some teabags could comprise the following: (1) an admixture of two haplotypes belonging to *C. sinensis* (potentially ascribable to var. *sinensis* and var. *assamica*, see [31]), (2) a mixture of *C. sinensis* and *C. pubicosta*, or (3) an admixture of *C. sinensis* and hybrids of *C. pubicosta* (♀)/*C. sinensis* (♂). These speculations are based on the presence of maternal inheritance in the chloroplast of *C. sinensis* according to Kaundun and Matsumoto [79].

In detail, *C. pubicosta* is a critical species, strictly allied to *C. sinensis*. The first specimens collected from Mount Bavi (currently in Vietnam) were tentatively included in *C. sinensis*. However, later, Merrill [80] recognized a new species by analyzing additional material, distinct from *C. sinensis* on account of relevant features in his opinion. Indeed, some of the vegetative features reported in the protologue [80] could be included in the variability of *C. sinensis*. For example, the acuminate-caudate leaves also occur in *C. sinensis* var. *assamica*. However, *C. pubicosta* appears distinct from *C. sinensis* s.l. based on other relevant traits, chiefly due to its three styles, which are free and not connate as in *C. sinensis*. Thus, *C. pubicosta* was placed in the sect. *Corallinae* Sealy, and *C. sinensis* belongs to the sect. *Thea* (L.) Benth. & Hook. [81]. However, this point is controversial, as other scholars [82], in contrast, also include *C. pubicosta* in the sect. *Thea*. According to the recent chloroplast genome phylogeny of Huang *et al.* [59], the authors likely put an end to the dispute, as based on their analysis, *C. pubicosta* was identified as a sister group to both *C. sinensis* var. *assamica* and *C. grandibracteata* terminal and therefore is better placed in the sect. *Thea*. According to Huang *et al.* [59], the character of connate vs. free styles was not consistent. Finally, considering the phylogenetic distance of the investigated taxa in *Camellia* (see, *C. sinensis* var. *assamica* vs. *C. sinensis* var. *sinensis*), some species falling in the *C. sinensis* varieties clade included *C. pubicosta* and *C. grandibracteata* (cfr. Fig 9 [59]) but as further varieties and not different species.

## Conclusion

After packaging, the species from which the tea samples are obtained are morphologically unrecognizable. Therefore, we cannot know with certainty using traditional methods if the source plants are actually those reported in the listed ingredients. For this reason, in the

present study, a standardized protocol of DNA barcoding characterization (*DNA Verity Test*, *DVT*) was developed to increase both efficiency and rapidity and to become current with the new molecular data submitted to GenBank. After the optimization of the *DVT* protocol (extraction efficiency, DNA polymerase fidelity, reliability amplification and the selection of barcoding markers), the 32 green/black tea samples of *C. sinensis* were analyzed in a blind test. The *DVT* protocol comprises two approaches used jointly, i.e., (1) a rapid pre-screening analysis through a genotyping approach using *trnH*<sup>(GUG)</sup>-*psbA* and *rps7-trnV*<sup>(GAC)</sup> for individuating any trace of contamination by other plants (revealed by anomalous peaks in the capillary row data) and (2) a Sanger sequencing approach using the *rbcL+matK±rps7-trnV*<sup>(GAC)</sup> to confirm both the presence of *C. sinensis* and identify food frauds (i.e., contamination, absence of certified ingredients).

We can summarize some of the main results as follows:

1. The *DVT* protocol is suitable for detecting adulteration in tea matrices (contamination or the absence of certified ingredients) and can be applied to study other similar systems.
2. Using BLAST analysis of the sequences of *rbcL+matK±rps7-trnV*<sup>(GAC)</sup> chloroplast markers, *C. sinensis* can be taxonomically characterized.
3. *Rps7-trnV*<sup>(GAC)</sup> can be employed to discriminate *C. sinensis* from *C. pubicosta*.
4. ITS2 is not an ideal DNA barcode for tea samples because of potential incomplete lineage sorting and hybridization/introgression phenomena in *C. sinensis* taxa.
5. This genotyping approach is an easy, inexpensive and rapid pre-screening method to detect anomalies in the tea templates.
6. Two herbal companies provided no authentic products with a contaminant (N11) or without some of the listed ingredients (V10).
7. The leaf matrices present in some teabags could comprise an admixture of different *C. sinensis* haplotypes and/or allied species (*C. pubicosta*).

## Acknowledgments

The authors would like to thank GeneAll Biotechnology Co. Ltd (Seoul, South Korea) and MW Biotech (Naples, Italy) for supplying the reagents for the DNA barcoding characterization; Tema Ricerca (Bologna, Italy) for technical support with GeneAll Biotechnology; Dr. Antonio Manco (Life Technologies), Dr. Sara Monari (Tema Ricerca), Drs. Paul Koo and Wanju Lee (GeneAll Biotechnology Co. Ltd) for their kind assistance with the *DNA Verity Test* project.

The authors would also like to thank the two anonymous reviewers for their helpful suggestions.

## Author Contributions

**Conceptualization:** ODC.

**Data curation:** ODC.

**Formal analysis:** ODC MC.

**Funding acquisition:** ODC BM JT FA MT MG.

**Investigation:** ODC MC ADM.



**Methodology:** ODC.

**Project administration:** ODC.

**Resources:** ODC.

**Supervision:** ODC EDG.

**Validation:** ODC MC.

**Writing – original draft:** ODC.

**Writing – review & editing:** ODC EDG.

## References

1. Nath TN. Determination of heavy metals and its impact on tea cultivated soil and tea leaf in Dibrugarh district of Assam, India. *J Adv Res.* 2013; 1: 561–567.
2. Sarma H, Islam NF, Borgohain P, Sarma A, Prasad MNV. Localization of polycyclic aromatic hydrocarbons and heavy metals in surface soil of Asia's oldest oil and gas drilling site in Assam, northeast India: Implications for the bio-economy. *Emerging Contaminants.* 2016; 2: 119–127.
3. Bianchi F, Comba P, Martuzzi M, Palombino R, Pizzuti R. Italian “Triangle of death”. *Lancet Oncol.* 2004; 5(12): 710. [https://doi.org/10.1016/S1470-2045\(04\)01645-6](https://doi.org/10.1016/S1470-2045(04)01645-6) PMID: 15581540
4. Senior K, Mazza A. Italian “Triangle of death” linked to waste crisis. *Lancet Oncol.* 2004; 5(9): 525–527. PMID: 15384216
5. Primrose S, Woolfe M, Rollinson S. Food forensics: methods for determining the authenticity of food-stuffs. *Trends Food Sci Tech.* 2010; 21: 582e590.
6. Ellis DI, Brewster VL, Dunn WB, Allwood JW, Golovanov AP, Goodacre R. Fingerprinting food: current technologies for the detection of food adulteration and contamination. *Chem Soc Rev.* 2012; 41: 5706–5727. <https://doi.org/10.1039/c2cs35138b> PMID: 22729179
7. De Medici D, Kuchta T, Knutsson R, Angelov A, Auricchio B, Barbanera M, et al. Rapid Methods for Quality Assurance of Foods: the Next Decade with Polymerase Chain Reaction (PCR)-Based Food Monitoring. *Food Anal Method.* 2015; 8(2): 255–271.
8. Galimberti A, Bruno A, Mezzasalma V, De Mattia F, Bruni I, Labra M. Emerging DNA-based technologies to characterize food ecosystems. *Food Res Int.* 2015; 69: 424–433.
9. Danezis GP, Tsagkaris AS, Camin F, Brusica V, Georgiou CA. Food authentication: Techniques, trends & emerging approaches. *Trends Analyt Chem.* 2016; 85: 123–132.
10. Peris M, Escuder-Gilbert L. Electronic noses and tongues to assess food authenticity and adulteration. *Trends Food Sci Tech.* 2016; 58: 40–54.
11. Spink J, Moyer DC. Defining the Public Health Threat of Food Fraud. *J Food Sci.* 2011; 76: R157–163. <https://doi.org/10.1111/j.1750-3841.2011.02417.x> PMID: 22416717
12. Arlt VM, Stiborova M, Schmeiser HH. Aristolochic acid as a probable human cancer hazard in herbal remedies: a review. *Mutagenesis.* 2002; 17(4): 265–277. PMID: 12110620
13. Newmaster SG, Grguric M, Shanmughanandhan D, Ramalingam S, Ragupathy S. DNA barcoding detects contamination and substitution in North American herbal products. *BMC Med.* 2013; 11: 222. <https://doi.org/10.1186/1741-7015-11-222> PMID: 24120035
14. Mezzasalma V, Ganopoulos I, Galimberti A, Cornara L, Ferri E, Labra M. Poisonous or non-poisonous plants? DNA-based tools and applications for accurate identification. *Int. J. Legal Med.* 2016; 131: 1–19. <https://doi.org/10.1007/s00414-016-1460-y> PMID: 27796590
15. Ekor M. The growing use of herbal medicines: issues relating to adverse reactions and challenges in monitoring safety. *Front Pharmacol.* 2014; 4: 177. <https://doi.org/10.3389/fphar.2013.00177> PMID: 24454289
16. Šoukand R, Quave CL, Pieroni A, Pardo-de-Santayana M, Tardío J, Kalle R, et al. Plants used for making recreational tea in Europe: a review based on specific research sites. *J Ethnobiol Ethnomed.* 2013; 9: 58. <https://doi.org/10.1186/1746-4269-9-58> PMID: 23941692
17. Menale B, De Castro O, Cascone C, Muoio R. Ethnobotanical investigation on medicinal plants in the Vesuvio National Park (Campania, Southern Italy). *J Ethnopharmacol.* 2016; 192: 320–349. <https://doi.org/10.1016/j.jep.2016.07.049> PMID: 27457693
18. Ravikumar C. Review on Herbal Teas. *J Pharm Sci & Res.* 2014; 6(5): 236–238.

19. World Health Organization. Food Safety: What you should know. World Health Day: 7 April 2015. SEARO Library. [http://www.searo.who.int/entity/world\\_health\\_day/2015/whd-what-you-should-know/en/#](http://www.searo.who.int/entity/world_health_day/2015/whd-what-you-should-know/en/#) (Accessed January 2017).
20. Lees M. Food Authenticity and Traceability. Cambridge: Elsevier Science; 2003.
21. EFSA. Scientific Cooperation (ESCO). Working Group on Botanicals and Botanical Preparations; Advice on the EFSA guidance document for the safety assessment of botanicals and botanical preparations intended for use as food supplements, based on real case studies on request of EFSA. EFSA J. 2009; 7(9): 280.
22. Bruni I, De Mattia F, Galimberti A, Galasso G, Banfi E, Casiraghi M, et al. Identification of poisonous plants by DNA barcoding approach. Int J Legal Med. 2010; 124(6): 595–603. <https://doi.org/10.1007/s00414-010-0447-3> PMID: 20354712
23. Bruni I, Galimberti A, Caridi L, Scaccabarozzi D, De Mattia F, Casiraghi M, Labra M. A DNA barcoding approach to identify plant species in multiflower honey. Food Chem. 2015; 170: 308–315. <https://doi.org/10.1016/j.foodchem.2014.08.060> PMID: 25306350
24. Galimberti A, De Mattia F, Losa A, Bruni I, Federici S, Casiraghi M, et al. DNA barcoding as a new tool for food traceability. Food Res Int. 2013; 50: 55–63.
25. Mishra P, Kumar A, Nagireddy A, Mani DN, Shukla AK, Tiwari R, Sundaresan V. DNA barcoding: an efficient tool to overcome authentication challenges in the herbal market. Plant Biotechnol J. 2016; 14(1): 8–21. <https://doi.org/10.1111/pbi.12419> PMID: 26079154
26. Gutman RL, Ryu BH. Rediscovering tea. An exploration of the scientific literature. Herbal Gram. 1996; 37: 33–48.
27. Sumpio BE, Cordova AC, Berke-Schlessel DW, Qin F, Chen QH. Green tea, the "Asian paradox," and cardiovascular disease. J Am Coll Surg. 2006; 202(5): 813–825. <https://doi.org/10.1016/j.jamcollsurg.2006.01.018> PMID: 16648021
28. Reygaert CW. An Update on the Health Benefits of Green Tea. Beverages. 2017; 3(1): 6.
29. Vancouver JF. What Lies in the Future of Tea Science in the Era of Genomics? J Tea Science Res. 2011; 1(1): 1–3.
30. Dhiman B, Singh M. Molecular Detection of Cashew Husk (*Anacardium occidentale*) Adulteration in Market Samples of Dry Tea (*Camellia sinensis*). Planta Med. 2003; 69(9): 882–884. <https://doi.org/10.1055/s-2003-43211> PMID: 14598225
31. Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP. Commercial teas highlight plant DNA barcode identification successes and obstacles. Sci Rep. 2011; 1: 42. <https://doi.org/10.1038/srep00042> PMID: 22355561
32. CBOL, Plant Working Group. A DNA barcode for land plants. Proc Natl Acad Sci USA. 2009; 106: 12794–12797. <https://doi.org/10.1073/pnas.0905845106> PMID: 19666622
33. Hollingsworth PM, Graham SW, Little DP. Choosing and Using a Plant DNA Barcode. PLoS ONE. 2011; 6(5): e19254. <https://doi.org/10.1371/journal.pone.0019254> PMID: 21637336
34. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S. Plant DNA barcoding: from gene to genome. Biol Rev Camb Philos Soc. 2015; 90(1): 157–166. <https://doi.org/10.1111/brv.12104> PMID: 24666563
35. Kress J, Erickson DL. A two-locus global DNA barcode for land plants: the coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE. 2007; 2(6): e508. <https://doi.org/10.1371/journal.pone.0000508> PMID: 17551588
36. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, et al. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE. 2008; 3(7): e2802. <https://doi.org/10.1371/journal.pone.0002802> PMID: 18665273
37. Sang T, Crawford DJ, Stuessy TF. Chloroplast DNA phylogeny, reticulate evolution and biogeography of *Paeonia* (Paeoniaceae). Am J Bot. 1997; 84: 1120–1136. PMID: 21708667
38. Tate JA, Simpson BB. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. Syst Bot. 2003; 28: 723–737.
39. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast *trnL*<sup>(UAA)</sup> intron for plant DNA barcoding. Nucleic Acids Res. 2007; 35(3): e14.
40. Chen S, Yao H, Han J, Liu C, Song J, Shi L, et al. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. PLoS ONE. 2010; 5(1): e8613. <https://doi.org/10.1371/journal.pone.0008613> PMID: 20062805
41. Chiou SJ, Yen JH, Fang CL, Chen HL, Lin TY. Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers. Planta Med. 2007; 73(13): 1421–1423. <https://doi.org/10.1055/s-2007-990227> PMID: 17909989

42. Vijayan K, Tsou C-H. Technical report on the molecular phylogeny of *Camellia* with nrITS: the need for high quality DNA and PCR amplification with Pfu-DNA polymerase. *Bot Stud.* 2008; 49: 177–188.
43. De Castro O, Brullo S, Colombo P, Jury S, De Luca P, Di Maio A. Phylogenetic and biogeographical inferences for *Panocratium* (Amaryllidaceae), with an emphasis on the Mediterranean species based on plastid sequence data. *Bot J Linn Soc.* 2012; 170: 12–28.
44. Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PLoS ONE.* 2012; 7(8): e43808. <https://doi.org/10.1371/journal.pone.0043808> PMID: 22952770
45. Di Maio A, De Castro O. SSR-patchwork: an optimized protocol to obtain a rapid and inexpensive SSR library using first generation sequencing technology. *Appl Plant Sci.* 2013; 1: 1200158.
46. Hall TA. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT *Nucleic Acids Symp Ser.* 1999; 41: 95–98.
47. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman JD. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25: 3389–3402. PMID: 9254694
48. Ratnasingham S, Hebert PD. Bold: the barcode of life data system. *Mol Ecol Notes.* 2007; 7: 355–364 <http://www.barcodinglife.org/> <https://doi.org/10.1111/j.1471-8286.2007.01678.x> PMID: 18784790
49. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Res.* 2013; 41: D36–D42. <https://doi.org/10.1093/nar/gks1195> PMID: 23193287
50. Schrader C, Schielke A, Ellerbroek L, John R. PCR inhibitors—occurrence, properties and removal. *J Appl Microbiol.* 2012; 113: 1014–1026. <https://doi.org/10.1111/j.1365-2672.2012.05384.x> PMID: 22747964
51. Graham HN. Green tea composition, consumption, and polyphenols chemistry. *Prev Med.* 1992; 21: 334–350. PMID: 1614995
52. Fazekas AJ, Steeves R, Newmaster SG, Hollingsworth PM. Stopping the stutter: Improvements in sequence quality from regions with mononucleotide repeats can increase the usefulness of non-coding regions for DNA barcoding. *Taxon.* 2010; 59: 694–697.
53. Yang JB, Li HT, Yang SX, Li DZ, Yang YY. The application of four DNA sequences to studying molecular phylogeny of *Camellia* (Theaceae). *Acta Bot Yunnanica.* 2006; 28: 108–114.
54. Vijayan K, Zhang W-J, Tsou C-H. Molecular taxonomy of *Camellia* (Theaceae) inferred from nrITS sequences. *Am J Bot.* 2009; 96: 1348–1360. <https://doi.org/10.3732/ajb.0800205> PMID: 21628283
55. Lee S-C, Wang C-H, Yen C-E, Chang C. DNA barcode and identification of the varieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences. *J Food Drug Anal.* 2016, in press;
56. Zhang Y, Skaar I, Sulyok M, Liu X, Rao M, Taylor JW. The Microbiome and Metabolites in Fermented Pu-erh Tea as Revealed by High-Throughput Sequencing and Quantitative Multiplex Metabolite Analysis. *PLoS ONE.* 2016; 11(6): e0157847. <https://doi.org/10.1371/journal.pone.0157847> PMID: 27337135
57. Shi C, Liu Y, Huang H, Xia E-H, Zhang H-B, et al. Contradiction between Plastid Gene Transcription and Function Due to Complex Posttranscriptional Splicing: An Exemplary Study of *ycf15* Function and Evolution in Angiosperms. *PLoS ONE.* 2013; 8(3): e59620. <https://doi.org/10.1371/journal.pone.0059620> PMID: 23527231
58. Yang J-B, Yang S-X, Li H-T, Yang J, Li D-Z. Comparative Chloroplast Genomes of *Camellia* Species. *PLoS One.* 2013; 8(8): e73053. <https://doi.org/10.1371/journal.pone.0073053> PMID: 24009730
59. Huang H, Shi C, Liu Y, Shu-Yan Mao, Gao L-Z. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol Biol.* 2014; 14: 151. <https://doi.org/10.1186/1471-2148-14-151> PMID: 25001059
60. Pang X, Liu C, Shi L, Liu R, Liang D, Li H, et al. Utility of the *trnH-psbA* Intergenic Spacer Region and Its Combinations as Plant DNA Barcodes: A Meta-Analysis. *PLoS ONE.* 2012; 7(11): e48833. <https://doi.org/10.1371/journal.pone.0048833> PMID: 23155412
61. Willerslev E, Davison J, Moora M, Zobel M, Coissac E, Edwards ME, et al. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature.* 2014; 506(7486): 47–51. <https://doi.org/10.1038/nature12921> PMID: 24499916
62. Spaniolas S, Bazakos C, Spano T, Zoghby C, Kalaitzis P. The potential of plastid *trnL*<sup>(UAA)</sup> intron polymorphisms for the identification of the botanical origin of plant oils. *Food Chem.* 2010; 122: 850–856.
63. Jiarui C, Friis I, Wilmot-Dear CM. *Pouzolzia* Gaudichaud-Beaupré. In: Wu ZY, Raven PH, Hong DY, editors. *Flora of China* 5. Beijing and St. Louis: Science Press and Missouri Botanical Garden Press; 2003.

64. Kravtsova TI, Friis I, Wilmot-Dear CM. Morphology and anatomy of fruits in *Pouzolzia* (Urticaceae) in Relation to Taxonomy. *Kew Bull.* 2003; 58(2): 297–327.
65. Wilmot-Dear CM, Friis I. The New World species of *Boehmeria* and *Pouzolzia* (Urticaceae, tribus Boehmerieae). A taxonomic revision. *Opera Bot.* 1996; 129: 1–103.
66. Wilmot-Dear CM, Friis I. The Old World species of *Pouzolzia* (Urticaceae, tribus Boehmerieae). A taxonomic revision. *Nord J Bot.* 2006; 24: 5–111.
67. Wilmot-Dear CM, Friis I. New world *Pouzolzia* and *Boehmeria* (Urticaceae): a new species and new generic record for Paraguay, *Pouzolzia amambaiensis*, and additional observations on already described species of both genera. *Nord J Bot.* 2011; 29: 691–695.
68. Wilmot-Dear CM, Acharya N, Kravtsova TI, Friis I. *Pouzolzia rugulosa* transferred from *Boehmeria*, and the distinction between *Boehmeria* and *Pouzolzia* (Urticaceae). *Edinb J Bot.* 2009; 66: 51–64.
69. Wilmot-Dear CM, Friis I, Monro AK. *Pouzolzia saxophila* sp. nov. (Urticaceae tribe Boehmerieae) from Bahia, Brazil. *Nord J Bot.* 2014; 24: 354–357.
70. The Plant List. Version 1.1. 2013. Available to: <http://www.theplantlist.org/> (Accessed December 2016).
71. Wilmot-Dear CM, Friis I. The Old World species of *Boehmeria* (Urticaceae, tribu Boehmerieae). A taxonomic revision. *Blumea.* 2013; 58: 85–216.
72. Bhattacharjya DK, Bora PC. Medicinal weeds of crop fields and role of women in rural health and hygiene in Nalbari district, Assam. *Indian J Tradit Know.* 2008; 7(3): 501–504.
73. Venkata Ratnam K, Venkata Raju RR. Traditional Medicine Used by the Adivasis of Eastern Ghats, Andhra Pradesh—For Bone Fractures. *Ethnobot Leaflets.* 2008; 12: 19–22.
74. Li P, Huo L, Su W, Lu R, Deng C, Liu L, et al. Free radical-scavenging capacity, antioxidant activity and phenolic content of *Pouzolzia zeylanica*. *J Serb Chem Soc.* 2011; 76(5): 709–717.
75. Singh RK, Srivastava RC, Adi Community, Monpa Community. Bioculturally important plant diversity of Arunachal Pradesh: Learning from Adi and Monpa communities about “Future crops of India”. *Indian J Tradit Know.* 2010; 9 (4): 754–759.
76. Wheeler L, Rivers MC. *Camellia grandibracteata*. The IUCN Red List of Threatened Species, 2015: e. T61859241A61859325. (Accessed December 2016)
77. Wheeler L, Rivers MC. *Camellia leptophylla*. The IUCN Red List of Threatened Species. 2015: e. T62033755A62033769.
78. Jiyin G, Parks CR, Du Y. Collected species of the genus *Camellia*: An illustrated outline. Zhejiang: Zhejiang Science and Technology Press; 2005.
79. Kaundun SS, Matsumoto S. Molecular evidence for maternal inheritance of the chloroplast genome in tea, *Camellia sinensis* (L.) O. Kuntze. *J Sci Food Agric.* 2011; 91: 2660–2663. <https://doi.org/10.1002/jsfa.4508> PMID: 21769876
80. Merrill ED. Records of Indo-Chinese plants III. *J Arnold Arbor.* 1942; 23: 156–197.
81. Min T, Bartholomew B. *Camellia* L. In: Wu ZY, Raven PH, editors. *Flora of China*. Vol. 12 (Hippocastanaceae through Theaceae). Beijing and St. Louis: Science Press and Missouri Botanical Garden Press; 2007.
82. Chang HD, Ren SX. *Flora of China*. Science Press. 1998; 49(3): 1–251.