Biodiversity of South Indian tea clones with detection of plant-based adulterants in tea dust using DNA barcoding

Steffi Avarave and Jibu Thomas

Algae Biomass Research Laboratory, Department of Biotechnology, Karunya Institute of Technology and Sciences (Deemed to be University), Coimbatore, Tamil Nadu, India

ABSTRACT

Tea is by and large a highly penetrated product in south India. Hence the adulteration risk in tea dust gets hiked in the markets. We constructed a standard database using plant plastid markers (rbcL, matK, trnH-psbA, rpoC, rpoB, ycf 1) and nuclear (ITS2) locus from prominent south Indian tea clones representing Assam, China, and Camobd varieties. These barcodes were used as reference algorithm to investigate the authenticity of 10 sampled commercial tea dust by recovering its DNA barcodes using rbcL, matK, and ITS2 loci. PCR amplification success, sequencing efficiency, genetic polymorphisms, BLAST search, and phylogenetic analysis were performed to enhance genotypic information on south tea cultivars and in authenticating the commercial samples of Camellia sinensis. Findings suggest that the chloroplast and nuclear loci can identify tea plant at the genus and varietal level respectively and rbcL as the potential marker for detecting plant-based admixtures coupled with TA cloning after DNA barcoding.

ARTICLE HISTORY

Received 6 August 2021
Accepted 24 October 2021

KEYWORDS

South Indian tea cultivars; commercial tea products; adulteration; DNA barcoding; TA cloning

1. Introduction

India has achieved prominent status on the global tea map. According to the Indian tea market report 2021, consumption in the country reached a volume of 1.10 million...
tons of tea in 2020. Tea market is further expected to expand at a compound annual growth rate of 4.2% in the forecast period of 2021–2026 to attain 1.40 million tons by 2026. The significance of the aqueous infusion of *Camellia sinensis* (Tea) among consumers is mainly of its functional components which have antioxidant, anti-bacterial, anti-viral, antiproliferative, and UV dermal protective effects (Sharma et al. 2020; Kuo et al. 2021). As the quantity of tea dust became demanding, the quality of the tea product became questionable which poses a major threat to consumer confidence. Markets are flooded with varied adulterated tea dust which spoiled the reputation of Indian tea even though the country stands second in tea cultivation (Stavrianidi et al. 2018). Significantly, the tea adulteration was considered a moral outrage since tea’s virtues were established as medicine, alimentary, economic, and social. Even though the adulteration of tea products has gotten a lot of attention in recent years, the extensive pulverizing of the vegetal material (two leaves and a bud) allows little physical characteristics to identify the plant-based substitute for *Camellia sinensis* supports as the major drawback (Faller et al. 2019). To address this problem the investigation was applied using the DNA barcoding technique, a widely adopted approach to detect the core tea plant from the possible adulterants, substitutions, fillers, and flavoring agents used in the tea products.

The authors primarily aimed to verify the authenticity of selected tea clones from south India as the southern region showed significance as a higher proportion of tea is grown and consumed compared to north, central, and east. The cultivated tea varieties of south India belong to *Camellia* species with three varieties viz., *Camellia sinensis var. sinensis*, *Camellia sinensis var. assamica*, and *Camellia sinensis var. lasiocalyx*. Due to the use of seedlings in the plantations, south Indian tea accessions are genetically diverse, and many of the existing accessions have been selected from huge populations primarily based on their phenotypic superiority as part of the selection programs (Saravanan et al. 2005; Li et al. 2020). Using the DNA barcoding technique, the genetic information can be gleaned from both coding and non-coding phylogenetic markers that have significant genetic variations among taxa. Dissection of signature motifs assists not in just detecting substitutions/contaminants but also in distinguishing distinct *Camellia* types. With successful DNA barcodes, a library of known species be generated and is matched, or assigned with the barcode sequence of the unknown sample against the barcode library for identification of adulterants or fillers (Pazi and Crawford 2020).

This paper outlines (1) Construction of standard reference library database for south Indian tea cultivars using plant chloroplast coding markers like *rbcL* (Ribulose 1, 5 bisphosphate carboxylase/oxygenase), *matK* (Maturase kinase), *rpoB*, *rpoC* (RNA polymerase β subunit), *ycf 1* (Hypothetical chloroplast open reading frame 1), non-coding *trnH-psbA* intergenic spacer, and nuclear *ITS2* (Internal transcribed spacer 2) gene. (2) Analysis of genetic information contents to differentiate the clonal varieties. (3) Retrieval of DNA barcodes from *rbcL*, *matK*, and *ITS2* for commercial products sampled for authentication. (4) Usage of cloning after DNA barcoding method to detect admixture samples among commercial products which are not detected by DNA barcoding technique alone.
2. Results and discussion

DNA of high molecular weight was verified from all reference tea clones shown in Table S1. An absorbance ratio (A260/A280) of 1.6–1.8 indicated insignificant levels of protein, phenol, or other contaminants in the extracted DNA (Figure S1). The Polymerase chain reaction (PCR) conditions for each locus investigated in the study were optimized and tabularized (Table S2). The Amplification efficiency of loci studied was found to be 100% and rbcL yielded a sequence length of ~740 bp, matK fragment of ~850 bp, trnH - psbA ~ 500 bp, rpoB ~ 550 bp, rpoC ~ 550 bp, ycf 1 ~ 900 bp, and nuclear gene ITS2 ~500 bp for all the samples analyzed (Figure S2). A reliable reference database with sufficient nucleobase variability to detect specimens at the genus and species level is considered prerequisite for successful DNA barcoding investigation. In this regard, authors recovered 165 barcodes of Camellia sinensis cultivars using universal primer sets (rbcL-40, matK-40, trnH - psbA-20, rpoB-20, rpoC-20, ycf 1-10, and ITS2-15). Generated accessions were submitted in the public database GenBank to assemble ‘Standard reference material for south Indian tea germplasm’ (Table S3). The barcodes were generated under the project title ‘National Tea research foundation tea DNA Barcodes’ (NTDB) in the Barcode of Life Data (BOLD) system.

The suitable candidate barcode that could enhance information on south Indian tea clones was analyzed. As the sequence tags in chloroplast regions are mostly conserved, only a very few sites are gleaned from the plastid region that could enhance information on south Indian tea clones. The conserved site observed in rbcL was 528/530 bp, matK- 726/742 bp, trnH-psbA- 253/254 bp, rpoC- 432/432 bp, rpoB- 328/330 bp, and ycf 1- 383/390 bp. The interspecific distance mean of rbcL, matK, trnH-psbA, rpoC, rpoB, and ycf 1 were 0.0002, 0.004, 0.004, 0, 0.0006, and 0.0004 respectively. Coding regions like rbcL, rpoB, and rpoC sites were invariant in the tea cultivars sampled in the investigation. Stoeckle et al. (2011) reported an A/C nucleobases variation at position 68 in the rbcL region associated with China and Indian tea cultivars. The authors confirmed a ‘C’ variant corresponding to the position with predicted amino acid threonine from all three varieties of south India (Figure S3a). The matK sequence of tea germplasm has a 9 bp deletion except for a single clone TRI2043 where 9 bp addition was observed (Figure S3b). Thuvarakia et al. (2017) claimed that tea cultivar TRI2043 was phenotypically distinct from other tea cultivars due to its high anthocyanin content and has high tolerance against diseases. Even genotypically the TRI2043 clone gleaned from the matK region had 9 bp addition and five SNPs which distinguish it from other tea cultivars. The authors observed an informative site (C/T) in the matK locus at a site corresponding to position 16 of the coding region with predicted amino acid either serine (16 C) or Phenylalanine (16 T). The ‘C’ variation was found in Sri Lankan clones, estate selections, and few UPASI clones that were widely planted in southern India. Seedling obtained from Ooty also carried the ‘C’ variant, suggesting their genetic similarity to these cultivars. Certain UPASI clones and seedlings sampled from Coonoor belonging to Assam or and Cambod varieties have the ‘T’ nucleobase variation. Non-coding trnH-psbA and coding ycf 1 regions showed ‘A’, and/or ‘T’, mono-nucleotide repeats and the size of the repeats ranged between 1 and 30 nucleotides respectively (Figure S3c, d). The Camellia spp used in southern states comes under one species so one of the key objectives of the study was to find a marker that can
discriminate the *Camellia spp* based on their varieties. Nuclear region marker *ITS2* was frequently used in DNA barcoding because of its high species discrimination power. Lee et al. (2017) opined *ITS2* gene could efficiently predict variations among varieties and cultivars apart from species differentiation. Consistent with the literature, when analyzing the nuclear region of the south tea cultivars a 9 bp indel and three parsimony informative sites (G/A, C/T, and A/G) were observed in the sequences aligned. The secondary structure of the *ITS2* tea cultivars showed variations in the stem-loop differentiating China variety from Assam and Cambod varieties (Figure S4a–c). The phylogeny was evaluated for all the loci using the unweighted pair group method with arithmetic mean (UPGMA) showed nuclear marker *ITS2* could distinguish plants at a varietal level (Figure S4d) compared with plastid markers where most of the tea clones fixed to the same clade (Figure S5). So the *ITS2* gene of nuclear ribosomal DNA is considered as an ideal locus for discriminating the varieties of the Theaceae family.

The secondary demand of the genomic profiling was to recover the barcode sequence from the commercial dust tea specimens. The sequence has to be recovered accurately so it can be analyzed with the standard reference databases. All 10 commercial samples tabularized in Table S4 were used in DNA resolution and yielded a PCR product of 740 bp for *rbcL*, 850 bp for *matK*, and 500 bp for *ITS2* respectively (Figure S6). All the barcodes generated using *rbcL* and *matK* were of high quality and were used for analysis effectively. Out of 10 commercial samples, the chromatogram of seven samples recovered from the *rbcL* gene was clean without multiple overlapping peaks and showed only one type of barcode (*Camellia sinensis*) in the BLAST analysis. Thus the tested seven samples confirmed that the tea dust was indeed what the label claimed it to be. Three samples (CS4, CS5, and CS7) of tea dust showed anomalous peaks in the chromatogram analyzed. The visible and additional to existing peaks in the pherogram indicated the presence of more than one unique barcode in each sample (Figure S7). These samples were labeled as mixed samples consisting of other plants from more than one species. The admixture samples were detected using TA-cloning and sequenced again for segregating the anomaly in other species identification. The cloned samples showed many other barcodes than the core plant *Camellia sinensis*. CS4 commercial sample generated barcode that was 99.60% identical to *Anacardium occidentale* which is a known adulterant in tea. It is considered a contaminant with species other than the labeled ingredient. From the CS5 commercial sample, three barcodes that showed 100% similarity to *Withania somnifera* (Ashwagandha), *Zingiber officinale* (Ginger), *Ocimum basilicum* (Tulasi) are identified apart from *Camellia sinensis*. CS7 green tea sample showed a sequence similar to *Citrus spp* (Figure S8). Overall, two samples were found to be mixed with flavoring additives and one was contaminated with plant-based adulterant. An investigation by Amane and Ananthanarayan (2019) stated the *rbcL* marker detected the macro contaminants or the plant-based admixtures of other species and also Lagiotis et al. (2020) concluded that the *rbcL* gene is effective for differentiating non-*Camellia* products and their adulterants in the findings. Supporting the hypothesis authors confirmed that the *rbcL* is the suitable locus to identify the genomic templates in mixed samples. But DNA barcoding after the cloning method is preferred for accurate identification of the templates as the barcoding technique may only detect the entire
substitution of one genus with another. The observation correlated with Silva et al. (2020) reported the cloning to be an effective method in identifying mixed specimens when the barcoding technique fails. The observation made in the matK sequences was that they have less variability among tea accessions and no multiple peaks were observed for the anomalous or admixture specimens as seen in electropherograms of the rbcl locus. This confirms the low universality of the primer but the marker segregated the tea samples which are associated with TRI2043 origin. Sequences retrieved from four tea dust (CS1, CS2, CS5, and CS10) showed 9 bp indel similar to TRI2043 tea cultivar. In the specimens sampled, the matK marker had the potential to identify the presence of Camellia sinensis and has nucleotide variations if the sample has more than two contaminants other than authentic tissue. In the matK genome phylogeny, all the authentic samples were branched within the standard tea cultivars except CS5 which is an admixture sample containing more than three DNA templates (Figure S9). The ITS2 was not an ideal DNA barcode for commercial tea products in this study because of its incomplete sequences. The results are in line with De Castro et al. (2017) who reported the inefficacy of the ITS2 marker in sequencing the tea bags. Commercial samples sequenced from ITS2 had potential multiple peaks and noise in the raw chromatograms. Indeed, BLAST comparison showed similarities to Camellia sinensis but we couldn’t retrieve the full ITS2 sequence length from the ITS2 database for analysis. Thus ITS2 has been neglected in analyses of the tea samples in the investigation. Accessions generated from commercial tea products were submitted in the GenBank database (Table S5).

3. Conclusion

Authentication of tea has become unavoidable because of high adulteration risks. The present study standardized an algorithm using plant chloroplast loci and a nuclear non-coding gene that served as reference tags for authentication of commercial products retrieved from rbcl, matK, and ITS2 barcodes. Cloning after DNA barcoding method was preferred to detect admixture genomic templates in a single specimen. The investigation assisted not only in detecting the mixed tea dust but also emphasized several genotypic signatures in the south Indian tea plastid and nuclear regions. Although authenticating pulverized botanicals is a tedious process due to deplorable conduct such as contaminant, filler usage, and mislabeling our DNA barcoding study was successful in discriminating the possible adulterants and flavoring additives used in the black and green teas. Hence, exploring this methodology will help in gaining consumer confidence in authenticating labeled contents in food products.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The authors thank the National Tea Research Foundation (NTRF), Government of India, (S.O: NTRF: 190/2016) for the financial support.
References

Amane D, Ananthanarayan L. 2019. Detection of adulteration in black gram-based food products using DNA barcoding. Food Control. 104:193–200. cont. 2019.04.041.

De Castro O, Comparone M, Di Maio A, Del Guacchio E, Menale B, Troisi J, Aliberti F, Trifuoggi M, Guida M. 2017. What is in your cup of tea? DNA Verity Test to characterize black and green commercial teas. PLoS One. 12(5):e0178262.

Faller AC, Ragupathy S, Shanmughanandhan D, Zhang Y, Lu Z, Chang P, Swanson G, Newmaster SG. 2019. DNA quality and quantity analysis of Camellia sinensis through processing from fresh leaves to a green tea. J AOAC Int. 102(6):1798–1807.

Kuo PC, Li YC, Wu RH, Tzen JT. 2021. Characterization of teaghrelin-like compounds from tea cultivars. Nat Prod Res. 35(1):57–62.

Lagiotis G, Stavridou E, Bosmali I, Osathanunkul M, Haider N, Madesis P. 2020. Detection and quantification of cashew in commercial tea products using High Resolution Melting (HRM) analysis. J Food Sci. 85(6):1629–1633.

Lee SC, Wang CH, Yen CE, Chang C. 2017. 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. 25(2):260–274.

Li YC, Tanapichatsakul C, Pripdeevech P, Hwang TL, Kuo PC, Tzen JT. 2020. Characterisation of teaghrelinlike principles from Assam tea cultivated in Thailand. Nat Prod Res. 19:1–7.

Pazi A, Crawford AJ. 2020. Molecular-based rapid inventories of sympatric diversity: A comparison of DNA barcode clustering methods applied to geography-based vs clade-based sampling of amphibians. J Biosci. 37(5):878–896.

Saravanan KM, John M, Raj Kumar R, Pius PK, Sasikumar R. 2005. Genetic diversity of UPASI tea clones (Camellia sinensis (L.) O. Kuntze) on the basis of total catechins and their fractions. Phytochemistry. 66(5):561–565.

Sharma R, Rana A, Kumar S. 2020. Phytochemical investigation and bioactivity studies of flowers obtained from different cultivars of Camellia sinensis plant. Nat. Prod. Res. 12:1–5.

Silva AJ, Kawalek M, Williams Hill DM, Hellberg RS. 2020. PCR cloning combined with DNA barcoding enables partial identification of fish species in a mixed-species product. Front Ecol Evol. 8:88–28.

Stavrianid A, Stekolshchikova E, Rodin I, Godovikov I, Shpigun O. 2018. Structure elucidation of sweet-tasting cycloartane-type saponins from ginseng oolong tea and Abrus precatorius L. leaves. Nat Prod Res. 32(20):2490–2493. 1416383

Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP. 2011. Commercial teas highlight plant DNA barcode identification successes and obstacles. Sci Rep. 1(1):42–47.

Thuvarakia R, Kottawa-Arachchib JD, Sumanasinghe VA. 2017. Characterization of new tea (Camellia sinensis L.) hybrid progenies based on morphological traits. Int J Tea Sci. 13:1–9.