ORIGINAL ARTICLE

DNA barcode identification of cultivated and wild tropical fruit species

Lee Hong Tnah[1] · Soon Leong Lee[1] · Chai Ting Lee[1] · Kevin Kit Siong Ng[1] · Chin Hong Ng[1] · Ngah Zawiah[1]

Abstract

With the rapid growth of the fruit industry worldwide, it is important to assess adulteration to ensure the authenticity and the safety of fruit products. The DNA barcoding approach offers a quick and accurate way of identifying and authenticating species. In this study, we developed reference DNA barcodes (*rbc*L, ITS2, and *trn*H-*psb*A) for 70 cultivated and wild tropical fruit species, representing 43 genera and 26 families. In terms of species recoverability, *rbc*L has a greater recoverability (100%) than ITS2 (95.7%) and *trn*H-*psb*A (88.6%). We evaluated the performance of these barcodes in species discrimination using similarity BLAST, phylogenetic tree, and barcoding gap analyses. The efficiency of *rbc*L, ITS2, and *trn*H-*psb*A in discriminating species was 80%, 100%, and 93.6%, respectively. We employed a multigene-tiered approach for species identification, with the *rbc*L region used for primary differentiation and ITS2 or *trn*H-*psb*A used for secondary differentiation. The two-locus barcodes *rbc*L+ITS2 and *rbc*L+*trn*H-*psb*A demonstrated robustness, achieving species discrimination rates of 100% and 94.3% respectively. Beyond the conventional species identification method based on plant morphology, the developed reference barcodes will aid the fruit agroindustry and trade, by making fruit-based product authentication possible.

Keywords DNA barcoding · Chloroplast DNA · Species identification · *rbc*L · ITS2 · *trn*H-*psb*A

Introduction

Tropical fruits, a botanically diverse group native to tropical climates (Underhill 2003), historically found their place in small-scale farming or were gathered from the wild for personal consumption. However, recent years have witnessed a significant transformation, turning tropical fruit production into a large and rapidly expanding industry. Mango, mangosteen, guava, pineapple, avocado, and papaya are among the major tropical fruits traded in the international markets (FAO 2022). Advancements in transportation systems and refrigerated storage have played a pivotal role in driving the industry's growth, enabling efficient global consumption and widespread exportation (Underhill 2003). In 2021, Asia, Africa, Central and South America, the Caribbean, and Oceania emerged as the primary exporters of tropical fruits, collectively contributing to the world export of 8379 thousand tons (FAO 2022). Notably, significant quantities of tropical fruits were annually imported by major players, such as the United States of America, the European Union, and China, with these top importers acquiring more than 5369 thousand tons (FAO 2022). While a substantial portion of tropical fruits is consumed fresh, the establishment of processing industries has added another dimension to the industry (Underhill 2003). To date, the tropical fruit industry is supporting the socioeconomic well-being of approximately 160 million people in the tropics (Zhu et al. 2022).

Malaysia is characterized by its rich flora diversity, known as one of the world's principal resources for tropical fruit trees (Saw et al. 1991), with a total of 520 tropical fruit species that produce edible fruits or seeds reported from its diverse ecosystems (Milow et al. 2014). In general, tropical fruit species can be categorized as cultivated and wild. Among the 64 widely cultivated tropical fruits in the country, 21 major fruit species are cultivated on a commercial scale, including mango, mangosteen, guava, rambutan, banana, pineapple, durian, and jackfruit (Rozhan 2019). In terms of wild tropical fruits, 76 species that produce wild edible fruits were recorded based on an inventory from the 50-ha research plot of Pasoh Forest Reserve

✉ Lee Hong Tnah
leehong@frim.gov.my

1 Forest Research Institute Malaysia, 52109 Kepong, Selangor Darul Ehsan, Malaysia

Springer

in Malaysia (Saw et al. 1991). The wild mangoes, *Mangifera* (Anacardiaceae), mangosteen, *Garcinia* (Clusiaceae), rambutans, *Nephelium* (Sapindaceae), and the forest breadfruits, *Artocarpus* (Moraceae) are among the main families that make up more than half the fruit tree flora (Saw et al. 1991). Besides being important contributors to human living and livelihoods, most wild edible fruits also serve as an important food source for animals, as reported in a recent compilation of 160 species of tropical fruit species in Malaysia (Zawiah et al. 2019). Furthermore, tropical fruits are a significant contributor to the country's agricultural gross domestic product. Annually, the industry generates over US$250 million in export profits and provides employment for more than 260,000 Malaysians (Rozhan 2019).

In the worldwide food market, tropical fruit products are widely available to consumers. Fruit juice is among the most steadily growing industries across developing and developed countries (Dasenaki and Thomaidis 2019). Dietary supplements derived from fruits have also gained substantial popularity, positioned as premium food items renowned for their remarkable health benefits, ranging from obesity treatment and cancer prevention to anti-aging and combating various chronic diseases (Hayamizu et al. 2003; Watanapokasin et al. 2010; Kong et al. 2021). In view of the growing market, ensuring the authenticity and the safety of fruit products become imperative, necessitating robust regulatory systems to detect and prevent adulteration and fraud (Jandrić et al. 2017). Notably, significant strides have been taken in leveraging advanced technologies and analytical methodologies, including chromatography, spectroscopy, and DNA-based techniques, to assess the authenticity of fruit products, signifying a crucial advancement in ensuring the authenticity of these products in the market (Dasenaki and Thomaidis 2019).

Since its inception in 2003, DNA barcoding has been a focal point of extensive research for species identification and authentication. This innovative approach relies on a concise standard genomic region, universally present in target lineages, possessing sufficient sequence variation to effectively discriminate among species (Hebert et al. 2003). The Plant Working Group of the Consortium for the Barcode of Life (CBOL) took a significant step by proposing two coding regions from the plastid genome, *rbc*L and *mat*K as a "core barcode" for land plants, to be supplemented with additional regions, such as *trn*H–*psb*A*, atp*F–*atp*H*, psb*K–*psb*I*, trn*L intron, and ITS (CBOL Plant Working Group 2009). Beyond the candidates, other portion of the genome, such as the second internal transcribed spacer (ITS2) from nuclear ribosomal DNA, is highly recommended and has been used in plant barcoding studies (Chen et al. 2010; Duan et al. 2019, Jones et al. 2021). Over the past decades, numerous research teams have piloted large datasets to evaluate the universality, sequence quality, and discriminatory power of the candidate barcode regions (Kress and Erickson 2007; Hollingsworth et al. 2009; China Plant BOL Group et al. 2011). Recognizing differences in their efficiency, a multi-gene-tiered approach is widely adopted, whereby a first-tier coding region common across the land plants provides resolution at a certain rank (e.g., family or genus), and a more variable (coding or non-coding) region provides resolution up to the species level (Newmaster et al. 2006). This tiered strategy ensures a comprehensive and accurate approach to DNA barcoding, catering to the diverse resolution requirements across different levels of taxonomy.

Despite the importance of tropical fruits in global food trade and the assessment of fruit product authenticity, the availability of DNA barcodes for tropical fruits remains notably limited. Presently, only a handful of DNA barcodes for fruit trees have been documented, primarily within genera, such as *Ananas* (Hidayat et al. 2012), *Annona* (Larranaga and Hormaza 2015), *Garcinia* (Anerao et al. 2021), *Selenicereus* (Huy et al. 2021), and *Mangifera* (Suparman et al. 2013; Fitmawati et al. 2016; Kumar et al. 2022). Recognizing this gap, this project was initiated with the purpose of facilitating species identification and fruit product authenticity assessment. We present a simple and accessible approach to barcode the cultivated and wild tropical edible fruits and evaluate the efficiency of species discrimination.

Materials and methods

Sample collection and DNA extraction

All fruit reference samples were obtained from the living collections of the Kepong Botanical Garden (N03.22440 E101.61578), Fruit Arboretum (N03.23043 E101.63397) of the Forest Research Institute Malaysia (FRIM), and the 50-ha forest research plot in Pasoh Forest Reserve (N02.98255, E102.31050). A total of 264 samples were collected, representing 70 fruit tree species, 43 genera, and 26 families. Details of the collected reference samples are listed in Table S1. Leaf samples were collected from two to 14 individuals per species. At least two samples per species were included to increase reliability and enable intraspecific variation detection. Total DNA was extracted according to Murray and Thompson (1980) with a single modification (2×CTAB) and further purified using a High Pure PCR Template Preparation Kit (Roche Diagnostics GmbH, Penzberg, Germany). For each sample, a properly pressed, labeled, and mounted voucher specimen was made for future reference.

PCR amplification and sequencing of DNA barcoding markers

In this study, we employed refined DNA barcodes, *rbc*L and *trn*H-*psb*A used in tropical plants (Tnah et al. 2019), and supplemented with ITS2, which has proved remarkably effective at discriminating diverse plant groups (Chen et al. 2010). The universal primer pairs of *rbc*L, ITS2, and *trn*H-*psb*A used for amplification and sequencing are listed in Table S2. The three candidate DNA barcodes were amplified separately through a 10 µl PCR reaction, consisting of Q5 High-Fidelity 2X Master Mix (New England Biolabs, Ipswich, MA, USA), 0.5 µM of each primer, and approximately 10 ng of template DNA. The PCR reactions were performed in a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA), for an initial denaturing step of 30 s at 98 °C, 35 cycles of 98 °C for 10 s, 50 °C annealing temperature for 30 s, and 72 °C for 30 s, followed by a final extension of 72 °C for 2 min. The PCR products were subsequently purified with ExoSAP-IT Express (Affymetrix, Cleveland, OH, USA). The cycle sequence reactions of the purified products were bi-directionally using the Big-Dye Terminator Sequencing Kit (Applied Biosystems) based on the standard dideoxy-mediated chain termination method. The sequencing thermal profile consists of 95 °C for 1 min, followed by 50 cycles of 95 °C for 10 s, 50 °C for 5 s, and 60 °C for 4 min. The sequenced products were then purified using BigDye XTerminator Purification Kit (Applied Biosystems) prior to electrophoresis on an ABI 3130xl Genetic Analyzer (Applied Biosystems).

Analysis of species recoverability and species discrimination

Nucleotide sequences were edited and assembled using SEQUENCHER version 5.4.6 (Gene Codes Corporation, Ann Arbor, MI, USA). Heterozygous bases found in the chromatograms were manually assessed and assigned IUPAC (International Union of Pure Applied Chemistry) codes. All DNA sequences obtained in this study have been deposited in GenBank (Table S1). Species recoverability (the ability to successfully retrieve a sequence from a species) of *rbc*L, ITS2, and *trn*H-*psb*A was compared among 264 samples of 70 fruit species. In order to ensure the accuracy of sequence identification, the edited barcodes were cross-checked with the nucleotide sequences deposited in the GenBank database. The performance of barcode markers in species discrimination was evaluated using similarity-based (Basic Local Alignment Search Tool, BLAST), nearest genetic distance (Neighbor-joining, NJ), and barcoding gap methods. For each individual data set, the discrimination ability for each barcode region was assessed as a percentage of the total number of species correctly identified divided by the total number of species examined.

In the similarity BLAST analysis, all sequences obtained from the study were formatted as both database and query, and subsequently queried individually against the database. When the query has a 100% match to the conspecific sequences, species discrimination is considered successful. On the contrary, species discrimination is considered a failure when it matches numerous species. As an alternative to sequence similarity-based identification, the NJ tree (Saitou and Nei 1987) was constructed using the program MEGA 11 (Tamura et al. 2021) based on sequence divergence estimated with Kimura 2-parameter (Kimura 1980). The multiple sequence alignments of ITS2 and *trn*H-*psb*A were performed using MAFFT (Multiple Alignment using Fast Fourier Transform; Katoh and Standley 2013). Bootstrapping analysis was conducted based on 1000 replicates to evaluate the robustness of the node for each tree. The number of species with conspecific individuals successfully clustered under respective monophyletic groups in the phylogenetic tree was calculated. In the barcoding gap assessment, pairwise Kimura 2-parameter distances were calculated within and between species using MEGA 11. For species with two or more individuals, the barcoding gap was assessed by plotting the maximum intraspecific distance against the minimum interspecific distance. The percentage of species having a barcoding gap was calculated based on those with the minimum interspecific distance greater than the maximum intraspecific distance.

Results and discussion

Reference DNA barcodes of tropical fruits

Despite the desirability of using DNA barcoding broadly in plant identification, the lack of a reliable DNA barcoding reference library is often the main impediment to its application. Therefore, the initial and crucial step involves the development of trustworthy reference barcodes. Securing access to accurately identified specimens from arboretums, botanical gardens, and forest research plots is indispensable for this purpose. In this study, fruit tree species spanning various families—Anacardiaceae, Fabaceae, Sapindaceae, Meliaceae, Moraceae, Myrtaceae, Phyllanthaceae, and others—are integral components of the living collections at FRIM and Pasoh Forest Reserve. Within Malaysia's tropical fruit diversity, approximately 160 species with major or medium economic status contribute to both self-consumption and market demands (Milow et al. 2014). The ongoing effort aims to expand the DNA barcode reference library to encompass a more extensive array of local fruit species in future.

We successfully generated a total of 264 *rbc*L barcodes (70 species), 253 ITS2 barcodes (67 species), and 223 *trn*H-*psb*A barcodes (62 species) (Fig. 1). The PCR products for *rbc*L had a consistent length of 605 bp across all species (GenBank accession numbers OK052728-OK052798). The length of the ITS2 sequence was 221–495 bp (GenBank accession numbers OK052867-OK052949). For ITS2, high-quality DNA barcodes were obtained; nevertheless, single-nucleotide polymorphisms (SNPs) and homo-polymers were observed in some species (Fig. 2). The intra-genomic variability of ITS2 resulted in a number of SNPs in the region ranging from 1 to 11, which were recoded with IUPAC codes. Partial readable sequences were observed in five species, due to homo-polymers composed of mononucleotide repeats, which caused premature termination of sequencing and resulted in unidirectional reads. The alignment length of *trn*H-*psb*A ranged from 215 to 664 bp (GenBank accession numbers OK052799-OK052866). The region was not amplified well in certain genera (e.g., *Aglaia, Artocarpus,*
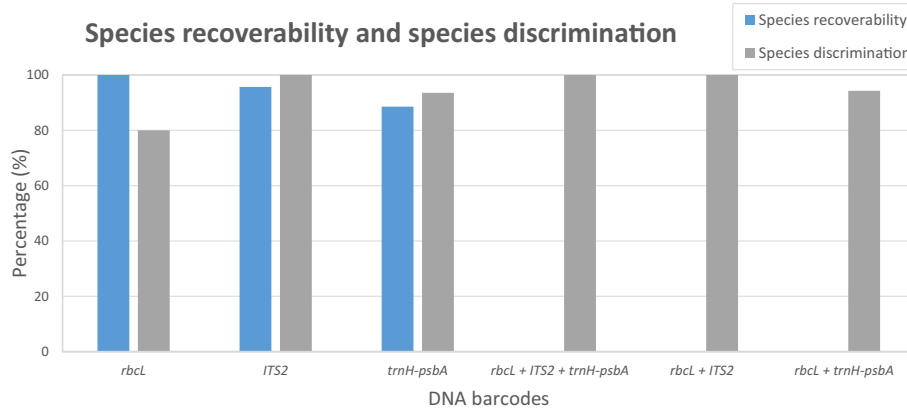

Fig. 1 Evaluation of species recoverability and species discrimination is depicted, comparing the efficacy of single barcodes, (*rbc*L, ITS2, *trn*H-*psb*A) and multigene-tiered barcodes, (*rbc*L+ITS2+*trn*H-*psb*A, *rbc*L+ITS2 and *rbc*L+*trn*H-*psb*A) through similarity BLAST. Species recoverability is represented by the percentage of successfully sequenced species within each barcode, while species discrimination assesses the capability of the barcodes to achieve identification at the species level

Fig. 2 The study reveals various types of polymorphisms in ITS2 regions. Chromatograms illustrate examples of (**A**) single-nucleotide polymorphism (SNP) and (**B**) homo-polymer regions found in different species, as indicated in the corresponding box
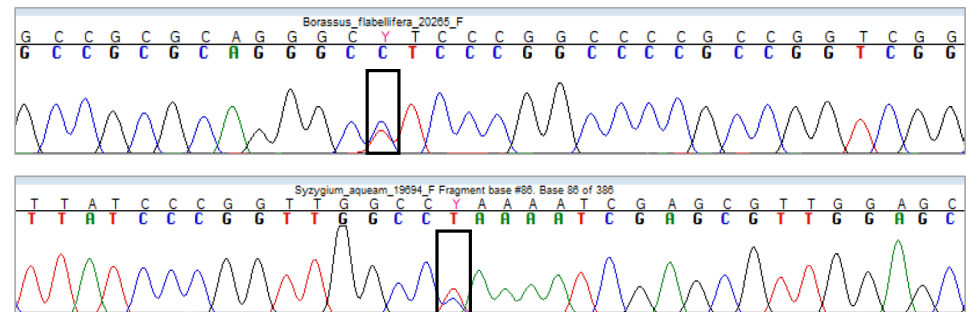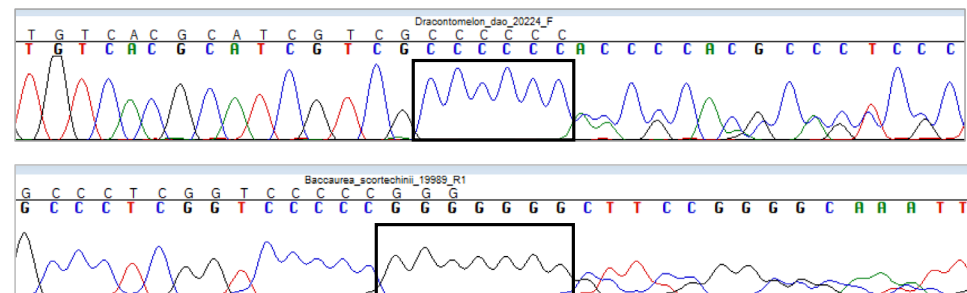
Borassus, Dracontomelon, Durio, and *Lansium*) and had homo-polymers that were difficult to sequence accurately.

Based on the results, out of the 70 species, *rbc*L has a higher species recoverability (100%) than ITS2 (95.7%) and *trn*H-*psb*A (88.6%) (Fig. 1). All three regions were recovered for 60 species, except for a small number of species in the ITS2 and *trn*H-*psb*A. In the cases reported on *rbc*L (de Vere et al. 2012) and *trn*H-*psb*A (Sass et al. 2007), empirical studies revealed that the addition of multiple primer combinations did not improve the overall recoverability. Moreover, as studies have shown that intra-genomic variability of ITS2 is common in plants (Song et al. 2012), further primer combinations might increase the risk of amplifying different copies within this multi-copy marker (Jones et al. 2021).

Evaluation of species discrimination (BLAST, phylogenetic tree, and barcoding gap)

In the present study, the *rbc*L emerged as the most effective region with good recovery and high sequence quality. Serving as the primary region in our multigene-tiered approach, it establishes a valuable backbone for the barcode dataset. In GenBank, the region provides a good baseline for comparison with other plant species, as it is one of the most characterized plastid coding regions with wide representation from all major groups. The evaluation of similarity BLAST (Fig. 1) and Neighbor-Joining trees (Figs. 3, 4, 5) revealed the discriminatory efficiency of *rbc*L, ITS2, and *trn*H-*psb*A as 80%, 100%, and 93.6%, respectively. Notably, *rbc*L demonstrates modest resolution at the species level, successfully discriminating 80% of the species, with the exception of 14 species from four genera (*Baccaurea,*

Fig. 3 Unrooted Neighbor-joining tree of 70 fruit species based on *rbc*L sequences, supported by 1000 bootstrap replicates. The circle denotes 14 unresolved species. Each species name is accompanied by the corresponding DNA ID number of the sample

Fig. 4 Unrooted Neighbor-joining tree of 67 fruit species based on ITS2 sequences, supported by 1000 bootstrap replicates. Species with more than one individual exhibited the presence of intraspecific vari- ations. Each species name is accompanied by the corresponding DNA ID number of the sample

Garcinia, Mangifera, and *Nephelium*) exhibiting low interspecific variation (Fig. 3). This observation aligns with the slow evolutionary rate of the *rbc*L region, making it the least divergent plastid gene in flowering plants (Kress et al. 2005). Consistent with other studies, such as those conducted by the US National Herbarium (Kress and Erickson 2007) and the UK flora (Jones et al. 2021), *rbc*L exhibits a modest discrimination ability, recorded at 76.3% and 64%, respectively.

In the assessment of the barcoding gap, the correlation between minimum interspecific and maximum intraspecific distances demonstrated high discrimination ability of the three regions (Fig. 6, Table S3). The study unveils that 80%, 97%, and 93.4% of species exhibit barcoding gaps in *rbc*L, ITS2, and *trn*H-*psb*A, respectively (Table S3)—a crucial criterion for establishing a reliable barcode. This alignment with the fundamental DNA barcoding principle, advocates for interspecific divergence to surpass intraspecific variation (Hebert et al. 2003). Calculated mean interspecific distances based on *rbc*L, ITS2, and *trn*H-*psb*A were 0.0143, 0.0958 and 0.1092, respectively. Despite their lower recovery rates, ITS2 and *trn*H-*psb*A barcodes exhibited higher species discrimination, outpacing *rbc*L in the interspecific genetic distance. Remarkably, both ITS2 and *trn*H-*psb*A are by far the best barcodes with the highest ability of species discrimination, and have been shown in numerous studies to be effective in the construction of DNA barcoding resources (Kress and Erikson 2007; Chen et al. 2010; Duan et al. 2019).

Fig. 5 Unrooted Neighbor-joining tree of 62 fruit species based on *trn*H-*psb*A sequences, supported by 775 bootstrap replicates. Species with more than one individual exhibited the presence of intraspecific variations. Each species name is accompanied by the corresponding DNA ID number of the sample

Finding an effective and robust species-level identification barcode across land plants remains a pressing challenge. Considering the ease of recovery, *rbc*L is the preferred option but it is hampered by lower species discrimination. If species discrimination greater than 80% is required, then there is a need to adopt a multigene-tiered approach, with the second or third tier being more taxon-specific. In our study, the implementation of three-locus plant barcodes (*rbc*L + ITS2 + *trn*H-*psb*A) yielded marginal or no improvement over two-locus plant barcodes (*rbc*L + ITS2 or *rbc*L + *trn*H-*psb*A) in terms of discriminated species (Fig. 1). This finding aligns with similar observations by Kress et al. (2010), who achieved 93% species discrimination using two-locus (*rbc*L + *trn*H-*psb*A) and three-locus (*rbc*L + *mat*K + *trn*H-*psb*A) barcodes in a Puerto Rican

Forest plot. In terms of cost efficiency, we opted for a two-locus approach rather than a three-locus. Besides, the two-locus plant barcodes, e.g., *rbc*L + *trn*H-*psb*A (Kress and Erickson 2007; Ferri et al. 2015; Tnah et al. 2019), *rbc*L + *trn*L-F (de Groot et al. 2011), ITS2 + *trn*H-*psb*A (Duan et al. 2019), among others, have been extensively used to establish reference libraries for various important plants.

In the context of two-locus plant barcodes, the universal core locus of the *rbc*L region serves as a foundation, while ITS2 and *trn*H-*psb*A are interchangeably employed as a second tier region. The independent application of the second tier alignment for a limited number of unresolved taxa mitigates difficulties in aligning highly divergent genera. In our study, both ITS2 and *trn*H-*psb*A regions exhibited

Fig. 6 The plot of minimum interspecific vs. maximum intraspecific ▶ distances in the barcoding dataset of (**a**) *rbc*L, (**b**) ITS2, and (**c**) *trn*H-*psb*A. Pairwise comparisons are calculated from the multiple alignment of each barcode. The diagonal indicates points of equality, with circles above the 1:1 line signifying the presence of a barcode gap

a sufficient number of autapomorphies, elevating their discrimination abilities to 100% and 94.3%, respectively (Fig. 1). Despite the fact that ITS2 had a higher species recovery rate, *trn*H-*psb*A was included as a complimentary second tier region. This study revealed instances where certain species were non-amplifiable in ITS2 but were amplified perfectly using *trn*H-*psb*A (e.g., *Myristica fragrans* & *Litchi chinensis*). Likewise, with the use of *trn*H-*psb*A alone, species recoverability in the genera *Aglaia, Artocarpus, Borassus, Dracontomelon, Durio,* and *Lansium* was limited. Recognizing the variation in recoverability and the limitation of individual DNA barcodes, the flexibility in choosing a second tier locus offers a substantial advantage. The approach has the advantage of keeping the complementary second tier regions at two for any particular group of taxa, thus maintaining the current efficiency of the system (Newmaster et al. 2006).

Applications of the reference DNA barcodes

While facilitating species identification, reference DNA barcodes play a pivotal role as indispensable tools for taxonomic exploration and phylogenetic analysis. Their utility becomes particularly evident when distinguishing closely related species that share highly similar morphological characteristics. The absence of diagnostic flowers and fruits during the juvenile stage presents challenges in making rapid and unequivocal identification, thereby impeding studies in diversity evaluation, germplasm collection, conservation, and management (Larranaga and Hormaza 2015). In plant genera where diagnostic features are lacking and morphological plasticity complicates the assessment of phylogenetic relationships, as seen in the *Mangifera* genus (Fitmawati et al. 2016), DNA barcodes provide a crucial solution. Discrepancies between morphological characteristics (Kostermans and Bompard 1993) and molecular studies (*rbc*L; Suparman et al. 2013, ITS; Fitmawati et al. 2016) highlight the necessity for taxonomic revision. In our current study, the cluster analysis of three *Mangifera* species based on ITS2 (Fig. 4) supports the previous molecular classification (Fitmawati et al. 2016). The acquisition of molecular data through DNA barcodes is thus imperative for supporting and supplementing morphology-based taxonomic classifications in such instances.

Considering the substantial global consumption of fruit juices and products, the detection and the prevention of adulteration and fraud are important and should
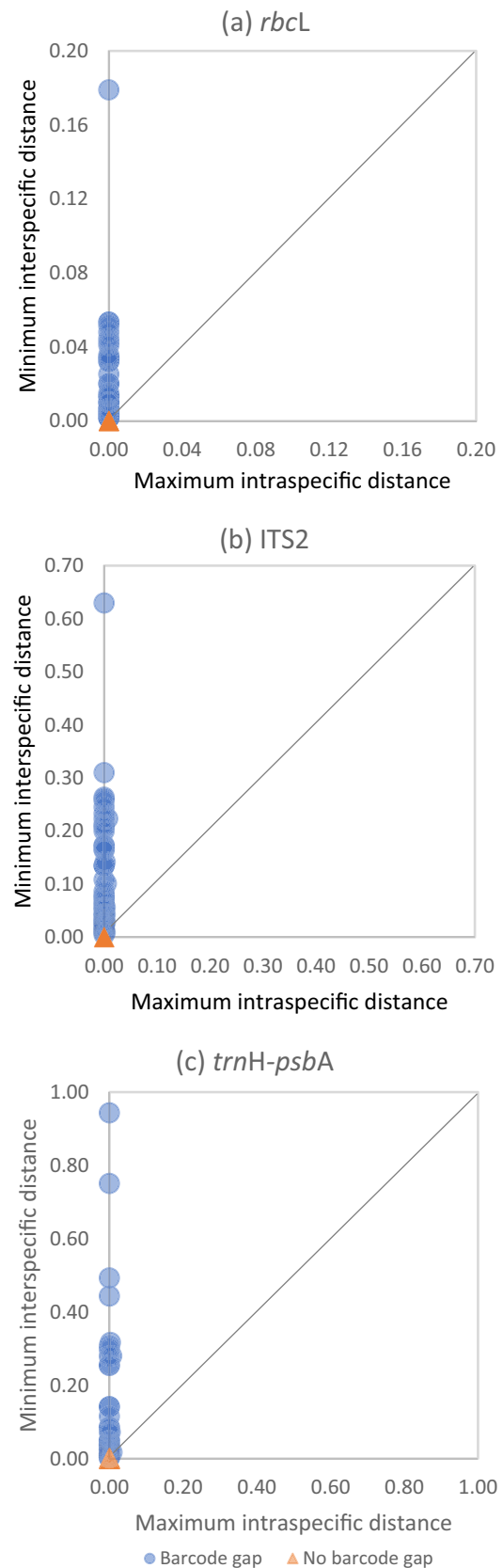
be given attention. It is crucial to establish robust control systems to safeguard consumers against impure and deceitfully presented fruit products (Dasenaki and Thomaidis 2019). When a query sample has been processed or is in powder form, taxonomic identification is impossible, let alone verification of product authenticity, unless there are unique biochemical markers for the particular species. DNA barcoding offers a solution to overcome this limitation. The technique has proven effective in identifying fruit residues in various products, including juices (Faria et al. 2013; Wu et al. 2018), pectin (Barcaccia et al. 2015), plant oil (Spaniolas et al. 2008), yogurt (Ortola-Vidal et al. 2007), purees, chocolates, and cookies (Sakai et al. 2010). Moreover, in an increasingly health-conscious population, tropical fruits are seen as an appropriate source of nutrition, added variety, and exotic appeal (Underhill 2003). As such, the DNA barcodes are useful for authentication assessment and plant-based adulterant detection, pertaining to fruit-based nutritional products.

Conclusions

Given the importance of tropical fruits in socioeconomic development, extensive research to catalog fruit genetic resources is needed to strengthen the industry. The establishment of reference DNA barcodes for 70 tropical fruit tree species in this study allows us to use DNA barcoding techniques for species identification and authenticity assessment of fruit products. We evaluated the effectiveness of the multigene tier approach, with the two-locus barcodes, rbcL + ITS2 and rbcL + trnH-psbA demonstrating excellent resolution in species discrimination. In concert with conventional species identification methods rooted in plant morphology, the established reference barcodes hold immense potential for the agroindustry and food trade. Their utilization will not only enhance the accuracy of species identification but also fortify fruit-based product authentication in the industries.

Declarations

References

Anerao J, Jha V, Shaikh N, Shivalkar A, Nityanand A, Sawant D, Rao GR, Mangaonkar K, Deodhar M, Desai N (2021) DNA barcoding of important fruit tree species of agronomic interest in the genus Garcinia L. from the Western Ghats. Genet Resour Crop Evol 68(8):3161–3177

Barcaccia G, Lucchin M, Cassandro M (2015) DNA barcoding as a molecular tool to track down mislabeling and food piracy. Diversity 8(1):2

CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106:12794–12797

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE 5:e8613

China Plant BOL Group, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci USA 49:19641–19646

Dasenaki ME, Thomaidis NS (2019) Quality and authenticity control of fruit juices - a review. Molecules 24(6):1014

de Groot GA, During HJ, Maas JW, Schneider H, Vogel JC, Erkens RH (2011) Use of rbcL and $trnL$-F as a two-locus DNA barcode for identification of NW-European ferns: an ecological perspective. PLoS ONE 6(1):e16371

de Vere N, Rich TC, Ford CR, Trinder SA, Long C, Moore CW, Satterthwaite D, Davies H, Allainguillaume J, Ronca S, Tatarinova T (2012) DNA barcoding the native flowering plants and conifers of Wales. PLoS ONE 7(6):e37945

Duan H, Wang W, Zeng Y, Guo M, Zhou Y (2019) The screening and identification of DNA barcode sequences for Rehmannia. Sci Rep 9(1):1–12

FAO (2022) Major tropical fruits statistical compendium 2021. Rome

Faria MA, Magalhães A, Nunes ME, Oliveira MB (2013) High resolution melting of $trnL$ amplicons in fruit juices authentication. Food Control 33(1):136–141

Fay MF, Swensen SM, Chase MW (1997) Taxonomic affinities of Medusagyne oppositifolia (Medusagynaceae). Kew Bull 52:111–120

Ferri G, Corradini B, Ferrari F, Santunione AL, Palazzoli F, Alu M (2015) Forensic botany II, DNA barcode for land plants: which markers after the international agreement? Forensic Sci Int: Genet 15:131–136

Fitmawati F, Hayati I, Sofiyanti N (2016) Using ITS as a molecular marker for Mangifera species identification in Central Sumatra. Biodiversitas 17(2):653–656

Hayamizu K, Ishii Y, Kaneko I, Shen M, Okuhara Y, Shigematsu N, Tomi H, Furuse M, Yoshino G, Shimasaki H (2003) Effects of Garcinia cambogia (Hydroxycitric Acid) on visceral fat accumulation: a double-blind, randomized, placebo-controlled trial. Curr Ther Res Clin Exp 64(8):551–567

Hebert PD, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proc R Soc Lond B Biol Sci 270:313–321

Hidayat T, Abdullah FI, Kuppusamy C, Samad AA, Wagiran A (2012) Molecular identification of Malaysian pineapple cultivar based on internal transcribed spacer region. APCBEE Proc 4:146–151

Hollingsworth ML, Andra Clark AL, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan R, Chase MW, Gaudeul M, Hollingsworth PM (2009) Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Resour 9(2):439–457

Huy TG, Men TT, Thi NPA, Khang DT (2021) Identification of dragon fruit (*Selenicereus*) species in Mekong Delta based on DNA barcode sequences. Biodiversitas 22(10):4216–4222

Jandrić Z, Islam M, Singh DK, Cannavan A (2017) Authentication of Indian citrus fruit/fruit juices by untargeted and targeted metabolomics. Food Control 72:181–188

Jones L, Twyford AD, Ford CR, Rich TC, Davies H, Forrest LL, Hart ML, McHaffie H, Brown MR, Hollingsworth PM, De Vere N (2021) Barcode UK: a complete DNA barcoding resource for the flowering plants and conifers of the United Kingdom. Mol Ecol Resour 21(6):2050–2062

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kong YR, Jong YX, Balakrishnan M, Bok ZK, Weng JKK, Tay KC, Goh BH, Ong YS, Chan KG, Lee LH, Khaw KY (2021) Beneficial role of *Carica papaya* extracts and phytochemicals on oxidative stress and related diseases: a mini review. Biology (basel) 10(4):287

Kostermans AJGH, Bompard JM (1993) The Mangoes: Their Botany, Nomenclature, Horticulture and Utilization. Academic Press, London

Kress WJ, Erickson DL (2007) A two-locus global DNA barcode for land plants: the coding *rbc*L gene complements the non-coding *trn*H–*psb*A spacer region. PLoS One 2:e508

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA 102(23):8369–8374

Kumar S, Kaushik RA, Jain D, Saini VP, Babu SR, Choudhary R, Ercisli S (2022) Genetic diversity among local mango (*Mangifera indica* L.) germplasm using morphological, biochemical and chloroplast DNA barcodes analyses. Mol Biol Rep 49:3491–3501

Larranaga N, Hormaza JI (2015) DNA barcoding of perennial fruit tree species of agronomic interest in the genus *Annona* (Annonaceae). Front Plant Sci 6:589

Levin RA, Wagner WL, Hoch PC, Nepokroeff M, Pires JC, Zimmer EA, Sytsma KJ (2003) Family-level relationships of Onagraceae based on chloroplast *rbc*L and *ndh*F data. Am J Bot 90:107–115

Milow P, Malek SB, Edo J, Ong HC (2014) Malaysian species of plants with edible fruits or seeds and their valuation. Int J Fruit Sci 14(1):1–27

Murray M, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in land plants: evaluation of *rbc*L in a multigene tiered approach. Botany 84(3):335–341

Ortola-Vidal A, Schnerr H, Rojmyr M, Lysholm F, Knight A (2007) Quantitative identification of plant genera in food products using PCR and Pyrosequencing technology. Food Control 18(8):921–927

Rozhan AD (2019) Trends in production, trade, and consumption of tropical fruit in Malaysia. FFTC Agricultural Policy Platform (FFTC-AP). Accessed in 31 May 2021, https://ap.fftc.org.tw/article/1381.

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sakai Y, Ishihata K, Nakano S, Yamada T, Yano T, Uchida K, Nakao Y, Urisu A, Adachi R, Teshima R, Akiyama H (2010) Specific detection of banana residue in processed foods using polymerase chain reaction. J Agric Food Chem 58(14):8145–8151

Sass C, Little DP, Stevenson DW, Specht CD (2007) DNA barcoding in the Cycadales: testing the potential of proposed barcoding markers for species identification of cycads. PLoS One 2(11):e1154

Saw LG, LaFrankie JV, Kochummen KM, Yap SK (1991) Fruit trees in a Malaysian rain forest. Econ Bot 45(1):120–136

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92(1):142–166

Song J, Shi L, Li D, Sun Y, Niu Y, Chen Z, Luo H, Pang X, Sun Z, Liu C, Lv A (2012) Extensive pyrosequencing reveals frequent intra-genomic variations of internal transcribed spacer regions of nuclear ribosomal DNA. PLoS One 7(8):e43971

Spaniolas S, Bazakos C, Ntourou T, Bihmidine S, Georgousakis A, Kalaitzis P (2008) Use of lambda DNA as a marker to assess DNA stability in olive oil during storage. Eur Food Res Technol 227(1):175–179

Suparman PA, Hidayat T (2013) Phylogenetic analysis of *Mangifera* based on *rbc*L sequences, chloroplast DNA. Sci Papers Ser B Hort 57:235–240

Tamura K, Stecher G, Kumar S (2021) MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38(7):3022–3027

Tnah LH, Lee SL, Tan AL, Lee CT, Ng KK, Ng CH, Nurul-Farhanah Z (2019) DNA barcode database of common herbal plants in the tropics: a resource for herbal product authentication. Food Control 95:318–326

Underhill S (2003) Fruits of tropical climates: commercial and dietary importance. In: Caballero B (ed) Encyclopedia of food sciences and nutrition, 2nd edn. Academic Press: Oxford, UK, pp. 2780–2785

Watanapokasin R, Jarinthanan F, Jerusalmi A, Suksamrarn S, Nakamura Y, Sukseree S, Uthaisang-Tanethpongtamb W, Ratananukul P, Sano T (2010) Potential of xanthones from tropical fruit mangosteen as anti-cancer agents: caspase-dependent apoptosis induction in vitro and in mice. Appl Biochem Biotechnol 162(4):1080–1094

Wu Y, Li M, Yang Y, Jiang L, Liu M, Wang B, Wang Y (2018) Authentication of small berry fruit in fruit products by DNA barcoding method. J Food Sci 83(6):1494–1504

Zawiah N, Nik Zanariah NM, Othaman H, Norwati M (2019) Simbiosis flora & fauna. Forest Research Institute Malaysia, Selangor, Malaysia, p 408

Zhu Z, Johnson J, Zaman QU, Wang H (2022) Challenges and opportunities to improve tropical fruits in Hainan. China Trop Plants 1(1):1