Analysis of DNA Barcoding Suitable for Tea Tree Field Genebank

Li Yanyan, Huang Wei, Lin Jiajia, Nie Chuanpeng
Wuyi University, Wuyishan, 354300
Corresponding author Email: niechp@wuyiu.edu.cn

Abstract
With the development of research for DNA barcoding, its application has attracted more and more attention. In this study, 100 tea tree samples were selected as subjects, the partial sequences of chloroplast matK and rbcL genes were used to investigate the molecular barcodes suitable for tea tree. The results showed that the rbcL sequences of 100 tea samples were identical, the matK sequences were different, the genetic distance ranged from 0.000 to 0.032, the sequences could be divided into 14 haplotypes, Hd and Pi were 0.604 and 0.23×10⁻², at the same time, the construction analysis of the sequence is carried out. The results showed that the matK sequence could be used in the development and utilization of DNA barcoding of tea tree field genebank.

Keywords
Tea tree; DNA barcoding; Field genebank

DNA barcoding is a new technology for rapid and accurate identification of species. Since Canadian taxonomist Paul Hebert put forward this concept, the scientific and technological circles have launched the "DNA Barcoding". DNA barcoding has become one of the most rapidly developing frontiers in Biology (Tautz et al., 2003). In recent years, the importance of DNA barcoding research has become increasingly prominent, which is widely used in the monitoring of biological diversity, phylogenetic analysis and also has a wide application prospect in the fields of medicine, food quality control, etc. (Sha et al., 2018).

In the process of establishing candidate DNA barcoding, the potential DNA barcoding regions recommended by botanists mainly concentrated in internal transcribed spacer (ITS) of ribosome gene and the coding or non-coding regions of chloroplast genome (Li et al., 2017). At present, the widely recognized sources of plant DNA barcoding are mainly ITS1, ITS2, rbcL, matK, trnH, psbA, etc. (Wu et al., 2020).

Wuyi University tea tree field genebank was established in 2015. By the end of 2017, the total built area is 5.33 hm². It laid a solid foundation for the breeding, development and utilization of high quality tea resources. More than 300 tea germplasm resources from 15 provinces have been collected and preserved in the resource nursery. The tea varieties have great research value. In this study, we studied the matK and rbcL partial sequences of 100 tea varieties, in order to find the molecular barcoding of appropriate tea variety, and provide certain reference for the later collection of tea resources, identification of tea varieties, preservation in the nursery and cultivation of new varieties.

1 Results and Analysis
1.1 Sequence amplification and determination
The amplified products of PCR were detected by gel electrophoresis, the location of the bands was about 600–700 BP, and the length of the detected sequences was consistent with that predicted. All the amplified sequences were identified as target fragments by BLAST.
1.2 Sequence processing and genetic distance analysis

The sequence of 100 \textit{rbcL} samples is the same. The \textit{matK} sequences of 100 samples were analyzed by clustalX, and have six differences.

\textbf{\textit{rbcL} sequence:}

\begin{verbatim}
TGATGAACCTTTTTAGTTAGGATGATATGGAAGGCGTATTTCAATATATTTGTATATGAAAATCAG
ATTTTTGAGATTAACAAAGATCGTTCTTTTCTGAGTGAAACTAGAAAGTTATTTTTATAGTTATCGGAAT
TCTGGTTATCTGAATAATGAGTATCACTTATGGCTAATCTCCTACTATGATCGTTACATGTATGACTCA
ATATATGTTGGAATAATCACAATATAGTTGCACTGACAGTTATCTTTCTCTAAATCCGTATTGATAC
TTACATTGTAAGTGGTAGTAACAAATACAGTGACAGTTACATTATGATGTTCCGTTTGGTGGAAGATG
GAAATAGTAGTGAAAGCGAGAGTTCCAGTATAATAGCTAGCAGAATGGGAGTGTCTCCTTGGTAAATTTAAG
AGAAATGCTAATGATCTGATGTAACTCAAAAATACAAGCTTTGTTGGTTCAATGCGAAAATTTGG
TTGGATTTATTTGAAATTTTGGTAAATTTTGGTAAATTTTGGTAAATTTTGGTAAATTTTGGTAAATTTT

\textbf{\textit{matK} sequence:}

\end{verbatim}

The genetic distance of 100 tea varieties was calculated by mega6.0. The genetic distance of \textit{rbcL} sequence was zero, and the genetic distance of \textit{matK} sequence was 0.000 ~ 0.032, most of them were between 0% ~ 0.5% (Figure 1).

Among them, the genetic distance of Yingshuang (Z11) and Jiangmuxiang (D7) was the farthest (0.032), indicating that the two tea cultivars had the farthest genetic relationship. Although the difference of \textit{matK} sequence is small, it can be used as an alternative fragment of tea DNA barcode.

1.3 \textit{matK} phylogenetic analysis

Mega6.0 software was used to construct phylogenetic tree based on K2P model with 1000 times of self-development. NJ system tree (Figure 2) mainly divides 100 tea varieties into three branches, of which Wumengzao (S35) and Yingshuang (Z11) are respectively divided into one big branch, and the remaining 98 tea varieties are divided into a
large branch, with a self-development support rate of 63%. There are two small branches in this large branch, and the small branches with 86% self-development support rate are Tianfucha 11 (S33) and Chongpi 71-1 (S1) and Xintianwan dacha (H16), and the other branch is composed of 22 tea varieties, including tea varieties from 5 regions, with a self-development support rate of 67%.

NJ tree based on matK partial sequences can reflect the genetic relationship among species (Zhu et al., 2014; Qi et al., 2019). The results showed that the extended matK fragment of tea plant is relatively conservative, the results of sequence analysis among varieties, especially DNA barcode analysis, may not be ideal, but it can be used as an alternative sequence. In future research, we should lengthen the length of this sequence, strive to find more diversity, and cooperate with other gene sequences, so as to carry out DNA barcode analysis among tea varieties.

Figure 2 NJ phylogenetic tree of tea tree based on matK sequence

1.4 matK haplotype polymorphism analysis and neutral test

Haplotype polymorphism (Hd) and nucleotide polymorphism (Pi) were analyzed by dnasp5.10 software. Through a series of calculations, the haplotype number of these 100 tea varieties was 14 (Table 1), the haplotype polymorphism (Hd) was 0.604, and the nucleotide polymorphism Pi was $0.23 \times 10^{-2}$; the total number of these 100 sequence sites was 818 except deletion, and there were 41 variation sites, and 11 effective sites accounted for 1.3% of the effective length of the sequence. The average level of chloroplast DNA genetic diversity was 0.67. The haplotype polymorphism (Hd) of 100 tea varieties in this study was lower than the average value, indicating that the genetic diversity among tea varieties was limited. The central test showed that Tajima's D value, Fu and Li's D value and Fu and Li's F value test were all negative values, all of which did not reach the significant level, which conformed to the neutral evolution pattern. That is to say, at the species level, the evolution pattern of matK sequence of tea varieties was consistent with the hypothesis of neutral evolution, which proved that the sequences could be used for system analysis. This result is also consistent with the research results of Liu et al. (2018) using rbcL sequence and trnH psbA sequence in the genetic diversity of tea resources, and both conform to the hypothesis of neutral evolution.
Table 1 Source of 100 tea tree and their haplotypes

| No. | Cultivars          | Origins | Haplotype | No. | Cultivars          | Origins | Haplotype |
|-----|--------------------|---------|-----------|-----|--------------------|---------|-----------|
| C1  | Bayutezao          | Chongqing| Hap2      | D6  | Hongdi             | Guangdong| Hap1      |
| C2  | Jinju              | Chongqing| Hap1      | D7  | Jiangtianxiang     | Guangdong| Hap8      |
| C3  | Nanchiang 2        | Chongqing| Hap11     | D8  | Laoxianweng        | Guangdong| Hap1      |
| C4  | Shushui 1          | Chongqing| Hap1      | D9  | Songzhong          | Guangdong| Hap1      |
| C5  | Shushui 2          | Chongqing| Hap2      | D10 | Tanchunxiang       | Guangdong| Hap2      |
| Z1  | Anjibaicha         | Zhejiang | Hap2      | D11 | Wuyedancong        | Guangdong| Hap1      |
| Z2  | Biyun              | Zhejiang | Hap1      | D12 | Xingren            | Guangdong| Hap1      |
| Z3  | Cuifeng            | Zhejiang | Hap2      | D13 | Yashidancong       | Guangdong| Hap1      |
| Z4  | Huangjiyina        | Zhejiang | Hap1      | D14 | Zhalanxiang        | Guangdong| Hap1      |
| Z5  | Jinfeng            | Zhejiang | Hap3      | D15 | Zonglvxiang        | Guangdong| Hap2      |
| Z6  | Jinfeng            | Zhejiang | Hap2      | S1  | Chongpi 71-1       | Sichuan  | Hap5      |
| Z7  | Julanxiang         | Zhejiang | Hap2      | S2  | Chongqingpipacha   | Sichuan  | Hap2      |
| Z8  | Pingyangtezao      | Zhejiang | Hap2      | S3  | Chuancha 2         | Sichuan  | Hap2      |
| Z9  | Shuiqucha          | Zhejiang | Hap2      | S4  | Chuancha 3         | Sichuan  | Hap2      |
| Z10 | Wuniuzuo           | Zhejiang | Hap2      | S5  | Chuancha 4         | Sichuan  | Hap2      |
| Z11 | Yingshuang         | Zhejiang | Hap14     | S6  | Chuancha 5         | Sichuan  | Hap2      |
| Z12 | Zhenong 113        | Zhejiang | Hap2      | S7  | Chuanhuang 1       | Sichuan  | Hap2      |
| Z13 | Zhenong 117        | Zhejiang | Hap9      | S8  | Chuanhuang 2       | Sichuan  | Hap2      |
| Z14 | Zhenong 701        | Zhejiang | Hap2      | S9  | Chuanbm 28         | Sichuan  | Hap2      |
| Z15 | Zhenong 702        | Zhejiang | Hap1      | S10 | Chuanbm 217        | Sichuan  | Hap4      |
| Z16 | Zhenong 901        | Zhejiang | Hap1      | S11 | Chuanmonghuangyazao| Sichuan  | Hap2      |
| Z17 | Zhongcha 102       | Zhejiang | Hap1      | S12 | Ermei 1            | Sichuan  | Hap10     |
| Z18 | Zhongcha 108       | Zhejiang | Hap2      | S13 | Ermeiwenchun       | Sichuan  | Hap2      |
| Z19 | Zhongcha 112       | Zhejiang | Hap7      | S14 | Gulinniuapicha     | Sichuan  | Hap2      |
| Z20 | Zhongcha 302       | Zhejiang | Hap2      | S15 | Huaqiu No.1        | Sichuan  | Hap2      |
| H1  | Baihaozhao         | Hunan    | Hap2      | S16 | Jianhexiangcha     | Sichuan  | Hap2      |
| H2  | Baqinghuangjin No.1| Hunan    | Hap1      | S17 | Mabianlv           | Sichuan  | Hap10     |
| H3  | Fufeng 20          | Hunan    | Hap1      | S18 | Mengshan 4         | Sichuan  | Hap2      |
| H4  | Fuhao              | Hunan    | Hap4      | S19 | Mengshan 9         | Sichuan  | Hap2      |
| H5  | Gaoqiaoaozao 1     | Hunan    | Hap2      | S20 | Mengshan 11        | Sichuan  | Hap2      |
| H6  | Gaoqiaoaozao 4     | Hunan    | Hap2      | S21 | Mengshan 23        | Sichuan  | Hap1      |
| H7  | Gaoqiaoaozao 6     | Hunan    | Hap2      | S22 | Mengshan 28        | Sichuan  | Hap1      |
| H8  | Huangjincha 1      | Hunan    | Hap2      | S23 | Mngshanbaihao      | Sichuan  | Hap1      |
| H9  | Huangjincha 2      | Hunan    | Hap2      | S24 | Mngshanbaihao 131  | Sichuan  | Hap12     |
| H10 | Jianghuakucha      | Hunan    | Hap2      | S25 | Mingshanzao 311    | Sichuan  | Hap13     |
| H11 | Jiaoniu No.1       | Hunan    | Hap2      | S26 | Nanjiang 1         | Sichuan  | Hap2      |
| H12 | Jinyao             | Hunan    | Hap2      | S27 | Nanjiang 4         | Sichuan  | Hap2      |
| H13 | Nanchuangdacha     | Hunan    | Hap2      | S28 | Shuke 1            | Sichuan  | Hap2      |
| H14 | Taidaye            | Hunan    | Hap2      | S29 | Shuke 3            | Sichuan  | Hap10     |
| H15 | Xiangfeilv         | Hunan    | Hap2      | S30 | Shuke 36           | Sichuan  | Hap2      |
| H16 | Xintianwandacha    | Hunan    | Hap5      | S31 | Taizicha           | Sichuan  | Hap2      |
| H17 | Yanwulong          | Hunan    | Hap2      | S32 | Tezao 213          | Sichuan  | Hap10     |
| H18 | Yulan              | Hunan    | Hap2      | S33 | Tianfucha 11       | Sichuan  | Hap5      |
| H19 | Yulv               | Hunan    | Hap2      | S34 | Tianfucha 28       | Sichuan  | Hap2      |
| H20 | Yusun              | Hunan    | Hap2      | S35 | Wumengzao          | Sichuan  | Hap6      |
| D1  | Baxianxiang        | Guangdong| Hap2      | S36 | Xiangshanzeo       | Sichuan  | Hap2      |
| D2  | Baiyedancong       | Guangdong| Hap3      | S37 | Yizao 1            | Sichuan  | Hap2      |
| D3  | Chengmen           | Guangdong| Hap2      | S38 | Zaobaijian 1       | Sichuan  | Hap2      |
| D4  | Fengkuangkuacha    | Guangdong| Hap2      | S39 | Zaobaijian 5       | Sichuan  | Hap4      |
| D5  | Gongxiang          | Guangdong| Hap10     | S40 | Ziyan              | Sichuan  | Hap2      |
According to the haplotype and phylogenetic tree analysis (Figure 3), Hap5 includes Tianfucha No. 11 (S33), Chongpi 71-1 (S1) and Xintianwan dacha (H16), which is consistent with the cluster branches of the two phylogenetic trees. The tea varieties included in Hap1 were all in the same cluster branch with more than 60% support rate of phylogenetic tree, Baiyedancong (D2) and Jinguang (Z5) of Hap3 were also in this branch. It can be seen that the haplotype type is basically consistent with the classification of phylogenetic tree. Wumengzao (S35) is hap6, Yingshuang (Z11) is hap14. They are divided into one group and there are no other tea species, which is consistent with the NJ phylogenetic tree. This also proves that the phylogenetic tree constructed by the adjacency method can better reflect the genetic and phylogenetic relationship of tea varieties than the phylogenetic tree constructed by the maximum parsimony method.

The haplotype phylogenetic tree was constructed by NJ method. It mainly divides 14 haplotypes into three branches, among which Hap6 and Hap14 are a single branch, and there is another branch in the branch with 63% self-development support rate, including Hap1, Hap3 and Hap12.

![Figure 3 NJ phylogenetic tree of 14 haplotype](image)

2 Discussion

Tea is a cross pollination plant, different varieties of tea will have frequent gene exchange, so its genetic background is relatively complex. In addition, tea germplasm resources are relatively rich and the cultivation history is long. People domesticate wild tea species by various ways, and interspecific hybridization is common. Therefore, it is of profound significance to identify tea varieties.

All the primers designed in this study amplified the target bands. The sequencing results of rbcL were consistent, and the sequencing results of matK were different, which indicated that matK could be used as a candidate sequence of DNA barcoding, but as a barcode, the fragments were still too few. In the future research, we will continue to explore chloroplast related genes and ITS Through the combination of related fragments to find the suitable DNA barcoding development for tea tree.

3 Materials and Methods

3.1 Test materials

In this study, 100 different tea varieties were collected from Wuyi University tea tree field genebank (Table 1) in September 2019. DNA extraction was completed on the same day of the samples collected (Huang et al., 2019). The
size and integrity of DNA were detected by 0.7% agarose gel electrophoresis. Finally, the concentration was diluted to 50 ng/μL and stored at -20°C for standby (Nie et al., 2017).

### 3.2 Primer design and PCR amplification

Genome extraction was carried out with plant genome Extraction Kit (Beijing Tiangen). According to a sequencing reaction, the primers which can amplify about 700 BP sequence were designed by primer premier 5.0 (Shanghai, Shenggong). 

- rbcL F: 5’ ATTCGGCGTCAAGGACAT 3’
- rbcL R: 5’ TGCCTGGATCAATCAAAAG 3’
- matK F: 5’ TTTTCTCCGCAAGCAATC 3’
- matK R: 5’ TTACGAGCCAAAGTTCTA 3’.

PCR reaction system (30 μL): PCR master mix 15 μL, primer (10 μmol/L) 1 μL, DNA template (50 ng/μL) 1 μL, sterilized ddH2O 12 μL.

PCR reaction conditions: the reaction system was denatured at 95°C for 4 min, then denatured at 94°C for 30 s, annealed at 50°C for 45 s, finally extended at 72°C for 90 s. The reaction system extended at 72°C for 5 min after 30 cycles.

The PCR products were detected by electrophoresis and sent to Shanghai Shenggong for purification and sequencing.

### 3.3 Data analysis

In order to determine the target fragment, homology of all amplified sequences must be detected by BLAST, the following analysis can be carried out (Nie et al., 2017). The DNA files obtained by sequencing were observed by chromas, and the qualified DNA files were sorted out and constructed into FASTA files. The sequences were compared with ClustalX. The genetic distance was calculated by mega6.0 and the trees were constructed by NJ. The haplotype number, haplotype polymorphism (Hd), nucleotide polymorphism (Pi) and neutral test of 100 tea cultivars were analyzed by DNASP5.10 (NIE et al., 2013).

**Authors’ contributions**

Li Yanyan is the experimental design and executor of this research; Lin Jiajia completes the data analysis and writes of the first draft of the paper; Nie chuanpeng is the designer, guiding the experimental design, data analysis, thesis writing and revision. All authors read and approved the final manuscript.

**Acknowledgments**

This research is jointly funded by Fujian Natural Science Foundation (2019J01826) and Wuyi University advanced talent introduction project (YJ201904, YJ202003).

**Reference**

Huang S.H., Wen L.X., Peng J.R., Tang Y.W., Long L.Y., and Mao L.Y., 2019, Genetic relationship analysis of wild tea tree germplasm resources in part of Guangxi based on EST-SSR markers, Guangxi Zhiwu (Guihaia), 39(6): 821-830

Li Y.Q., Zhao S.G., and Jiao T., 2017, Establishment of DNA barcode of leguminousae forages’ six species, Fenzhi Zhiwu Yuzhong (Molecular Plant Breeding), 15(1): 322-329

Liu Z., Cheng Y., Zhao Y., Yang P.D., and Yang Y., 2018, Genetic diversity and relationship study of Hunan Tea germplasm resources based on chloroplast rbcL and trnH-psbA sequence, Redai Zuowu Xuebao (Chinese Journal of Tropical Crops), 39(1): 40-45

Luo K., Chen S.L., Chen K.L., Song J.Y., Yao H., Ma X.Y., Zhu Y.J., Pang X.H., Yu H., Li X.W., and Liu Z., 2010, Assessment of candidate plant DNA barcodes using the Rutaceae family, Zhongguo Kexue (Science China), 30(4): 342-351

Nie C.P., Zhao J., Li Y.Y., and Wu X.B., 2013, Diversity and selection of MHC class IIb gene exon3 in Chinese alligator, Mol. Biol. Rep., 40: 295-301

Qi H.S., Dai J.N., Yang Y.N., Li S.Y., Wang J., Shi L.C., Wu Y.G., Lai H.G., Hu X.W., and Yu J., 2019, DNA barcoding identification of Camellia spp. seed based on trnH-psbA and matK sequences, Fenzhi Zhiwu Yuzhong (Molecular Plant Breeding), 17(15): 5057-5065

Sha W., Liu L., Zhang M.J., and Ma T.Y., 2018, Application of DNA barcoding in molecular phylogeny of bryophytes, Fenzhi Zhiwu Yuzhong (Molecular Plant Breeding), 16(22): 7438-7442
Tautz D., Arctander P., and Thomas R.H., 2003, A plea for DNA taxonomy, Trend Evol., 18: 70-74
https://doi.org/10.1016/S0169-5347(02)00041-1

Wu F., Pei N.C., Liao B.W., Guan W., Jiang Z.M., Li M., 2020, Assessment of Major Mangrove Plants from Guangdong Province Using DNA Barcode, Journal of Northeast Forestry University, 48(4): 42-49

Xiong Y., Li W.Y., Yang C., Luo B.S., and Yang Q.S., 2019, Bibliometric and visualization analysis of DNA barcoding in plants, Guangxi Zhiwu (Guilhaia), 39(4): 557-568

Zhu Y.D., Cao M.G., Xu Z., Wang K., and Zhang W., 2014, Phylogenetic relationship between xinjiang wild apple (Malus sieversii Roem,) and Chinese apple (Malus domestica subsp. chinesis) based on ITS and matK sequences, Yuanyi Xuebao (Acta Horticulturae Sinica), 41(2): 227-239