Developing high-resolution DNA barcodes for discrimination of timber species using the complete chloroplast genome

Dr. Lichao JIAO

Research Institute of Wood Industry (CRIWI)
Chinese Academy of Forestry (CAF)
Beijing, China
May 20-22, 2019
Introduction

DNA barcoding

Utility and potential for wood species identification
Wood DNA Extraction

Selection of DNA Barcode

Establishment of DNA Database
The selection of high-resolution DNA barcodes

Species identification

➢ Plant DNA barcode

-- mitochondrial cytochrome c oxidase subunit 1 (CO1)
-- nuclear gene (the internal transcribed spacer region, ITS)
-- universal chloroplast DNA barcodes
Wood DNA barcode

--universal chloroplast DNA barcodes (*rbcL*, *matK*, and *psbA-trnH*)
--barcodes combination

A potential strategy is to locate high-resolution DNA barcodes by searching complete chloroplast genomes.
Wood DNA barcode

-- highly variable DNA regions

-- the recovery success rate

A novel strategy concerning the design of high-resolution DNA barcodes for species discrimination of wood tissues based on the complete chloroplast genomes was reported.
Fig. 1  A schematic representation of developing wood DNA barcodes using the complete chloroplast genome
Materials and methods

Plant materials

Fresh leaf samples

-- *Pterocarpus indicus* (the second-class category of the National List of Local Protected Flora in China)

-- *P. santalinus* (CITES II)

-- *P. tinctorius* (wood anatomy similar to *P. santalinus*)

Xylarium (wood collection) of the Chinese Academy of Forestry

30 wood specimens/7–13 individuals per species

-- *P. indicus*  
-- *P. santalinus*  
-- *P. tinctorius*
Sequencing, assembly, and annotation of the complete chloroplast genomes of three *Pterocarpus* species

--fresh leaf samples
--DNeasy Plant Mini Kit (Qiagen)
--Illumina HiSeq 4000

Identification of highly variable regions on the chloroplast genome

--alignment of the chloroplast genomes of the three *Pterocarpus*
--evaluate the nucleotide diversity (Pi)
DNA extraction of wood specimens from xylarium
--modified DNA extraction method based on plant DNA protocol

PCR amplification, sequencing, and phylogenetic analysis of wood specimens from the xylarium
--neighbor-joining (NJ) tree
Results and discussion

Fig. 2a Gene map of three *Pterocarpus* chloroplast genome.

Chloroplast genome features

- *P. indicus* 158,347 bp
- *P. santalinus* 158,965 bp
- *P. tinctorius* 159,116 bp
Table 1 Summary of complete chloroplast genome features of three *Pterocarpus* taxa

<table>
<thead>
<tr>
<th></th>
<th><em>P. indicus</em></th>
<th><em>P. santalinus</em></th>
<th><em>P. tinctorius</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Chloroplast genome length</td>
<td>158,347</td>
<td>158,965</td>
<td>159,116</td>
</tr>
<tr>
<td>LSC length</td>
<td>88,237</td>
<td>88,458</td>
<td>88,631</td>
</tr>
<tr>
<td>IR length</td>
<td>18,730</td>
<td>19,121</td>
<td>19,139</td>
</tr>
<tr>
<td>SSC length</td>
<td>25,690</td>
<td>25,694</td>
<td>25,673</td>
</tr>
<tr>
<td>Total gene number</td>
<td>112</td>
<td>112</td>
<td>112</td>
</tr>
<tr>
<td>No. of protein-coding genes</td>
<td>78</td>
<td>78</td>
<td>78</td>
</tr>
<tr>
<td>No. of tRNA genes</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>No. of rRNA genes</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Total G+C content (%)</td>
<td>36.36</td>
<td>36.31</td>
<td>36.28</td>
</tr>
</tbody>
</table>
Plastome-wide identification of high-variable DNA regions in *Pterocarpus*

**Fig. 2b** Sliding window analysis of the complete chloroplast genomes of three *Pterocarpus* species (window length: 600 bp, step size: 100 bp). X axis: position of the midpoint of a window. Y axis: nucleotide diversity of each window.
Table 2 Features of seven highly variable regions of three *Pterocarpus* complete chloroplast genome

<table>
<thead>
<tr>
<th>Marker</th>
<th>Length</th>
<th>Variable sites</th>
<th>Nucleotide diversity(Pi)</th>
<th>Number of Haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>number</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>rpl32-ccsA</td>
<td>638</td>
<td>19</td>
<td>2.98</td>
<td>0.0217</td>
</tr>
<tr>
<td>rpl20-clpP</td>
<td>758</td>
<td>22</td>
<td>2.90</td>
<td>0.0214</td>
</tr>
<tr>
<td>trnC-rpoB</td>
<td>874</td>
<td>19</td>
<td>2.17</td>
<td>0.0211</td>
</tr>
<tr>
<td>ycf1b</td>
<td>903</td>
<td>26</td>
<td>2.88</td>
<td>0.0193</td>
</tr>
<tr>
<td>accD-ycf4</td>
<td>846</td>
<td>23</td>
<td>2.72</td>
<td>0.0192</td>
</tr>
<tr>
<td>ycf1a</td>
<td>706</td>
<td>20</td>
<td>2.83</td>
<td>0.0191</td>
</tr>
<tr>
<td>psbK-accD</td>
<td>851</td>
<td>22</td>
<td>2.59</td>
<td>0.0183</td>
</tr>
<tr>
<td>ndhF-rpl32</td>
<td>805</td>
<td>17</td>
<td>2.11</td>
<td>0.0158</td>
</tr>
<tr>
<td>trnH-psbA</td>
<td>293</td>
<td>4</td>
<td>1.36</td>
<td>0.0150</td>
</tr>
<tr>
<td>matK</td>
<td>1536</td>
<td>16</td>
<td>1.04</td>
<td>0.0069</td>
</tr>
<tr>
<td>trnL-F</td>
<td>414</td>
<td>3</td>
<td>0.72</td>
<td>0.0052</td>
</tr>
<tr>
<td>rbcL</td>
<td>1428</td>
<td>3</td>
<td>0.21</td>
<td>0.0014</td>
</tr>
</tbody>
</table>
Evaluation of DNA recovery rate and phylogenetic analysis of wood specimens from the xylarium

PCR success rate (>70%)
- rpl32-ccsA (80.0%)
- ycf1b (76.7%)

Sequencing
- rpl32-ccsA
  occurrence of mixed peaks

Fig. 3a Recovery success rates of the seven highly variable DNA regions from wood specimens.
Evaluation of DNA recovery rate and phylogenetic analysis of wood specimens from the xylarium

- **PCR success rate (>70%)**
  - *rpl32-ccsA* (80.0%)
  - *ycf1b* (76.7%)

- **Sequencing**
  - *rpl32-ccsA*
  - Occurrence of mixed peaks

**Fig. 3a** Recovery success rates of the seven highly variable DNA regions from wood specimens.
Fig. 3b Taxon identification tree constructed using neighbor-joining analysis of P distance showing of ycf1b

The phylogenetic tree based on the ycf1b region also indicated that each Pterocarpus wood species clustered together to form a monophyletic group.

The ycf1b region could possibly become a useful DNA barcode for species identification of Pterocarpus wood.
Conclusions

- The chloroplast genomes of three *Pterocarpus* species were reported, and comparative genomic analysis of three *Pterocarpus* chloroplast genome sequences with detailed gene annotation was performed.

- The work verified the result of our previous study that the single universal or widely reported DNA barcodes (*rbcL*, *matK*, and *ndhF-rpl32*) did not have high enough resolution for species identification of *Pterocarpus*, since the nucleotide diversity of these barcodes was relatively low.

- The seven regions with the most variation, i.e., *rpl32-ccsA*, *rpl20-clpP*, *trnC-rpoB*, *ycf1b*, *accD-ycf4*, *ycf1a*, and *psbK-accD*, from the complete chloroplast genome of *Pterocarpus* were identified.
After comprehensive evaluation of species discrimination ability and the recovery success rate, the $ycf1b$ region was identified as a prioritized high-resolution chloroplast barcode and as a potential molecular marker for species identification of *Pterocarpus* wood.

This work reports a strategy for developing specific chloroplast DNA barcodes of wood tissues, especially for the target group, on the scale of the chloroplast genome.

Future work will include more species/specimens of the *Pterocarpus* genus and involve the exploration of nuclear DNA barcodes.
Acknowledgements

--The International Association of Wood Anatomists (IAWA)

--International Union of Forest Research Organizations (IUFRO)

--Chinese Academy of Forestry (CAF)

--Research Institute of Wood Industry, Chinese Academy of Forestry

--Forest Products Laboratory, Forest Service, USDA
Thanks for Your Comments

Dr. Lichao Jiao
jiaolc@caf.ac.cn

Research Institute of Wood Industry (CRIWI)
Chinese Academy of Forestry (CAF)