Identification of *Uvaria* sp by barcoding coupled with high-resolution melting analysis (Bar-HRM)

M. Osathanunkul¹, P. Madesis², S. Ounjai¹, K. Pumiputavon¹, R. Somboonchai³, P. Lithanatudom¹, T. Chaowasku¹, J. Wipasa⁴ and C. Suwannapoom⁵

¹Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand
²Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thessaloniki, Greece
³Program of Geosocial Based Sustainable Development, Faculty of Agricultural Production, Maejo University, Chiang Mai, Thailand
⁴Research Institute for Health Sciences Chiang Mai University, Chiang Mai, Thailand
⁵School of Agriculture and Natural Resources, University of Phayao, Phayao, Thailand

Corresponding author: M. Osathanunkul
E-mail: omaslin@gmail.com

Received August 10, 2015
Accepted October 15, 2015
Published January 13, 2016
DOI http://dx.doi.org/10.4238/gmr.15017405

**ABSTRACT.** DNA barcoding, which was developed about a decade ago, relies on short, standardized regions of the genome to identify plant and animal species. This method can be used to not only identify known species but also to discover novel ones. Numerous sequences are stored in online databases worldwide. One of the ways to save cost and time (by omitting the sequencing step) in species identification is to use available barcode data to design optimized primers for further analysis, such as high-resolution melting analysis (HRM). This study aimed to determine the effectiveness of the hybrid method Bar-HRM (DNA barcoding combined with HRM) to identify species that share similar external morphological features, rather than conduct traditional taxonomic identification that
require major parts (leaf, flower, fruit) of the specimens. The specimens used for testing were those, which could not be identified at the species level and could either be *Uvaria longipes* or *Uvaria wrayias*, indicated by morphological identification. Primer pairs derived from chloroplast regions (*mat*K, *psbA-trnH*, *rbcL*, and *trnL*) were used in the Bar-HRM. The results obtained from *psbA-trnH* primers were good enough to help in identifying the specimen while the rest were not. Bar-HRM analysis was proven to be a fast and cost-effective method for plant species identification.

**Key words:** Barcoding; High-resolution melting; Species identification; *Uvaria* sp; Chloroplast genes

**INTRODUCTION**

Over the last decade, DNA barcoding has enabled researchers to change the process of species identification. DNA barcoding is based on sequence diversity within a short and standardized gene region for species discrimination. The method can not only identify known species but also discover novel ones (Hebert et al., 2003; Mitchell, 2008). DNA barcoding involves isolating DNA from the organism of interest, performing polymerase chain reaction (PCR) of a selected region, sequencing the PCR product, and identifying the organism by comparing the sequence to a database (Hajibabaei et al., 2006; Hebert et al., 2004, 2010; Van der Bank et al., 2012; Franzini et al., 2013).

The rich online database of sequences can be used for designing optimized primers for further investigation such as high-resolution melting (HRM) analysis, and can be highly effective in saving cost and time for species identification because the relatively expensive and time-consuming sequencing step could be skipped.

Annonaceae is the most species-rich family of Magnoliales, comprising about 2400 species belonging to 108 genera (Rainer and Chatrou, 2006). It has long been referred to as a complex phylogenetic and taxonomic group. The Annonaceae classifications were proposed based on different sources of data: morphological characters (Hutchinson, 1923; Fries, 1959) and phylogenetic analysis (Doyle and Le Thomas, 1994; Mols et al., 2004; Pirie et al., 2006; Couvreur et al., 2008; Chatrou et al., 2012) but the classification not yet stable as newly discovered diversity of the family is being reported. Morphological characters that have been used in the family classification were floral, fruit, and seed morphology. The most widely used factor for Annonaceae classification seems to be the one introduced by Fries (1959), primarily based on floral characters.

In this study, we focused on the genus *Uvaria* L., which is one of the largest paleotropical genera in the family Annonaceae (Keßler, 1993), with 210 species. *Uvaria* L. are mainly woody climbers and scandent shrubs, found in the wet tropical forests in Africa, Madagascar, continental Asia, Malaysia, northern Australia, and Melanesia. The genus has its center of diversity in continental Southeast Asia and Malaysia (Meade, 2005).

*Uvaria* L. is characterized by a combination of morphological features, including stellate hairs (throughout the plant), generally valvate sepals, partially or fully imbricate petals, numerous stamens and carpels, inaperturate monad pollen, and generally many seeded apocarpous fruits (Meade, 2005; Zhou et al., 2009).

The major distinguishing feature of the two closely related *Uvaria* species - *Uvaria*
Identification of *Uvaria* sp by a molecular hybrid method

**MATERIAL AND METHODS**

**DNA mining of barcode regions**

Sequences for selected plastid regions (*matK, psbA-trnH, rbcL*, and *trnL*) were retrieved from NCBI GenBank (National Center for Biotechnology Information, USA). Low-quality sequences and accessions without specified herbarium vouchers were discarded. Analyses of all downloaded
sequences were performed using multiple-sequence alignments made using the SeqMan Pro software program (DNASTAR, Inc., Madison, WI, USA), ClustalX2 (Larkin et al., 2007), and MEGA5 (Tamura et al., 2011). Variable characters and GC (guanine-cytosine) content were recorded.

Plant materials and DNA isolation

Three *Uvaria* species were used as plant materials for the analysis. The species and their locations are listed in Table 1. Plant samples were ground with liquid nitrogen, and 100 mg of the material was used for DNA extraction using a NucleoSpin Plant® II (Macherey-Nagel, Düren, Germany) kit following manufacturer instructions. The DNA was stored at -20°C for further use.

<table>
<thead>
<tr>
<th>Species Abbreviation</th>
<th>Location</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uvaria cf. longipes</td>
<td>L1</td>
<td>Khao Hin Son Botanical Garden 13.749527, 101.505463</td>
</tr>
<tr>
<td></td>
<td>L2</td>
<td>Phanomsarakham, Chachoengsao 13.790384, 100.372378</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Taweewattana, Bangkok</td>
</tr>
<tr>
<td>Uvaria siamensis</td>
<td>S</td>
<td>Dept. of Biology Cultivated Plantation Faculty of Science, Chiang Mai University</td>
</tr>
<tr>
<td>Uvaria cf. wrayi</td>
<td>W1</td>
<td>Khao Hin Son Botanical Garden 13.749527, 101.505463</td>
</tr>
<tr>
<td></td>
<td>W2</td>
<td>Phanomsarakham, Chachoengsao 13.790384, 100.372378</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Khao Hin Son Botanical Garden Phanomsarakham, Chachoengsao</td>
</tr>
<tr>
<td>Uvaria sp</td>
<td>U</td>
<td>Taweewattana, Bangkok          13.790384, 100.372378</td>
</tr>
</tbody>
</table>

Table 1. Species and location of plants included in this study.

HRM method

DNA amplification using real-time PCR was performed using the Eco Real-Time PCR (RT-PCR) system (Illumina®, San Diego, CA, USA) in order to establish characteristic melting temperatures (Tm) to enable the distinction of the different medicinal plants. The reaction mixture (total volume 10 µL) for the RT-PCR and HRM analysis consisted of 5 µL 2X THUNDERBIRD® SYBR® qPCR Mix, 0.2 µM forward primer, 0.2 µM reverse primer, and 1 µL 25 ng/µL DNA. The primer pairs used for each region are listed in Table 2. SYBR fluorescence dye was used to monitor the accumulation of the amplified product during PCR and the HRM process to derive the Tm value.

<table>
<thead>
<tr>
<th>Primer HRM</th>
<th>5’-3’</th>
<th>Ta (°C)</th>
<th>Expected size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRM_matK1F</td>
<td>CTCTATTACGATACATCCCTCT</td>
<td>57</td>
<td>170</td>
</tr>
<tr>
<td>HRM_matK1R</td>
<td>TTCTTTGATACGAACTTATG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRM_psbA-trnH F</td>
<td>ATGGGTTATTTGTTTTG</td>
<td>57</td>
<td>115-150</td>
</tr>
<tr>
<td>HRM_psbA-trnH R</td>
<td>TGTATTATATATATACATCATACTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRM_rbcLBF</td>
<td>GGTACATGGAGAACTTGTGGA</td>
<td>57</td>
<td>150</td>
</tr>
<tr>
<td>HRM_rbcLBR</td>
<td>ACAGAACCTCTCTCAAAAGGTTCTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRM_trnL1F</td>
<td>TGGGCAATCTGAGCCAAATC</td>
<td>57</td>
<td>120</td>
</tr>
<tr>
<td>HRM_trnL1R</td>
<td>AACAGCTCCATTGAGCTCTGACCACT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The RT-PCR amplification was performed on a 48-well Helixis plate using an initial denaturing step at 95°C for 5 min followed by 35 cycles of 95°C for 30 s, 57°C for 30 s, and 72°C for 20 s. Fluorescence data were acquired at the end of each extension step during the PCR.
cycles. Subsequently, the PCR amplicons were denatured for HRM at 95°C for 15 s, and then annealed at 50°C for 15 s to form random-DNA duplexes. The RT-PCR HRM protocol collected fluorescence data at 0.1°C temperature increments. The EcoStudy Software v 5.0 was used to plot a normalized curve of decreasing fluorescence with increasing temperature. The negative derivative of fluorescence (F) over the temperature (T) (dF/dT) curve gives the Tm. To generate normalized melting curves and difference melting curves (Wittwer et al., 2003), pre- and post-melt normalization regions were set to define the main temperature boundaries of the normalized and difference plots with U. siamensis set as the reference species.

RESULTS

DNA mining

Sequences for four selected plastid regions (matK, psbA-trnH, rbcl, and trnL) were retrieved from GenBank (NCBI) for each of the species in the genus Uvaria (Annonaceae) (Table S1). The primary number of resulting sequences downloaded from GenBank of matK, psbA-trnH, rbcl, and trnL were 86, 67, 94, and 98, respectively. The length of the downloaded sequences ranged from 751 to 1353 bp for matK, 209 to 428 bp for psbA-trnH, 446 to 1566 bp for rbcl, and 194 to 1421 bp for trnL. Across all search results from GenBank, the largest group of sequences in the Uvaria genus (98 records) retrieved belonged to trnL region on the chloroplast genome. However, the length of rbcl was the longest at 1566 bp and the shortest was psbA-trnH, which is of 209 bp (Table 3).

<table>
<thead>
<tr>
<th>Regions</th>
<th>Search result from GenBank</th>
<th>Length (bp)</th>
<th>Sequences in analysis dataset</th>
<th>Length (bp)</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>matK</td>
<td>86</td>
<td>751-1353</td>
<td>73</td>
<td>360</td>
<td>61</td>
</tr>
<tr>
<td>psbA-trnH</td>
<td>67</td>
<td>209-428</td>
<td>49</td>
<td>360</td>
<td>48</td>
</tr>
<tr>
<td>rbcl</td>
<td>94</td>
<td>446-1566</td>
<td>77</td>
<td>360</td>
<td>63</td>
</tr>
<tr>
<td>trnL</td>
<td>98</td>
<td>194-1421</td>
<td>50</td>
<td>360</td>
<td>45</td>
</tr>
</tbody>
</table>

Low-quality sequences and accessions without specified herbarium vouchers were discarded. After sorting out of irrelevant, unverified, ambiguous, and outnumbered sequences followed by alignment and manual editing, the number of remaining sequences in analysis dataset of matK, psbA-trnH, rbcl, and trnL were reduced to 73, 49, 77, and 50 covering 61, 48, 63, and 45 species, respectively. After filtering sequences with criteria mentioned above, rbcl was the largest group of samples covering the highest number of species (Table 3).

All processed sequences were subjected to a comparison of variable sites with 360 bp being the total length of the analyzed sequences of each region. At 31.7% (114 of 360 bp), the psbA-trnH fragment sequences were observed to have higher nucleotide variation than the fragment sequences of the other regions (Figure 1). The relative nucleotide variation within fragment sequences was as follows: psbA-trnH>matK>rbcl>trnL (Table 4 and Figure 1). However, within the variable sites, if any site contains at least two types of nucleotides, and if at least two of them occur with a minimum frequency of two, it is then classified as parsimony-informative site. The percentage of parsimony-informative sites found on each plastid region was 76.1% (35 of 46), 64% (73 of 114), 69.1% (29 of 42), and 30% (12 of 40) for matK, psbA-trnH, rbcl, and trnL,
respectively (Table 4 and Figure 2). The HRM primers derived from these four selected regions were used in HRM analysis.

**Table 4.** Characteristics of sequences and primers used for high-resolution melting analysis.

<table>
<thead>
<tr>
<th>Regions</th>
<th>matK</th>
<th>psbA-trnH</th>
<th>rbcl</th>
<th>trnL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available species</td>
<td>70</td>
<td>66</td>
<td>68</td>
<td>70</td>
</tr>
<tr>
<td>Conserved sites (%)</td>
<td>314 (87.2)</td>
<td>205 (56.9)</td>
<td>318 (88.3)</td>
<td>320 (88.9)</td>
</tr>
<tr>
<td>Variable characters (%)</td>
<td>46 (12.8)</td>
<td>114 (31.7)</td>
<td>42 (11.7)</td>
<td>40 (11.1)</td>
</tr>
<tr>
<td>Parsimony-informative site (%)</td>
<td>35 (76.1)</td>
<td>73 (64.0)</td>
<td>29 (69.1)</td>
<td>12 (30.0)</td>
</tr>
<tr>
<td>Conserved forward primer/total (%)</td>
<td>14/26 (53.8)</td>
<td>20/24 (83.3)</td>
<td>22/22 (100)</td>
<td>21/21 (100)</td>
</tr>
<tr>
<td>Conserved reverse primer/total (%)</td>
<td>12/23 (52.2)</td>
<td>24/28 (85.7)</td>
<td>24/25 (96.0)</td>
<td>27/27 (100)</td>
</tr>
<tr>
<td>Average %GC content</td>
<td>35.9</td>
<td>29.9</td>
<td>47.4</td>
<td>36.1</td>
</tr>
</tbody>
</table>

**Figure 1.** Relative nucleotide variation within fragment sequences of the four chloroplast regions: A. *matK*; B. *psbA-trnH*; C. *rbcl*; and D. *trnL*. Details on nucleotide variants found on the variation sites were shown as two to the maximum of four.
Evaluation of four primer pairs used in Bar-HRM

Four primer pairs of four chloroplast regions were used in HRM analysis (Table 2). These four primer sets amplified products from matK, psbA-trnH, rbcL, and trnL and were expected to yield amplicons to the sizes of 170, 115-150, 150, and 120 bp, respectively. The forward and reverse matK primers matched the consensus sequence of the downloaded species at the binding sites in only 14 of 26 sites (53.8%) and 12 of 23 sites (52.2%), respectively (Table 4). Similarly, psbA-trnH primers matched the consensus sequence of the downloaded species at the binding sites in 14 of 24 sites (58.3%) of forward primer and 16 of 28 sites (57.1%) of reverse primer. The average %GC content of amplicons was calculated in order to predict the variation in melting curves for the different markers. The psbA-trnH had the lowest average %GC content, with 29.9%, followed by matK, trnL, and rbcL, with 35.9, 36.1, and 47.4%, respectively (Table 4).

HRM analysis was performed in triplicates to establish the Tm for each primer set. The shapes of the melting curves were analyzed using the EcoStudy Software v 5.0 to distinguish the different plant species. Performing HRM analysis using a set of primers of the four selected regions would yield amplicons ranging from 100 to 170 bp as mentioned above. Reed and Wittwer (2004) found that amplicons suitable for HRM analysis should be 300 bp or less for optimal results.
However, both sequence length and the nucleotide variation within sequences influence the dissociation energy of the base pairs and result in different Tm values. The \textit{psbA-trnH} amplicon sequences were observed to have higher nucleotide variation (31.7\%) than the amplicons of the other regions, and thus the \textit{psbA-trnH} primer pair in this study was expected to be a suitable primer for HRM analysis for identifying tested plant species. On the other hand, in this study, none of amplicons could be detected from HRM experiments using \textit{matK} primer set. High universality at the initial bases of the primer site is crucial for primer annealing and subsequent elongation initiation by the DNA polymerase. The forward and reverse \textit{matK} primers were found to match the consensus sequence of the \textit{Uvaria} species at just about half of the primer length (Table 4). This could explain why we cannot get any results (amplicons) from performing HRM using the \textit{matK} primer pair. In contrast, \textit{rbcL} and \textit{trnL} primer pairs have 100\% and \textit{psbA-trnH} 83.3\% match with the \textit{Uvaria} species sequences at the binding sites, therefore, amplicons from these primer sets could be generated.

**Species identification of the target plant**

The \textit{rbcL} and \textit{trnL} primer pairs were used in HRM analysis. The amplicons were analyzed to determine the Tm and the results presented by means of conventional derivative plots, which show the Tm value for the amplified fragment from each species. The melt curve is generated by slowly melting the DNA of tested plant species through a range of temperatures in the presence of a dsDNA binding dye. The melting temperature peaks of the \textit{Uvaria} plant species are calculated as Tm. HRM analysis with the \textit{rbcL} primer pair failed to distinguish between the two closely related \textit{Uvaria} species (\textit{U. longipes} and \textit{U. wrayi}) as the melt curves of the two species could not be distinguished from each other (Figure 3A), although the melt curve of \textit{U. siamensis} is clearly distinct from the two species. Also none of the tested species could be distinguished using \textit{trnL} primers (Figure 3B). Therefore, this primer pair of \textit{rbcL} and \textit{trnL} is not suitable for identifying the target sample. The \textit{rbcL} did not give satisfactory results in this study although it has been suggested as a core marker for plant DNA barcode. Also, Bar-HRM using \textit{rbcL} was reported to be useful in authenticating herbal species in commercial products (Osathanunkul et al., 2015). In contrast, the individual melting curves were reproducibly achieved using \textit{psbA-trnH} primers from each of the three different \textit{Uvaria} species in triplicate analyses (Figure 4A and B). Thus, the method is applicable for identifying the target plant species. In Figure 4A, there are clearly separate groups of the plant samples as \textit{U. longipes}, \textit{U. wrayi}, and \textit{U. siamensis}. Melt curve of the target plant was found to be in a \textit{U. wrayi} cluster with a 90\% confidence interval. This indicates that the target plant could be identified as \textit{U. wrayi}. The whole process from start until species identification took just about 2 h. Such findings suggest that Bar-HRM may be used as a powerful aiding tool for plant species identification and for discriminating closely related species, rather than the traditional method of checking morphological characteristics, which would involve waiting for about two years for the flowering to occur.

**Confirmation of Bar-HRM**

DNA of all samples (two samples of \textit{U. longipes} (L1 and L2), two samples of \textit{U. wrayi} (W1 and W2), one of \textit{U. siamensis} (S), and the unknown sample (U)) were sent to be sequenced for confirming the Bar-HRM results. To identify similarity of the samples, a comparison of homologous
sequences using sequence alignment techniques was carried out. Sequences of L1 and L2 contain 100% base similarity in all four regions (\textit{matK}, \textit{psbA-trnH}, \textit{rbcl}, and \textit{trnL}). Likewise, W1, W2, and U share 100% similarity in their nucleotide composition. All four gene sequences were deposited in GenBank (Table 5). Therefore, these data could be used to support our Bar-HRM results that indicated the unknown sample as \textit{U. wrayi}.

**Table 5. Sequences deposited in GenBank with accession number.**

<table>
<thead>
<tr>
<th>Species</th>
<th>Samples</th>
<th>\textit{matK}</th>
<th>\textit{psbA-trnH}</th>
<th>\textit{rbcl}</th>
<th>\textit{trnL}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Uvaria longipes}</td>
<td>L1</td>
<td>KR824531</td>
<td>KR905569</td>
<td>KR905574</td>
<td>KR905579</td>
</tr>
</tbody>
</table>

...
DISCUSSION

In this study, we describe the development of a hybrid method, Bar-HRM, for identification of a plant sample of *Uvaria* species, of which not all important morphological features could be collected. The target species plant was at an early age stage, with no flowers produced. The two closely related species, *U. longipes* and *U. wrayi*, share several morphological similarities. Flowers of these plants are critical in species identification of these plants. However, it usually takes several years for the plant to flower, and in our case we would have had to wait for two more years.

The method developed here was proven to be effective in distinguishing between three *Uvaria* species (*U. longipes*, *U. siamensis*, and *U. wrayi*). The DNA extracted from tested samples yielded a specific amplification product with the *psbA-trnH*, *rbcL*, and *trnL* primers but not *matK* set. The *matK* locus is one of the most variable plastid coding regions and has high interspecific divergence and good discriminatory power. However, it can be difficult to amplify with the standard barcoding primers due to high substitution rates at the primer sites (CBOL Plant Working Group, 2009; Hollingsworth, 2011). In addition to the *matK* and *rbcL* barcodes, other regions, *psbA-trnH* and *trnL* were suggested as supplementary DNA barcodes for plants (Hollingsworth, 2011; Li et al., 2011). Only the use of *psbA-trnH* primers reported here is good for the tested *Uvaria* species identification. The normalized HRM curves for the amplicons, from the two *Uvaria* species and the target species, based on HRM analysis with barcode marker *psbA-trnH* were easily to spot, and the target sample could be successfully assigned to the *U. wrayi* species.

Thus, Bar-HRM analysis was proven to be a fast and cost-effective method for plant species differentiation and/or identification.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the CMU Research Group Grant (Chiang Mai University).

REFERENCES

http://dx.doi.org/10.1073/pnas.0905845106

http://dx.doi.org/10.1111/j.1095-8339.2012.01235.x

http://dx.doi.org/10.1016/j.ympev.2008.01.018

http://dx.doi.org/10.1080/12538078.1994.10515148

http://dx.doi.org/10.1371/journal.pone.0002802

http://dx.doi.org/10.1111/j.1467-2379.2008.02052.x
Identification Uvaria sp by a molecular hybrid method


**Supplementary material**

**Table S1.** *Uvaria* plant sequences of four plastid regions (*matK*, *psbA-trnH*, *rbcL*, *trnL*) were retrieved from GenBank (NCBI) for each of the species with Accession number.