CHAPTER III

DNA BARCODES OF ELEVEN *ARISTOLOCHIA* SPECIES

3.1 Introduction

At present, biomolecular technologies are popular tools for plant identification. DNA barcode is a latest focused technique. The objective of DNA barcoding is to find one or a few DNA regions that can distinguish among the species existing on the earth and to obtain the DNA information to produce a large-scale database of creatures of the world (Hollingsworth 2011). In recent years, DNA barcoding of land plants using four standard regions (*rbc*L, *mat*K, ITS and *trn*H–*psb*A) has been recommended as a tool for identification at species level for land plant (Li, Cao et al. 2011). In this study, the *rbc*L gene, *mat*K gene, ITS region and *trn*H-*psb*A region were studied for the identification and discrimination of different eleven species in the genus *Aristolochia* using the DNA barcoding technique.

3.2 Materials and methods

3.2.1 Plant materials

Forty-two samples of eleven *Aristolochia* taxa, considered as authentic samples were collected from various locations from Thailand (Table 2). All specimens were identified by Assoc. Prof. Thatree Phadungcharoen at the Department of Pharmacognosy and Pharmaceutical Botany, Faculty of Pharmaceutical Science, Chulalongkorn University, Thailand. Herbarium voucher specimens were prepared and kept at the Museum of Natural Medicines, Faculty of Pharmaceutical Science, Chulalongkorn University.

Table 2 Plant materials and their accession numbers for DNA barcodes.

Species	Place of collection (Thailand, Province)	Voucher number	GenBank accession number			
			*rbc*L	*mat*K	ITS	*trn*H-*psb*A
A. anguicida Jacq.	Chiang Mai	MUS-5405	KP903720	KP998777˙	KP998791	KP998805
	Chiang Mai	MUS-5406	-	-	-	-
A. gigantea Mart. et Zucc.	Nakhon Pathom	MUS-5376	-	-	-	-
	Bangkok	MUS-5377	-	-	-	-
	Chiang Mai	MUS-5396	-	-	-	-
	Bangkok	MUS-5393	KP998764	KP998778	KP998792	KP998806
	Bangkok	MUS-5394	-	-	-	-
	Bangkok	MUS-5395	-	-	-	-
	Chiang Mai	MUS-5397	-	-	-	-
A. grandiflora Sw.	Phitsanulok	MUS-5379	-	-	-	-
	Petchabun	MUS-5390	-	-	-	-
	Lampang	MUS-5392	-	-	-	-
	Lampang	MUS-5391	KP998765	KP998779	KP998793	KP998807
	Phitsnulok	MUS-5380	-	-	-	-

Table 2 Plant materials and their accession numbers for DNA barcodes (continued).

Species	Place of collection (Thailand, Province)	Voucher number	GenBank accession number			
			*rbc*L	*mat*K	ITS	*trn*H-*psb*A
A. *kerrii* Craib	Chiang Mai	MUS-5415	-	-	-	-
	Chiang Mai	MUS-5413	KP998766	KP998780	KP998794	KP998808
A. *littoralis* D. Parodi	Bangkok	MUS-5404	KP998767	KP998781	KP998795	KP998809
A. *pierrei* Lecomte	Sakon Nakhon	MUS-5407	-	-	-	-
	Sakon Nakhon	MUS-5408	-	-	-	-
	Sakon Nakhon	MUS-5409	KP998768	KP998782	KP998796	KP998810
	Sakon Nakhon	MUS-5410	-	-	-	-
	Sakon Nakhon	MUS-5411	-	-	-	-
A. *pothieri* Pierre ex Lecomte	Bangkok	MUS-5374	KP998769	KP998783	KP998797	KP998811
	Bangkok	MUS-5381	-	-	-	-
	Bangkok	MUS-5382	-	-	-	-
	Bangkok	MUS-5402	-	-	-	-
	Bangkok	MUS-5403	-	-	-	-
	Bangkok	MUS-5416	KP998776	KP998790	KP998804	KP998818

Table 2 Plant materials and their accession numbers for DNA barcodes (continued).

Species	Place of collection (Thailand, Province)	Voucher number	GenBank accession number			
			*rbc*L	*mat*K	ITS	*trn*H-*psb*A
A. ringens Vahl	Bangkok	MUS-5375	KP998770	KP998784	KP998798	KP998812
	Bangkok	MUS-5383	-	-	-	-
	Nakhon Pathom	MUS-5384	-	-	-	-
	Nakhon Pathom	MUS-5385	-	-	-	-
	Chiang Mai	MUS-5387	-	-	-	-
	Chiang Mai	MUS-5388	-	-	-	-
	Bangkok	MUS-5389	-	-	-	-
	Bangkok	MUS-5412	-	-	-	-
A. tagala Cham.	Chiang Mai	MUS-5400	KP998772	KP998786	KP998800	KP998814
	Bangkok	MUS-5401	KP998775	KP998789	KP998803	KP998817
	Bangkok	MUS-5386	KP998774	KP998788	KP998802	KP998816
A. tentaculata Schmidt in Fedde	Bangkok	MUS-5398	KP998773	KP998787	KP998801	KP998815
	Bangkok	MUS-5414	-	-	-	-
A. sp.	Bangkok	MUS-5399	KP998771	KP998785	KP998799	KP998813

3.2.2 Genomic DNA extraction

Total genomic DNA was extracted from 80 to 100 mg of fresh leaves from each sample. Extraction of the genomic DNA utilized DNeasy™ Plant Mini Kit (Qiagen, Germany) and Genomic DNA Mini Kit (Plant) (Geneaid, Taiwan) according to the manufacturer's protocol. Briefly, fresh leaves were ground into fine powder in liquid nitrogen by a mortar and a pestle, and suspension and lysis buffers were then added. The lysate was applied to a column to remove precipitates and cell debris by centrifugation. The flow-through fraction was applied to a spin column. The spin column was centrifuged, washed and eluted with elution buffer. Genomic DNA quality were determined by 1% agarose gel electrophoresis, stained with ethidium bromide and visualized under UV light. The extracted DNA samples were stored at -20°C until further use.

3.2.3 Primer design

A pair of the universal primers of each target DNA (Table 3), rbcL_af and rbcL_R23, ITS1 (or ITS3) and ITS4 and trnHR and psbAF, positioned on the conserve regions were used to amplify the complete *rbc*L gene (Figure 2), ITS regions (Figure 4) and *trn*H-*psb*A regions (Figure 5) of the genus *Aristolochia*, respectively. In order to amplify and sequence the complete *mat*K gene of *Aristolochia* plants, the *trn*K-*mat*K sequences of related species in the genus *Aristolochia* including *A. pierrei* (accession number DQ296649), *A. grandiflora* (accession number DQ532052), *A. gigantea* (accession number JX485569) and *A. gigantea* (accession number DQ882187) were aligned and flanking conserved regions were selected. Three primers, matK-Aris-96F, matK-Aris-50F and matK-Aris-201R, were designed based on the sequences of the *trn*K-*mat*K regions obtained from GenBank (Figure 3). In some cases of ITS region, the ITS sequences of related species in the genus *Aristolochia*, including *A. cucurbitifolia* (accession number AM501925), *A. kaempferi* (accession number AM501928),

A. mollissimo (accession number JQ255433), *A. mollissima* (accession number JQ255434), *A. shimadai* (accession number AM501926), *A. kaempferi* (accession number AM501930), *A. kwangsiensis* (accession number FJ980372), *A. faveolata* (accession number AM501927) and *A. zollingeriana* (accession number AM501929) were aligned and flanking conserved regions were selected to design primer ITS-Aris-371F (Figure 4). The obtained sequences of each region were aligned. The flanking conserved regions were selected to design sequencing primers. The locations of amplification primers and the sequencing primers on *rbc*L, *mat*K, ITS and *trn*H–*psb*A are shown in Figure 2, Figure 3, Figure 4 and Figure 5 respectively. Details of these primers are presented in Table 3. The solid line represents the universal primers. The dotted line represents the primers designed in this study. The length of lines represents the length of obtained sequence.
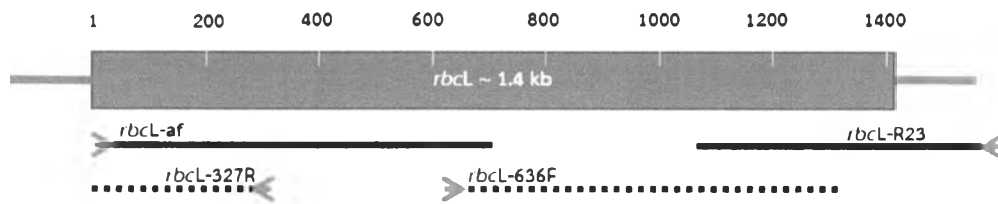
Figure 2 Schematic diagram of the chloroplast *rbc*L gene and relative positions of the PCR amplification primers and sequencing primers used in this study. The arrows represent the directions of the primers.

Figure 3 Schematic diagram of the chloroplast *mat*K gene and relative positions of the PCR amplification primers and sequencing primers used in this study. The arrows represent the directions of the primers.
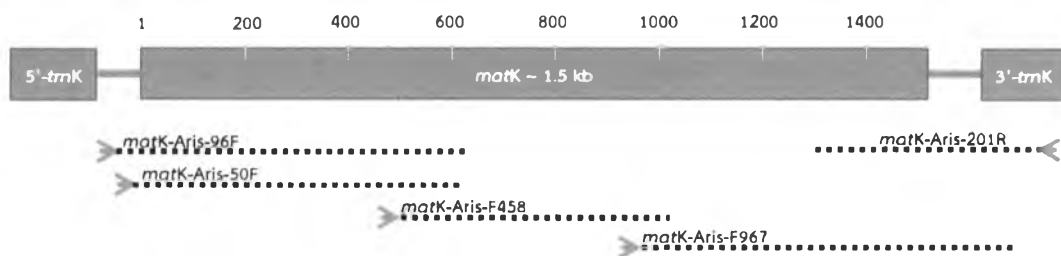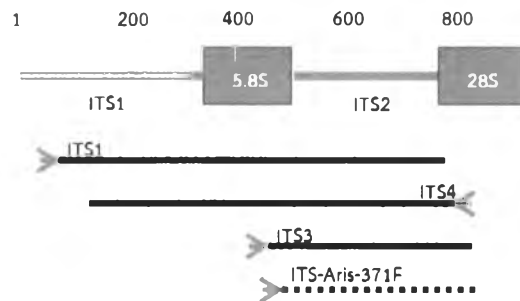
Figure 4 Schematic diagram of ITS region and relative positions of the PCR amplification primers and sequencing primers used in this study. The arrows represent the directions of the primers.

Figure 5 Schematic diagram of the chloroplast intergenic spacer *trn*H-*psb*A and relative positions of the PCR amplification primers and sequencing primers used in this study. The arrows represent the directions of the primers.
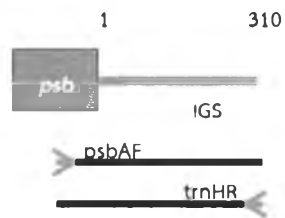
Table 3 Primers used for the generation of DNA barcodes.

Barcode	Primer	Sequence (5'⟶3')	References
*rbc*L	rbcL_aF	ATG TCA CCA CAA ACA GAG ACT AAA GC	(Kress and Erickson 2007)
	rbcL-Aris-327R	TTC AAA AAG GTC TAA AGG GTA AGC	-
	rbcL_636F	GCG TTG GAG AGA TCG TTT CT	(Ohi-Toma, Sugawara et al. 2006)
	rbcL_R23	TTT TAG TAA AAG ATT GGG CCG	(Ohi-Toma, Sugawara et al. 2006)
*mat*K	matK-Aris-96F	ATC CCC TAT TCC TTC AGT TCA A	-
	matK-Aris-50F	CCT TGT TTT GAC TGT ATC GCA C	-
	matK-Aris-F458	ATA CCC CAC CCC ATC CAT CTG	-
	matK-Aris-F967	CAC TTG TGG TCT CAA CCG GG	-
	matK-Aris-201R	GCA CAC GGC TTT CCC TAT G	-
ITS	ITS1	TCC GTA GGT GAA CCT GCG G	(White, Bruns et al. 1990)
	ITS3	GCA TCG ATG AAG AAC GCA GC	(White, Bruns et al. 1990)
	ITS4	TCC TCC GCT TAT TGA TAT GC	(White, Bruns et al. 1990)
	ITS-Aris-371F	AAT TGC AGA ATC CCG CGA AC	-
*trn*H-*psb*A	psbAF	GTT ATG CAT GAA CGT AAT GCT C	(Li, Ling et al. 2010)
	trnHR	CGC GCA TGG TGG ATT CAC AAA TC	(Li, Ling et al. 2010)

3.2.4 PCR amplification of the barcode regions

The complete *rbc*L regions of the *Aristolochia* samples were amplified with a primer pair of rbcL_aF and rbcL_R23. The complete *mat*K regions of the *Aristolochia* samples were amplified with a primer pair of matK-Aris-50F (or matK-Aris-96F) and matK-Aris-R201. The ITS regions of the *Aristolochia* samples were amplified with a primer pair of ITS1 (or ITS3 or ITS-Aris-371F) and ITS4. The *trn*H-*psb*A of the *Aristolochia* samples regions were amplified with a primer pair of trnHR and psbAF. The primers used in this study are shown in Table 3. The PCR amplification was performed in 50 µL of GoTaq® Flexi DNA Polymerase reaction mixture, consisting of 5X PCR buffer, 25 mM MgCl$_2$, 2.5 mM of each dNTP, 10 mM of each primer, 5U *Taq* polymerase (Promega, USA), and 10–100 ng of total genomic DNA as a template. PCR amplifications were carried out in a C1000™ Thermal Cycler (Bio-Rad, USA) using cycling conditions at 96°C for 3 min, followed by 30 cycles of 96°C for 60 s, 55°C for 1 min and 72°C for 2 min (for *rbc*L and *mat*K) and 45 sec (for ITS and *trn*H-*psb*A), and final extension at 72°C for 10 min. The amplified products were detected by 1.2% agarose gel electrophoresis in 1X TAE buffer were then visualized by ethidium bromide staining under UV light.

3.2.5 Cloning technique

In some cases, the PCR products of ITS and *trn*H-*psb*A regions were purified using Wizard® SV Gel and PCR Clean-Up System (Promega, USA), and were then cloned using the pGEM®-T Easy Vector Systems (Promega). The cloning followed standard procedures with 1 µl vector, 1 µl ligase, 5 µl Buffer (all provided with each kit) and 3 µl PCR product. Colony PCR were performed, and then the sequencing process was determined with the amplification primers after plasmid isolation and purification through automatic DNA isolation system PI-50 (Kurabo, Japan).

3.2.6 DNA sequencing of the barcode regions

The PCR products of the *rbc*L, *mat*K, ITS and *trn*H-*psb*A region were purified and then the sequencing process was performed by capillary sequencing (AIT Biotech, Singapore) with sequencing primers for the region. The sequences were aligned, edited, analyzed and corrected using BioEdit Sequence Alignment Editor Version 7.2.5 (Hall 1999). The obtained sequences were assembled for their consensus sequences using DNASTAR® (Version 8.0.2) program. The sequences were then submitted to DDBJ/EMBL/GenBank nucleotide sequence databases with their accession numbers listed in Table 2.

For phylogenetic analysis, the complete *mat*K sequences of the *Aristolochia* spp. were aligned by ClustalX Version 2.1 (Larkin, Blackshields et al. 2007) and MEGA Version 6 (Koichiro Tamura, Glen Stecher et al. 2013). *Thottea dependens* (DQ882194.1) and *Thottea siliquosa* (JN415679.1) were included as outgroup for matK gene sequences.

3.3 Results

3.3.1 Sequence analysis of the barcode regions of eleven *Aristolochia* species

The properties of selected DNA loci are shown in Table 4 including accession number, length, GC content (%) and nucleotide variation (%). The degrees of sequence variations between the *Aristolochia* samples were in the order of ITS > *trn*H-*psb*A > *mat*K > *rbc*L. The large insertions/deletions in ITS caused high variations (28.96%) between the eleven *Aristolochia* plants and were much higher than *trn*H-*psb*A (13.35%), *mat*K *(11.22%)* and *rbc*L *(3.29%)* *(*Table 4*)*.

The complete *rbc*L gene of eleven *Aristolochia* plants was amplified using the amplification primers rbcL_af and rbcL_23R by PCR technique. The PCR products of *rbc*L gene about 1,500 bp in length were obtained. The purified products were sequenced individually using the sequencing primers listed in Table 3. The complete

*rbc*L gene sequences of eleven *Aristolochia* plants were 1,428 bp in length. Sequence distance (percent divergence) were calculated using the program MEGA Version 6. The *rbc*L sequences from all samples of the same species showed completely identical sequence. Forty-seven different sequence variations were found of 1,428 total aligned sites (3.29%). The sequence divergence among eleven *Aristolochia* plants varied from 0% to 1.92%. A pairwise comparison between *A.* sp and *A. anguicida* showed the lowest nucleotide sequence divergence at 0%. Whereas, a pair of *A. pothierri* and *A. anguicida, A. pothierri* and *A. gigantea, A. pothieri* and *A. littoralis* and a pair of *A.* sp *and A. pothieri* showed the highest nucleotide sequence divergence at the same percentage of 1.92%. The nucleotide sequence divergence amoung the eleven *Aristolochia* plants are shown in Table 5. The obtained *rbc*L gene sequences have been deposited in GenBank (Appendix B) and the accession numbers are listed in Table 4.

The complete *mat*K gene of eleven *Aristolochia* plants was amplified using the amplification primers matK-96F (or matK-50F) and matK-201R by PCR technique. The PCR products of *mat*K gene about 1,700 bp in length were obtained. The purified products were sequenced individually using the sequencing primers listed in Table 3. The complete *mat*K gene sequences of the samples were 1,518-1,554 bp in length. The complete *mat*K sequences were aligned by MEGA Version 6. The nucleotide sequences obtained from two samples of *A. pothieri* (KP998811 and KP998818) were completely identical. The sequences obtained from two samples of *A. tagala* (KP998786 and KP998788) were completely identical, but intraspecies variation was found from another one (KP998789) at positions 906 and 1,063 (Appendix B). One hundred and seventy-six different sequence variations were found of 1,569 total aligned sites (11.22%). Sequence distance (percent divergence) were calculated by using the program MEGA Version 6. The sequence divergence among eleven *Aristolochia* plants varied from 0.19% to 4.68%. A pairwise comparison between

A. tagala and *A. pierrei* showed the lowest nucleotide sequence divergence at 0.19%, whereas, a pair of *A. pothierri* and *A. littoralis* showed the highest nucleotide sequence divergence at 4.68%. The nucleotide sequence divergence amoung eleven *Aristolochia* plants are shown in Table 6. The obtained *mat*K gene sequences have been deposited in GenBank (Appendix B) and the accession numbers are listed in Table 4.

The ITS region of *A. grandiflora* and *A. pothieri* were amplified using the amplification primers ITS*1* and ITS4. A primer pair of ITS-Aris-371F and ITS4 was used for *A. anguicida*, *A. kerrii*, *A. pierrei* and *A. tagala*. The other *Aristolochia* plants were amplified by ITS3 and ITS4. The PCR products of ITS region about 750 bp and 450 bp in length were obtained. The purified products were sequenced individually using the sequencing primers listed in Table 3. Some of the amplified products were purified, cloned in competent *Escherichia coli* cells and then sent for DNA sequencing process. The ITS sequences of the samples were 360-751 bp in length. The ITS sequences were aligned by MEGA Version 6. The nucleotide sequences obtained from two samples of *A. pothieri* were completely identical. However, intraspecies variation was found in the sequences obtained from three samples of *A. tagala* at positions 576, 715 and 737 (Appendix B). Two hundred and forty-six different sequence variations were found of 846 total aligned sites (28.96%). Sequence distance (percent divergence) were calculated using the program MEGA Version 6. The sequence divergence among the eleven *Aristolochia* plants varied from 0.50% to 31.32%. A pairwise comparison between *A.* sp and *A. angucida* showed the lowest nucleotide sequence divergence at 0.50%, whereas, a pair of *A. pothierri* and *A. grandiflora* showed the highest nucleotide sequence divergence at 31.32%. The nucleotide sequence divergence between eleven *Aristolochia* plants are shown in Table 7. The obtained ITS gene sequences have been deposited in GenBank (Appendix B) and the accession numbers are listed in Table 4.

The *trn*H-*psb*A region of eleven *Aristolochia* plants was amplified using the amplification primers psbAF and trnHR by PCR technique. The PCR products of *trn*H-*psb*A region about 300-400 bp in length were obtained. The purified products were sequenced individually using the sequencing primers listed in Table 3. Some of the amplified products were purified, cloned in competent *Escherichia coli* cells and then sent for DNA sequencing process. The *trn*H-*psb*A sequences of the samples were 300-369 bp in length. The *trn*H-*psb*A sequences were aligned by MEGA Version 6. The nucleotide sequences obtained from two samples of *A. pothieri* were completely identical. While the sequences obtained from two samples of *A. tagala* (KP998816 and KP998817) were completely identical, intraspecies variation was found when compared to another sample (KP998814) at positions 140, 309 and 310 (Appendix B). Fifty-three different sequence variations were found of 397 total aligned sites (13.35%). Sequence distance (percent divergence) were calculated using the program MEGA Version 6. The sequence divergence among the eleven *Aristolochia* plants varied from 0% to 11.38%. A pairwise comparison between *A. tagala* and *A. anguicida*, *A. tagala* and *A. pierrei*, *A. sp and A. anguicida*, *A.* sp and *A. tagala* showed the lowest nucleotide sequence divergence at 0%. A pair of *A. grandiflora* and *A. anguicida* showed the highest nucleotide sequence divergence at 11.38%. The nucleotide sequence divergence between eleven *Aristolochia* plants are shown in Table 8. The obtained *trn*H-*psb*A sequences have been deposited in GenBank (Appendix B) and the accession numbers are listed in Table 4.

The pairwise percent sequence divergence in *rbc*L, *mat*K, ITS and *trn*H-*psb*A among three *Aristolochia* species used as Krai-Krue are shown in Table 9.

Table 4 Properties of selected DNA loci (*rbc*L, *mat*K, ITS and *trn*H-*psb*A) of *Aristolochia* plants used in this study.

Barcode	*rbc*L				*mat*K			
Property	Accession no.	Length (bp)	GC content (%)	Variation (%)	Accession no.	Length (bp)	GC content (%)	Variation (%)
A. anguicida	KP903720	1428	44.96		KP998777	1539	34.50	
A. gigantea	KP998764	1428	45.03		KP998778	1554	34.23	
A. grandiflora	KP998765	1428	45.03		KP998779	1524	33.92	
A. kerrii	KP998766	1428	44.40		KP998780	1527	34.51	
A. littoralis	KP998767	1428	45.03		KP998781	1539	34.18	
A. pierrei	KP998768	1428	45.10		KP998782	1518	34.58	
A. pothieri	KP998769	1428	44.82	3.29	KP998783	1524	34.19	11.22
A. pothieri	KP998776	1428	44.82		KP998790	1524	34.19	
A. ringens	KP998770	1428	44.89		KP998784	1548	34.75	
A. tagala	KP998772	1428	44.96		KP998786	1518	34.52	
A. tagala	KP998775	1428	44.96		KP998789	1518	34.39	
A. tagala	KP998774	1428	44.96		KP998788	1518	34.52	
A. tentaculata	KP998773	1428	45.03		KP998787	1524	34.84	
A. sp	KP998771	1428	44.96		KP998785	1539	34.44	

Table 4 Properties of selected DNA loci (rbcL, matK, ITS and trnH-psbA) of *Aristolochia* plants used in this study (continued).

Barcode	ITS				trnH-*psb*A			
Property	Accession no.	Length (bp)	GC content (%)	Variation (%)	Accession no.	Length (bp)	GC content (%)	Variation (%)
A. anguicida	KP998791	399	69.42		KP998805	300	40.33	
A. gigantea	KP998792	461	65.73		KP998806	305	40.00	
A. grandiflora	KP998793	751	64.98		KP998807	308	38.96	
A. kerrii	KP998794	360	76.39		KP998808	319	39.81	
A. littoralis	KP998795	437	67.73		KP998809	305	39.67	
A. pierrei	KP998796	379	70.71		KP998810	315	40.32	
A. pothieri	KP998797	696	70.69	28.96	KP998811	369	34.96	13.35
A. pothieri	KP998804	708	70.48		KP998818	369	34.96	
A. ringens	KP998798	432	68.06		KP998812	305	40.00	
A. tagala	KP998800	378	71.69		KP998814	318	39.94	
A. tagala	KP998803	422	68.96		KP998817	319	40.13	
A. tagala	KP998802	422	68.48		KP998816	319	39.81	
A. tentaculata	KP998801	436	66.97		KP998815	310	40.00	
A. sp	KP998799	431	67.95		KP998813	283	40.28	

Table 5 Pairwise percent sequence divergence in the complete *rbc*L gene among eleven species in the genus *Aristolochia*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2	0.14													
3	1.20	1.20												
4	1.70	1.70	1.56											
5	0.14	0.14	1.13	1.70										
6	1.41	1.41	1.41	1.27	1.41									
7	1.92	1.92	1.49	1.49	1.92	0.63								
8	1.92	1.92	1.49	1.49	1.92	0.63	0.00							
9	0.21	0.21	1.13	1.63	0.21	1.34	1.84	1.84						
10	1.56	1.56	1.27	1.13	1.56	0.14	0.49	0.49	1.49					
11	1.56	1.56	1.27	1.13	1.56	0.14	0.49	0.49	1.49	0.00				
12	1.56	1.56	1.27	1.13	1.56	0.14	0.49	0.49	1.49	0.00	0.00			
13	1.06	1.06	0.99	1.41	1.06	1.27	1.34	1.34	0.99	1.13	1.13	1.13		
14	0.00	0.14	1.20	1.70	0.14	1.41	1.92	1.92	0.21	1.56	1.56	1.56	1.06	

1	A. anguicida	5	A. littoralis	9	A. ringens	13	A. tentaculata
2	A. gigantea	6	A. pierrei	10	A. tagala	14	A. sp
3	A. grandiflora	7	A. pothieri	11	A. tagala		
4	A. kerrii	8	A. pothieri	12	A. tagala		

Table 6 Pairwise percent sequence divergence in the complete *mat*K gene among eleven species in the genus *Aristolochia*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2	0.66													
3	3.17	3.53												
4	3.30	3.60	3.74											
5	0.80	0.61	3.74	3.96										
6	2.35	2.80	2.87	2.66	2.90									
7	4.21	4.37	4.18	4.17	4.68	2.14								
8	4.21	4.37	4.18	4.17	4.68	2.14	0.00							
9	1.09	1.57	2.76	3.15	1.72	2.10	3.69	3.69						
10	2.24	2.70	2.67	2.46	2.80	0.19	2.04	2.04	2.00					
11	2.29	2.75	2.72	2.51	2.85	0.28	2.14	2.14	2.04	0.09				
12	2.24	2.70	2.67	2.46	2.80	0.19	2.04	2.04	2.00	0.00	0.09			
13	1.30	1.64	3.02	3.22	1.78	2.20	3.80	3.80	1.05	2.10	2.15	2.10		
14	1.68	1.73	3.96	4.10	1.78	2.80	2.99	2.99	2.42	2.69	2.74	2.69	2.60	

1	A. anguicida	5	A. littoralis	9	A. ringens	13	A. tentaculata
2	A. gigantea	6	A. pierrei	10	A. tagala	14	A. sp
3	A. grandiflora	7	A. pothieri	11	A. tagala		
4	A. kerrii	8	A. pothieri	12	A. tagala		

Table 7 Pairwise percent sequence divergence in ITS region among eleven species in the genus *Aristolochia*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2	5.46													
3	12.66	13.01												
4	22.95	27.01	22.38											
5	3.86	1.86	13.45	24.68										
6	17.35	19.21	19.45	17.47	18.52									
7	22.77	21.64	31.32	21.70	21.36	16.75								
8	22.77	21.64	31.32	21.70	21.36	16.75	0.00							
9	5.20	4.04	13.27	25.07	3.07	17.28	21.18	21.18						
10	15.90	18.47	17.25	16.37	17.43	4.12	15.85	15.85	16.20					
11	15.32	16.00	15.82	15.30	15.09	4.38	13.71	13.71	14.01	0.80				
12	15.32	15.91	15.16	15.99	15.01	3.82	13.92	13.92	13.93	0.27	0.48			
13	8.89	7.85	14.81	25.34	7.56	16.23	21.28	21.28	7.33	15.95	13.78	13.71		
14	0.50	4.54	11.25	22.95	3.08	17.35	20.18	20.18	4.30	15.53	13.43	13.35	7.65	

1	A. anguicida	5	A. littoralis	9	A. ringens	13	A. tentaculata
2	A. gigantea	6	A. pierrei	10	A. tagala	14	A. sp
3	A. grandiflora	7	A. pothieri	11	A. tagala		
4	A. kerrii	8	A. pothieri	12	A. tagala		

Table 8 Pairwise percent sequence divergence in *trn*H-*psb*A among eleven species in the genus *Aristolochia*.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2	2.08													
3	11.38	8.58												
4	6.85	5.56	7.33											
5	2.50	0.36	8.82	5.84										
6	4.67	3.03	6.18	3.68	3.35									
7	8.30	6.35	8.25	7.23	6.61	3.63								
8	8.30	6.40	8.43	7.38	6.66	3.63	0.00							
9	3.78	1.43	9.42	6.34	1.79	3.81	7.14	7.19						
10	0.00	1.83	9.67	5.89	2.16	4.17	7.09	7.14	3.32					
11	5.10	3.27	5.97	3.71	3.58	0.35	4.30	4.39	4.02	4.38				
12	4.67	2.96	5.95	3.56	3.28	0.00	4.19	4.19	3.72	4.08	0.33			
13	2.49	1.07	8.50	5.64	1.41	3.32	6.39	6.43	1.79	2.15	3.46	3.20		
14	0.00	1.83	9.67	5.89	2.16	4.17	7.09	7.14	3.32	0.00	4.38	4.08	2.15	

1	A. anguicida	5	A. littoralis	9	A. ringens	13	A. tentaculata
2	A. gigantea	6	A. pierrei	10	A. tagala	14	A. sp
3	A. grandiflora	7	A. pothieri	11	A. tagala		
4	A. kerrii	8	A. pothieri	12	A. tagala		

Table 9 Pairwise percent sequence divergence in *rbc*L, *mat*K, ITS and *trn*H-*psb*A among three *Aristolochia* species used as Krai-Krue.

		A. pothieri	A. pothieri	A. pierrei	A. tagala	A. tagala	A. tagala	
rbcL	A. pothieri		0.00	2.14	2.04	2.14	2.04	matK
	A. pothieri	0.00		2.14	2.04	2.14	2.04	
	A. pierrei	0.63	0.63		0.19	0.28	0.19	
	A. tagala	0.49	0.49	0.14		0.09	0.00	
	A. tagala	0.49	0.49	0.14	0.00		0.09	
	A. tagala	0.49	0.49	0.14	0.00	0.00		
ITS	A. pothieri		0.00	3.63	7.09	4.30	4.19	trnH-psbA
	A. pothieri	0.00		3.63	7.14	4.39	4.19	
	A. pierrei	16.75	16.75		4.17	0.35	0.00	
	A. tagala	15.85	15.85	4.12		4.38	4.08	
	A. tagala	13.71	13.71	4.38	0.80		0.33	
	A. tagala	13.92	13.92	3.82	0.27	0.48		

3.3.2 Phylogenetic analysis of *Aristolochia* species based on *mat*K sequences

To examine the phylogenetic relationships among the eleven *Aristolochia* taxa in the present study, the complete *mat*K gene sequences of all samples were analyzed. The phylogenetic tree was conducted based on the percent divergence of the complete *mat*K gene from eleven *Aristolochia* plants (Figure 6). A total aligned length of 1,578 bp, including the outgroup sequences, formed the final data set. The monophyly of the genus *Aristolochia* was highly supported in NJ tree with bootstrap value 100 %. The tree revealed that all *Aristolochia* spp. were monophyletically grouped together. The eleven *Aristolochia* species were divided into two groups (Figure 6). The first group consisted of *A. tagala*, *A. pierrei*, *A. pothieri* as one subgroup whereas the bootstrap value is lower than 50 in the other subgroup of *A. grandiflora* and *A. kerrii*. The second group with two subgroups were comprised of *A. tentaculata* and *A. ringens* as one subgroup and *A. sp*, *A. anguicida A. littoralis* and *A. gigantea* all included as the other one subgroup.
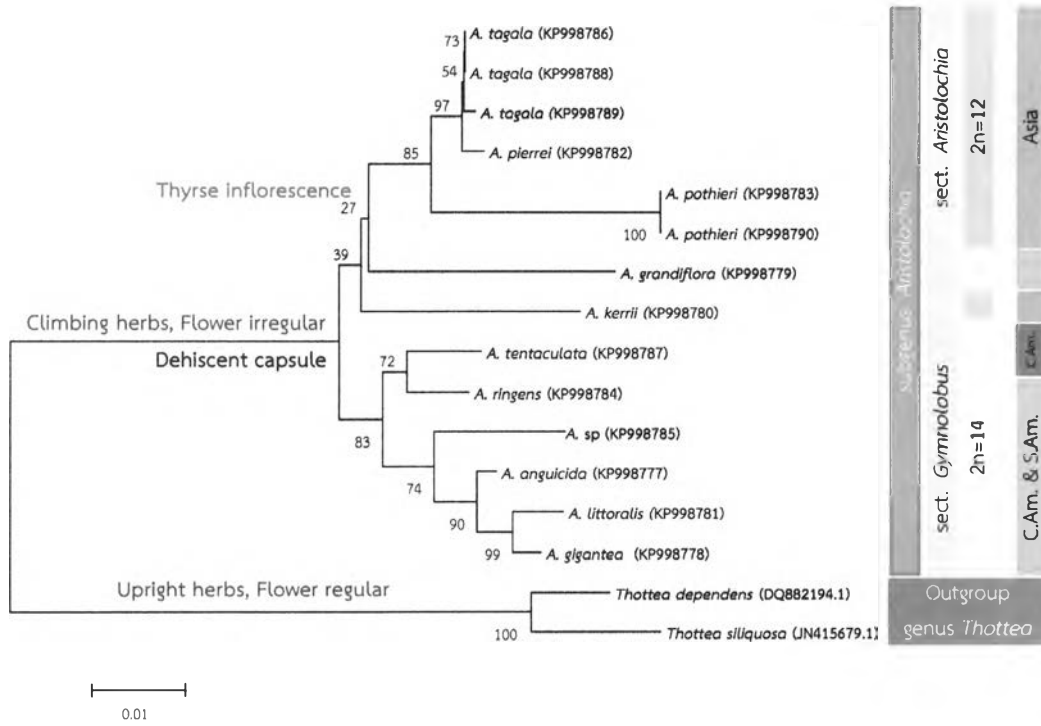
Figure 6 Phylogenetic assessment of eleven *Aristolochia* species constructed with the *mat*K sequences using the Neighbor-joining algorithm (bootstrap values are shown below the branches) with *Thottea dependens* and *T. siliquosa* as the outgroup. The sequence data of the species followed by accession numbers in brackets were retrieved from the GenBank DNA database.

3.4 Discussion

Currently the molecular identification technique is extremely useful for the identification of medicinal herbs. It is included in the latest edition of the Pharmacopoeia of the People's Republic of China and its online supplementary note (Commission 2010b). Amongst various molecular techniques, nucleotide sequencing of standardized region becomes an alternative approach for herbal material identification, as a complement to other methods for better quality control. The chloroplast coding region, *rbc*L and *mat*K sequences, are proposed to be the core standardized region whereas the intergenic spacer *trn*H-*psb*A region and internal transcribed spacer (ITS) (or ITS2) are proposed as additional region for land plant identification at species level (Hollingsworth 2011, Fazekas, Kuzmina et al. 2012). Interestingly, some *Aristolochia* have very limited DNA sequences available in GenBank database. For instance, a partial *trn*H-*psb*A sequence (HG963744.1) of *A. anguicida*, a partial *rbc*L sequence (AB180121.1) of *A. kerrii*, a partial *mat*K sequence (AB060778.1) and a complete *mat*K sequence (DQ296649.1) of *A. pierrei*, and 4 accessions of partial *mat*K sequences (AB211567.1, AB060780.1, AB060779.1 and KF498584.1) of *A. tagala* are provided. Unfortunately, no nucleotide data for *A. pothieri* are available on the existing database and there is no nucleotide sequence of ITS region of the *Aristolochia* plants for use in this study. The limited number of DNA sequences has restricted the development of rapid molecular identification techniques for these herbs.

DNA barcoding based on five candidate regions (*rbc*L, *mat*K, ITS, *trn*H-*psb*A and *trn*L-*trn*F) have been continuously studied for identification of *Aristolochia* plants used in China, for example, *A. fangchi, A. manshuriensis, A. contorta* and *A. debilis* (Li, Au et al. 2014), *A. mollissima* (Li, Au et al. 2012), and *A. californica, A. championii, A. contorta, A. debilis, A. heterophylla* and *A. kaempferi* (Li, Ling et al. 2010). However, prior to this study, there was no DNA barcoding study in other *Aristolochia* species used in Thailand.

In this study, DNA barcoding technique was used for identification of eleven *Aristolochia* species. The four DNA regions (*rbc*L, *mat*K, ITS and *trn*H-*psb*A) were used as suitable DNA regions. The results demonstrated that all four candidate DNA barcodes have performed very well in *Aristolochia* plants. The interspecific genetic distances were obviously greater than the intraspecific distance among the eleven *Aristolochia* species as indicated by *mat*K and *rbc*L. On the other hand, the intraspecific variations of ITS and *trn*H-*psb*A regions were greater than the interspecific variations. The degree of sequence variations amoung the *Aristolochia* samples were in the order of ITS > *trn*H-*psb*A > *mat*K > *rbc*L. The large insertions/deletions in ITS caused high variations (28.96%) amoung the eleven *Aristolochia* plants and were much higher than *trn*H-*psb*A (13.35%), *mat*K (11.22%) and *rbc*L (3.29%).

The complete *rbc*L gene was easily amplified with a universal primer pair and the sequencing result was very good. The obtained fragments were found to be approximately 1,428-1,485 bp in length. The *rbc*L sequence from all samples of the same species showed completely identical sequence. The nucleotide information of the *rbc*L gene among these *Aristolochia* species were highly conservative. *A. anguicida* and *A.* sp could not be distinguished from each other based on the sequence divergences.

The complete *mat*K gene was amplified using primers matK-Aris-96F (or matK-Aris-50F) and matK-Aris-201R based on the sequences of *trn*K-*mat*K region. The *trn*K-*mat*K genes of the eleven *Aristolochia* plants were about 1,610-1,758 bp in length which is consistent with a previous report showing that the *mat*K coding region in the most angiosperms is around 1.5-1.6 kb in length (Neuhaus and Link 1987). As a result, two bases of intraspecies variation were observed in samples of *A. tagala.* collected from different locations. Notably, intra-species polymorphism with *mat*K sequences was generally limited to substitution of one or a few bases (Parmentier, Duminil et al. 2013). According to GenBank database, the variations in the same species were observed

in *A. tagala* (AB060779, AB060780 and AB211567), *A. indica* (AB060771 and AB211579), *A. ringens* (AB211583, AB211584 and AB211586), *A. gigantea* (DQ882187 and JX485569), *A. acuminata* (DQ296646 and DQ532063), *A. albida* (AB211566, DQ296648 and DQ532064), *A. arborea* (DQ532044 and JX485577), *A. baetica* (DQ296653 and DQ882189), *A. cucurbitifolia* (AB060741 and AB180183), *A. kaempferi* (AB060743, AB180157, AB180159, AB180156, AB180158, AB180161, AB180162, AB180163, AB180160, AB180175, AB180165, AB180172, AB180173, AB180174, AB180167, AB180168, AB180171, AB180169, AB180170, AB180166 and DQ532042), *A. kwangsiensis* (AB060745, AB060746, AB060747, AB180176 and AB180179), A. maxima (AB060782 and AB060783), *A. moupinensis* (AB060750, AB060751, AB211562), *A. odoratissima* (AB060798, AB211561 and AB211585), *A. onoei* (AB060753, AB060754, AB060755, AB060756, AB060757, AB180164, AB189730 and AB353065), *A. pistalochia* (DQ296652 and AF543724), *A. ringens* (AB060800 and AB060801), *A. rotunda* (DQ296665 and DQ532061), *A. saccata* (AB211569 and AB353067), *A. salvadorensis* (JX485576 and DQ882191), *A. shimadai* (AB060760, AB060761, AB060762, AB060763). However, the DNA sequences of the *matK* gene among these *Aristolochia* species can be used for discrimination of these plants.

The entire ITS region was amplified using a universal primer pair of ITS1 and ITS4; however, the amplification was not successful in many species. ITS2 is an optional barcode for better amplification and sequencing compared to the entire ITS region (Chen, Yao et al. 2010). ITS3 and ITS-Aris-371F located on 5.8s rDNA was used in the amplification step. The cloning process was also needed for some species. Sequencing analysis failed probably caused by fungal contamination or natural endophytes (Hollingsworth 2011). The entire ITS regions of *A. pothierri* and *A. grandiflora* were observed at 696-751 bp in length, while ITS2 regions of other species were about 360-461 bp in length. The sequencing result indicated that ITS showed the highest

variations (28.96%) because of the large insertions/deletions. It can be used for plant discrimination at species level.

The length of *trn*H-*psb*A region of plant samples were about 300-369 bp. The cloning process was needed to obtain the good quality of sequencing process. The sequencing result indicated that *trn*H-*psb*A is the second region showing high variation (13.35%). It also can be used for plant discrimination at species level.

For unknown species, *A.* sp, could not be identified by morphological characteristics. The BLAST results between *A.* sp and other plant species which are existing on the GenBank database. The closest species referred by *trn*H-*psb*A region are *A. littoralis* (GU135396.2) and *A. ringens* (KP763860.1) at 89% query covery and identity at 97 and 96%, respectively, while the ITS region showed that the closest species are *A. ringens* (KP763867.1 and KP763865.1) at 99% query covery and at 92% identity. From *mat*K BLAST results, the closest species at 100% query covery and identity at 97% are *A. cruenta* (DQ882186.1), *A. cf. cordiflora* (DQ532056.1), *A. nummularifolia* (DQ532053.1), *A. eriantha* (DQ882185.1 and DQ532054.1) and *A. gigantea* (DQ882187.1) and the *rbc*L region indicated that the closed species at 97% query covery and 99% identity are *A. reniformis* (AB205600.1), *A. eriantha* (AB205590.1), *A. grandiflora* (AB205592.1), *A. micrantha* (AB205595.1), *A. maxima* (AB205594.1), *A. gigantea* (AB205591.1), *A. burelae* (AB205587.1), *A. zollingeriana* (AB205599.1), *A. clematitis* (AB205588.1) and *A. pentandra* (AB205596.1). As the BLAST results described above, there is no complete match sequence for *A.* sp through DNA sequences. Although in this study, *A.* sp shows the complete matches to *A. anguicida* by *rbc*L sequences and *A. tagala* by *trn*H-*psb*A sequences, *A.* sp could not be identified by morphological characteristics and nucleotide assessment by now.

The systematic of *Aristolochia* has been based on morphological characters such as perianth tubes, leaves, number of styles and anthers on gynostemium and fruit.

The genus *Aristolochia* consists of three subgenera, the subgenera *Siphisia*, *Aristolochia* and *Pararistolochia*. The subgenus *Siphisia* is characterized by dehiscent capsules, a trilobed each accompanying two anthers, while the subgenus *Aristolochia* is defined by dehiscent capsules, lobes of the perianth unilatally appressed in the bud and breaking up into one to three segments, six or fewer lobes of the gynostemium and six or fewer anthers and chromosome 2n=12, 14 or 16. The subgenus *Pararistolochia* is distinguished from the others by fleshy indehiscent fruits, trilobed perianth, valvate in bud, sometimes with one or three long tails, six to twelve lobes of the gynostemium, six to twenty-four anthers and chromosome 2n=12 (González 1999, Murata, Ohi et al. 2001, Ohi-Toma, Sugawara et al. 2006).

The morphological features indicated that all samples in this study belong to the subgenus *Aristolochia*. The molecular phylogeny potentially divided the eleven *Aristolochia* species into two clades, following the taxonomic suggestion of González (González 1999): the section *Aristolochia* clade of Asia species (*A. tagala, A. pierrei, A. pothieri* and *A. kerrii*) and the section *Gymnolobus* clade of central and south America (*A. tentaculata, A. ringens, A. anguicida, A. littoralis* and *A. gigantea*). *A. grandiflora* was the only taxon congruent with section *Aristolochia* and not grouped specifically with section *Gymnobolus*. This result related to morphological features; section *Aristolochia* is distinguished from section *Gymnobolus* by the thyrsoid branching in the inflorescence.

The plants in subsect. *Podanthemum*, sect. *Aristolochia* were all included in a single clade consisting of *A. pierrei* and *A. tagala* which was supported by a relatively high bootstrap (97%). This group was consistent with morphological characteristics which are the unique stipitate utricle of the perianth tube and the chromosome number of 2n=12. While *A. grandiflora* arranged to sect. *Hexandrae* was divided into other close clade. The section *Gymnolous*, subsection *Hexandrae*, ser. *Hexandrae* consisting of *A. tentaculata, A. anguicida, A. ringens, A. littoralis* and *A. gigantea* were all included in a single clade at bootstrap value 83%, which was

consistented with morphological features which are hexamerous stemium and chromosome number of 2n=14.

The results from phylogenetic analysis were well correlated to morphological criteria, habitats and chromosome numbers. Our results agree well with the previously published *mat*K phylogenetic tree (Murata, Ohi et al. 2001, Ohi-Toma, Sugawara et al. 2006) and *trn*L-*trn*F phylogenetic tree (Neinhuis, Wanke et al. 2004). In addition, this is the first study of *A. pothieri*.

The nucleotide variations of the three *Aristolochia* plants used as Krai-Krue herb are shown in Table 9. The nucleotide sequences of three DNA regions which are *rbc*L, *mat*K and ITS, could be used for identification of these plants at species level except the nucleotide sequences of *trn*H-*psb*A spacer. From phylogenetic analysis related to the habitat, *A. pothieri*, *A. pierrei* and *A. tagala* which are botanical origin of Krai-Krue herbs are Asia species. The results also relate to traditional uses of these herbs, *A. pothieri* and *A. pierrei* is used in Thailand, *A. tagala* is widely used in many countries such as Thailand, India, Philippines and Maleysia (Sathornviriyapong, Picheansoonthon et al. 2007, Heinrich, Chan et al. 2009).

3.5 Conclusion

The DNA sequencing technique could be used for the establishment of DNA database and the identification of organisms. The *rbc*L, *mat*K, ITS and *trn*H-*psb*A contained diagnostic polymorphic sites and could be used to distinguish the *Aristolochia* plants from other species or its adulterants/substitutes. The *mat*K sequence was chosen for phylogenetic analysis. The ITS region appeared to be suitable DNA regions for the species identification and provided the simple and rapid tools for further study. The DNA information will be helpful for forensic investigation and safety control by the herbal industries and regulatory authorities and also for new plant species discovery.