

**Molecular Authentication of Baihuasheshecao  
and Icefish**

**YU, Jing**

A Thesis Submitted in Partial Fulfillment  
of the Requirements for the Degree of  
Master of Philosophy  
in  
Biology

The Chinese University of Hong Kong  
November 2011

Thesis/ Assessment Committee

Professor CHU, Ka Hou (Chair)

Professor CHIU, Siu Wai (Thesis Supervisor)

Professor SHAW, Pang Chui (Committee Member)

Dr. MARK, Simmons (External Examiner)

## Abstract

Species identification of food or traditional Chinese medicine is very important for prevention of adulterants in food and herb commodities, protection of health and economic benefits for consumers and strengthening the force of supervision by government. However, deficiency of taxonomic knowledge, morphological changes induced by processing a raw material and limitation of technological skills hampered the accuracy of species identification. Recently, one molecular method, forensically informative nucleotide sequencing (FINS) technology, has been successfully used as an alternative technology for more accurate, easy and low cost species identification.

Baihuasheshecao (*Hedyotis diffusa* Willd.) is highly reputed as an herb for cancer treatment. However, many adulterants are derived from related species, such as *H. corymbosa* (L.) Lam, *H. tenelliflora* Blume and *H. pinifolia* Wall. ex G. Don. In this study, Baihuasheshecao was identified by molecular techniques using FINS approach based on nuclear internal transcribed spacer (ITS: ITS1-5.8S-ITS2), plastid *trnH-psbA* intergenic spacer (*trnH-psbA* region), plastid *trnL* intron and the *trnL-F* intergenic spacer (*trnL-F* region). A total of 41 plant samples of *Hedyotis* L. species and 8 samples of Baihuasheshecao commodities were collected from Hong Kong, Macao, Taiwan, mainland China and USA. FINS analysis revealed that three samples of Baihuasheshecao commodities purchased from Hong Kong and USA were adulterants derived from *H. corymbosa*. Among the three DNA regions analyzed, the *trnL-F* region was the most efficient DNA marker for discriminating Baihuasheshecao from the adulterants and other closely related species.



Icefishes are recognized as traditional delicious food in China. In the market, they are usually sold as fresh, dried or pickled commodities, the latter formats lacking sufficient morphological characters for identification. In this study, the FINS approach with 16S rDNA and 12S rDNA regions was used to identify 18 samples purchased from Hong Kong and two samples from Taiwan. Based on sequence distance matrices and the sequence-based phylogenetic trees, the analyses revealed that eight specimens fell into the clade of *Neosalanx*, two were *Hemisalanx brachyrostralis*, and seven were *Protosalanx chinensis*. These belong to family Salangidae of order Osmeriformes, in contrast to two pickled samples from family Engraulidae in the order Clupeiformes. The results indicate the culture of flexible adaptation of similar morpho-forms as a food/herbal product among Chinese.

The molecular study was also used to unveil the phylogenetic relationship in *Hedyotis*. The classification of *Hedyotis* in *Flora Republicae Popularis Sinicae* has placed much emphasis on using the type of fruit dehiscence as a key character in subdividing the genus into different sections. However, the capsule characters can be variable and subject to adaptations to dispersal strategies. In this study, one nuclear (ITS1, 5.8s and ITS2 rDNA) and five plastid regions (*trnL* intron, *trnL-F* intergenic spacer and *trnH-psbA* intergenic spacer, *rbcL*, *matK* and *rps16*) were used to infer phylogenetic relationship of 87 samples representing 20 *Hedyotis* species from Guangzhou, Hong Kong, Macao and Taiwan. The result reveals that section delimitation based solely on the type of capsule dehiscence is both confusing and unsafe.



## 摘要

食品與傳統中藥材的物種鑒定對於防止商品中出現偽劣品、替代品，保護消費者健康和經濟利益及幫助政府部門加強監督管理是十分重要的。然而由於普通消費者和監管者們都缺乏一定的分類學知識，外加食物或藥材經過處理後的形態發生了一定的改變和缺少合適方便的鑒定技術都限制了準確的物種鑒定。目前應用於食品和藥材的物種鑒定 DNA 技術包括 FINS 法醫資訊核苷酸序列技術 (FINS 技術)，這種技術一直被認為是一種十分準確，方便及廉價的物種分子鑒定技術。本研究旨在利用該技術對於市場中常見的一種中藥材及食物進行物種鑒定。

中藥白花蛇舌草 (*Hedyotis diffusa*) 因可以用於癌症治療而備受關注。近年來在市場上發現了不少白花蛇舌草的偽品，分別來自它的近源種傘房花耳草 (*H. corymbosa*)、纖花耳草 (*H. tenelliflora*)、松葉耳草 (*H. pinifolia*)。為了保證臨床用藥的療效穩定及藥品的品質安全，非常有必要尋找一種快速、準確的方法用於白花蛇舌草的鑒定。本次實驗以 FINS 分子鑒定方法通過三種不同的 DNA 序列，rDNA - ITS 區序列 (ITS1、5.8S 和 ITS2 全長序列)，葉綠體基因組 (cpDNA) *trnL* 內含子及 *trnL*-F 間隔區序列 (*trnL* intron plus *trnL*-F intergenic spacer region) *trnH-psbA* 序列 (*trnH-psbA* intergenic spacer) 來鑒定白花蛇舌草的真偽。本實驗的 41 個耳草屬的樣品和 8 個白花蛇舌草採集且購買自香港、澳門、臺灣、中國大陸及美國五個地區。經由 FINS 分子鑒定方法揭示，購買自美國和香港的三個藥品的樣本是來自傘房花耳草的偽品，而其

他 5 個藥品均為正品。同時研究發現 *trnL-F* 序列是三個 DNA 序列中，用於鑒別白花蛇舌草及其近源品種效率最高的 DNA 序列。

銀魚在中國一直被認為傳統意義上的美食。在市場中，一般有新鮮的銀魚，銀魚幹及醃制的銀魚出售，但是由於受過處理的醃制銀魚缺少物種鑒定的基本特徵而給鑒定帶來一定的麻煩。本研究旨在採用 FINS 技術，利用線粒體 16S 和 12S 核糖體 DNA 序列鑒定來自 18 個香港市場和兩個臺灣市場的銀魚市售樣本。根據遺傳距離矩陣及鄰位元相連法 (NJ) 構建的系統樹狀圖的結果，分析表明有 20 個樣品中有 8 個樣品屬於新銀魚屬，七個樣品是大銀魚，兩個樣品是短吻間銀魚，其餘兩個醃制的銀魚樣品屬於鯡形目的鯷科。同時研究表明，在中國的傳統食品及中藥材中，相似形態的不同物種的命名具有一定的相容性。

耳草屬植物的基因序列除了用於物種鑒定外，還能用來分析耳草屬植物間的親源關係。在中國植物志中，耳草屬的屬內分組以蒴果的開裂方式為主要依據。但是，蒴果的形態十分多樣，而且其形態特徵可以發生一定改變以適應特定環境下的種子擴散方式。本研究分析了六段基因序列：1. 一段核基因組的 ITS-1, 5.8S, ITS-2 序列；2. 五段葉綠粒基因組序列：*trnL* 內含子及 *trnL-trnF* 間隔區序列，*trnH-psbA* 間隔序列，*rbcL*，*matK* 和 *rps16*。用於研究的 20 個物種分別採集自廣州，香港，澳門和臺灣各地的 87 個樣品。本研究結果表明了僅僅依賴蒴果的開裂方式來進行屬的分組不僅會造成分類的混亂而且是不可靠的。此外，該研究也分析了本次採集到的耳草屬物種與其他鈕扣草族植物間的親緣關係。



## Acknowledgements

First of all, my deepest and sincere gratitude goes first to my supervisors, Professor Paul P. H. But and Professor S.W. Chiu, for their constant support and guidance throughout the research. I would also like to expand my gratitude to Professors K. H. Chu, Mark P. Simmons, and P. C. Shaw for being my thesis committee members. I thank Mr. T. K. Woo and Prof. P. P. H. But for their help in collection of Icefishes, Baihuasheshecao, and *Hedyotis* samples. The support and assistant from Hong Kong Herbarium and Division of Nature Studies and Conservation in Department of Gardens and Green Areas in Macao is also very helpful. I am grateful to Professor R.-J. Wang for his work on *Hedyotis* species identification. Next, I thank the Department of Biology of the Chinese University of Hong Kong, which provided me with abundant facilities and resources. For the staff and colleagues in the Department of Biology and Institute of Chinese Medicine, I am beholden to Mr. T. K. Woo, Mr. K. L. Wong, Ms. Pamela W. H. Chan, Ms. H. Lam, Mr. M. Li, Ms. Y. L. Wang, Mr. Y. M. Chan, Mr. H. Zhang etc. for their generous help and support. I can never thank Mr. Joe K. H. Ling enough for teaching me molecular technologies and analysis methods. Last, but not least, I express my gratitude to my family and friends for their love, encouragement and support during my research.



## Contents

<b>Abstract</b> .....	<b>I</b>
<b>摘要</b> .....	<b>III</b>
<b>Acknowledgements</b> .....	<b>V</b>
<b>Contents</b> .....	<b>VI</b>
<b>List of Figures</b> .....	<b>VIII</b>
<b>List of Tables</b> .....	<b>X</b>
<b>Abbreviations and Symbols</b> .....	<b>XII</b>
<b>CHAPTER 1 Introduction</b> .....	<b>1</b>
1.1 Phylogenetic study of <i>Hedyotis</i> .....	2
1.1.1 Rubiaceae .....	2
1.1.2 Controversial taxonomic issues.....	9
1.2 Traditional Chinese medicine (TCM).....	19
1.2.1 Introduction .....	19
1.2.2 Baihuasheshecao .....	20
1.2.3 Authentication of Baihuasheshecao.....	23
1.3 Icefishes in the Hong Kong market.....	23
1.3.1 Introduction .....	23
1.4 Molecular approach.....	25
1.4.1 Introduction to molecular phylogeny .....	25
1.4.2 FINS (Forensically Informative Nucleotide Sequencing) for species identification .....	27
1.4.3 DNA sequence markers.....	28
1.5 Objectives.....	38
<b>CHAPTER 2 MATERIALS AND METHODOLOGY</b> .....	<b>41</b>
2.1 Materials.....	42
2.2 DNA extraction .....	50
2.3 Polymerase chain reaction (PCR) method.....	51
2.4 Gel electrophoresis.....	54
2.5 PCR production purification .....	54
2.6 Ligation and transformation and transformation of PCR product.....	56
2.7 DNA sequencing and sequence analyses.....	58
<b>CHAPTER 3 USING FORENSICALLY INFORMATIVE NUCLEOTIDE SEQUENCING (FINS) TECHNOLOGY FOR SPECIES IDENTIFICATION</b> 64	
3.1 Authentication of Baihuasheshecao by FINS Analysis .....	65
3.1.1 Authentication using FINS technology .....	65
3.1.2 Relative effectiveness of DNA regions for FINS analysis .....	70
3.1.3 Phylogenetic interpretation.....	72
3.2 Authentication of Salangids (Icefishes) by FINS Analysis .....	74
3.2.1 Analysis based on mitochondrial ribosome DNA region .....	74
3.2.2 Analysis based on mitochondrial 16S rRNA.....	77

3.2.3 Analysis based on combined regions.....	79
3.2.4 Phylogenetic analysis .....	81
3.2.5 Discussion .....	85
3.3 Conclusions .....	88
<b>CHAPTER 4 PHYLOGENTIC STUDY OF HEDYOTIS IN CHINA AND THEIR POSITION IN SPERMACOCEAE.....</b>	<b>89</b>
4.1 Phylogentic study of <i>Hedyotis</i> species in Chinese .....	90
4.1.1 Nuclear ITS region .....	90
4.1.2 Plastid <i>trnL</i> intron and <i>trnL</i> -F intergenic spacer region .....	94
4.1.3 Plastid <i>trnH-psbA</i> intergenic spacer region.....	98
4.1.4 Plastid <i>rbcL</i> region .....	102
4.1.5 Plastid <i>matK</i> region .....	106
4.1.7 Combined analysis .....	114
4.2 The phylogenetic position of <i>Hedyotis</i> (species in China) in the tribe of Spermaceae s. l.....	121
4.2.1 Plastid <i>trnL</i> -F intergenic spacer region .....	121
4.2.2 Plastid <i>rbcL</i> region .....	133
4.2.3 Plastid <i>rps16</i> region.....	141
4.3 Discussion .....	153
4.3.1 Comparison of phylogenetic utility of the six DNA regions .....	153
4.3.2 Diplophragma section .....	154
4.3.3 <i>Hedyotis</i> , <i>Dimetia</i> , <i>Euoldendandia</i> and <i>Gonotheca</i> sections.....	156
4.3.4 The position of <i>Hedyotis</i> (species in China) in Spermaceae.....	158
4.4 Conclusions .....	160
<b>REFERENCES.....</b>	<b>161</b>
<b>APPENDIX .....</b>	<b>173</b>



## List of Figures

Figure 1- 1. The species richness of Rubiaceae in TDWG (Taxonomic Databases Working Group) level 3 regions.....	3
Figures 1- 2. Pictures of representative species in five sections ( <i>Hedyotis</i> , <i>Diplophragma</i> , <i>Dimetia</i> , <i>Euoldenlandia</i> and <i>Gonotheca</i> ) of <i>Hedyotis</i> . ....	17
Figure 1- 3. Picture of Baihuasheshecao sample.....	20
Figure 1- 4. Pictures of the four <i>Hedyotis</i> species. ....	22
Figure 1- 5. General appearance of salangids.....	25
Figure 1- 6. Schematic diagram of plastid genome. ....	29
Figure 1- 7. Diagram of <i>rbcL</i> gene in plastid DNA.....	30
Figure 1- 8. Location of <i>matK</i> in plastid DNA. ....	31
Figure 1- 9. Location of the <i>rps16</i> region in plastid DNA. ....	32
Figure 1- 10. Diagram of <i>trnL-F</i> region.....	33
Figure 1- 11. Diagram of a single unit of nrDNA.....	35
Figure 1- 12. Diagram of the Piscine mitochondrial genome.....	37
Figure 3- 1. Phylogenetic tree based on nuclear ITS region constructed using maximum parsimony method.....	67
Figure 3- 2. Phylogenetic tree based on plastid <i>trnL-F</i> region constructed using maximum parsimony method. ....	68
Figure 3- 3. Phylogenetic tree based on plastid <i>trnH-psbA</i> region constructed using maximum parsimony method. ....	69
Figure 3- 4. Phylogenetic tree based on 12s rRNA sequences constructed using maximum parsimony method.. ....	82
Figure 3- 5. Phylogenetic tree based on 16s rRNA sequences constructed using maximum parsimony method. ....	83
Figure 3- 6. Phylogenetic tree based on combined 12s rRNA and 16s rRNA sequences constructed using maximum parsimony method from 20 fish samples.....	84
Figure 4- 1. Jackknife 50% majority-rule strict consensus tree of the <i>Hedyotis</i> based on ITS region.. ....	92
Figure 4- 2. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of the <i>Hedyotis</i> based on ITS region.....	93
Figure 4- 3. Jackknife 50% majority-rule strict consensus tree of the <i>Hedyotis</i> based on <i>trnL</i> intron and <i>trnL-F</i> intergenic spacer region.....	96
Figure 4- 4. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of <i>trnL</i> intron and <i>trnL-F</i> intergenic spacer.....	97
Figure 4- 5. Jackknife 50% majority-rule strict consensus tree of the <i>Hedyotis</i> based on <i>trnH-psbA</i> intergenic spacer. ....	100
Figure 4- 6. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of <i>trnH-psbA</i> intergenic spacer. ....	101
Figure 4- 7. Jackknife 50% majority-rule strict consensus tree of the <i>Hedyotis</i> based on <i>rbcL</i> region.....	104



Figure 4- 8. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of <i>rbcL</i> region. ....	105
Figure 4- 9. Jackknife 50% majority-rule strict consensus tree of the <i>Hedyotis</i> based <i>matK</i> region. ....	108
Figure 4- 10. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of <i>matK</i> region.....	109
Figure 4- 11. Jackknife 50% majority-rule strict consensus tree of the <i>Hedyotis</i> based on <i>rps16</i> region. ....	112
Figure 4- 12. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of <i>rps16</i> region.....	113
Figure 4- 13. Jackknife 50% majority-rule strict consensus tree of the <i>Hedyotis</i> based on the combined datasets.....	118
Figure 4- 14. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of combined region.. ....	119
Figure 4- 15. Jackknife 50% majority-rule strict consensus tree with 1000 replicates based on the data of <i>trnL-F</i> intergenic spacer region.....	126
Figure 4- 16. Part one of the magnified Figure 4-15.....	127
Figure 4- 17. Part two of the magnified Figure 4-15. ....	128
Figure 4- 18. Part three of the magnified Figure 4-15.....	129
Figure 4- 19. Jackknife 50% majority-rule strict consensus tree based on the data of <i>trnL-F</i> intergenic spacer region. ....	130
Figure 4- 20. Part three of the magnified Figure 4-19.....	131
Figure 4- 21. Part three of the magnified Figure 4-15.....	132
Figure 4- 22. Jackknife 50% majority-rule strict consensus tree with 1000 replicates based on the data of <i>rbcL</i> region. ....	136
Figure 4- 23. Part one of the magnified Figure 4-22.....	137
Figure 4- 24. Part two of the magnified Figure 4-22. ....	138
Figure 4- 25. Part three of the magnified Figure 4-22.....	139
Figure 4- 26. Jackknife 50% majority-rule strict consensus tree based on the data of <i>rbcL</i> region.....	140
Figure 4- 27. Jackknife 50% majority-rule strict consensus tree with 1000 replicates based on the data of <i>rps16</i> region.....	146
Figure 4- 28. Part one of the magnified Figure 4-27.....	147
Figure 4- 29. Part two of the magnified Figure 4-27. ....	148
Figure 4- 30. Part three of the magnified Figure 4-27.....	149
Figure 4- 31. Jackknife 50% majority-rule strict consensus tree based on the data of <i>rps16</i> region.....	150
Figure 4- 32. Part three of the magnified Figure 4-31.....	151
Figure 4- 33. Part three of the magnified Figure 4-31.....	152



## List of Tables

Table 1- 1. Morphological comparison of the four <i>Hedyotis</i> species. ....	22
Table 2- 1. List of herb samples used in this study is shown below.....	43
Table 2- 2. List of field samples used in this study.....	44
Table 2- 3. Codes for sample of icefishes are listed below. ....	49
Table 2- 4. List of reagents and equipments used in modified CTAB method.....	51
Table 2- 5. List of amplified regions, primer sequences, amplified size and their sources. ....	53
Table 2- 6. List of reagents and equipment used for LBA plate and LB were prepared during the ligation and transformation.....	57
Table 2- 7 List of ligation reaction system used in this study.....	57
Table 2- 8. List of accession number of Baihuasheshecao samples is shown below. ....	59
Table 2- 9. List of accession number of field samples.....	59
Table 2- 10. Accession No. of 12S and 16S regions of icefishes. ....	62
Table 3- 1. Estimates of evolutionary divergence of the 12s rRNA sequences. ....	76
Table 3- 2. Estimates of evolutionary divergence of 16s rRNA sequences.....	78
Table 3- 3. Estimates of evolutionary divergence of the 12s rRNA and 16s rRNA sequences.....	80
Table 4- 1. Total taxa (TT), total characters (TC), constant characters (CC), parsimony- uninformative variable character (PUIC), parsimony-informative characters (PIC), tree length (TL), consistency index (CI), homoplasy index (HI), CI excluding uninformative characters (CIE), HI excluding uninformative characters (HIE), retention index (RI), rescaled consistency index (RC) for different datasets. ....	120
Table 4- 2. List of sequences downloaded from GenBank of the National Center for Biotechnology Information (NCBI). ....	121
Table 4- 3. List of sequences downloaded from Genbank of the National Center for Biotechnology Information (NCBI). ....	133
Table 4- 4. List of sequences downloaded from Genbank of the National Center for Biotechnology Information (NCBI). ....	141
Table 6- 1 Sequences alignment matrix of herb samples based on ITS region. ....	174
Table 6- 2 Sequences alignment matrix of herb samples based on <i>trnL</i> intron and <i>trnL</i> -F intergenic spacer region.....	182
Table 6- 3 Sequences alignment matrix of herb samples based on <i>trnH-psbA</i> intergenic spacer region.....	193
Table 6- 4 Sequences alignment matrix of icefish samples based on 12s rRNA.....	198
Table 6- 5 Sequences alignment matrix of icefish samples based on 16s rRNA.....	200
Table 6- 6 Sequences alignment matrix of icefish samples based on combined datasets. ....	204
Table 6- 7 Sequences alignment matrix of <i>Hedyotis</i> based on ITS region. ....	208
Table 6- 8 Sequences alignment matrix of <i>Hedyotis</i> based on <i>trnL</i> intron and <i>trnL</i> -F	

intergenic spacer region.....	220
Table 6- 9 Sequences alignment matrix of <i>Hedyotis</i> based on <i>trnH-psbA</i> intergenic spacer region.....	235
Table 6- 10 Sequences alignment matrix of <i>Hedyotis</i> based on <i>rbcL</i> region.....	245
Table 6- 11 Sequences alignment matrix of <i>Hedyotis</i> based on <i>matK</i> region.....	257
Table 6- 12 Sequences alignment matrix of <i>Hedyotis</i> based on <i>rps16</i> region.....	273
Table 6- 13 Sequences alignment matrix of <i>Hedyotis</i> based on combined datasets. ....	285
Table 6- 14 Sequences alignment matrix of Spermacoceae based on <i>trnL-F</i> intergenic spacer region.....	357
Table 6- 15 Sequences alignment matrix of Spermacoceae based on <i>rbcL</i> region. ....	382
Table 6- 16 Sequences alignment matrix of Spermacoceae based on <i>rps16</i> region.....	398



## Abbreviations and Symbols

12S/12S rRNA	mitochondrial 12S ribosomal RNA
16S/16S rRNA	mitochondrial 16S ribosomal RNA
AFCD	Agriculture, Fisheries and Conservation Department
BIG	Beijing genomics of institute
BLAST	basic local alignment search tool
bp	basepair
CH <sub>3</sub> COONa	sodium acetate
CITES	Convention on International Trade in Endangered Species of Wild Fauna and Flora
cm	centimeter
CTAB	cetyltriethylammonium bromide
CUHK	The Chinese University of Hong Kong
dd water	distilled-deionized water
DHA	docosahexaenoic acid
DNA	deoxyribonucleic acid
dNTP	2'-deoxynucleoside 5'-triphosphate
dNTPs	deoxytriphosphates (A, T, G, C)
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediamine tetraacetate
FISH-BOL	Fish Barcode of Life Initiative
g	gram
g force	gravity force
HCl	hydrochloric acid
HK	Hong Kong
Hong Kong SAR	Hong Kon Special Administrative Region
HPLC	high performance liquid chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
ITS	internal transcribed spacer
ITS-1	internal transcribed spacer-1
ITS-2	internal transcribed spacer-2
IUCN	International Union for the Conservation of Nature and Natural Resources
kb	kilobase pair
KCl	potassium chloride
LB	luria broth
LBA	LB with ampicillin
M	molar
<i>matK</i>	maturaseK gene
mg	milligram
MgCl <sub>2</sub>	magnesium chloride
MilliQ water	millipore filtered water

min	minute(s)
ml	milliliter
mM	millimolar
mRNA	messenger RNA
mtDNA	mitochondrial DNA
N <sub>2</sub>	nitrogen
NaCl	sodium chloride
NaOAc	sodium acetate
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
NCCAM	National Center for Complementary and Alternative Medicine
No.	number
nrDNA	nuclear ribosomal DNA
°C	degree Celsius
PCR	polymerase chain reaction
PVP	polyvinylpyrrolidone
<i>rbcL</i>	large subunit of the ribulose-bisphosphate carboxylase gene
rDNA	ribosomal DNA
Ref.	reference
RNA	ribonucleic acid
RNase	ribonuclease
<i>rps16</i>	ribosomal protein 16 small subunit gene
rRNA	ribosomal RNA
SDS	sodium dodecyl sulfate
sec	second(s)
TAE	tris-acetate-EDTA
TCM	traditional Chinese medicines
TL	Total length
TLC	thin layer chromatography
Tris	tris [hydroxymethyl] aminomethane
Tris-HCl	tris-hydrochloride
tRNA	transfer RNA
<i>trnH-psbA</i>	plastid <i>trnH-psbA</i> intergenic spacer
<i>trnL</i>	plastid tRNA-Leu ( <i>trnL</i> ) gene
<i>trnL-F</i>	plastid tRNA-Leu ( <i>trnL</i> ) gene and tRNA-Phe ( <i>trnF</i> ) gene intergenic spacer region
U	unit
UV	ultraviolet
w/v	weight per volume
xg	times gravitational force
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
μl	microliter
μM	micromolar

# **CHAPTER 1**

## **Introduction**



## **1.1 Phylogenetic study of *Hedyotis***

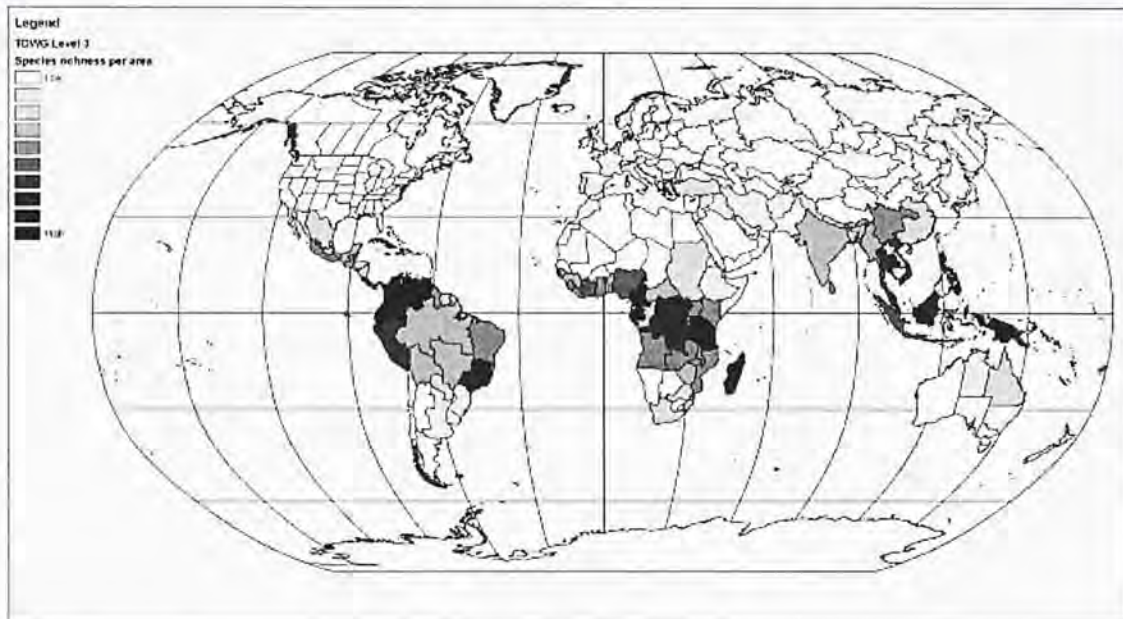
### **1.1.1 Rubiaceae**

#### **1.1.1.1 Introduction**

Family Rubiaceae, which is commonly called the coffee family, contains more than 13,000 species in 611 genera (Goevarts, Ruhsam et al. 2006) and is the fourth largest angiosperm family in the world. As one of the members of Gentianales, Rubiaceae share the common features with other families in this order, like the presence of colleters and lack of internal phloem. Rubiaceae is also diagnosed by opposite, entire leaves with stipules, inferior ovaries, presence of nuclear endosperm, and often possession of alkaloids (Bremer 1996b; Davis, Govaerts et al. 2009). According to recent advances, Rubiaceae have three subfamilies, Cinchonoideae, Ixoroideae and Rubioideae. The most easily recognized and described subfamily is Rubioideae (Bremer 1996b), which contains the majority of the species. This subfamily is identified by the combination of herbaceous, valvate corolla lobes and the presence of raphides.

The Rubiaceae is less frequent and less diverse in temperate regions relative to subtropical and tropical regions (Davis, Govaerts et al. 2009). Some species can also be found in subpolar regions of the Arctic and Antarctic. **Figure 1- 1** shows the species richness of Rubiaceae in the world.

**Figure 1- 1. The species richness of Rubiaceae in TDWG (Taxonomic Databases Working Group) level 3 regions.** Rubiaceae's relative species richness is rescaled by the size of that region using a power-law species area relationship and standardized to 10,000 km<sup>2</sup> (Davis, Govaerts et al. 2009).



In terms of biological and morphological diversities of species in Rubiaceae, they vary in different life forms and reproductive characteristics. The life forms of species range from delicate herb, robust herb, epiphyte, climber, shrub to tree. Their fruit types range from dry capsules for wind dispersion seed, dry dehiscent and indehiscent mericarps, and berries or drupes for animal dispersal (Bremer, Andreasen et al. 1995; Bremer 1996a).

Rubiaceae are also well known as an economically important flowering family, for its famous species like *Coffea canephora* and *Coffea arabica* for coffee production and the bark of *Cinchona* for a source of alkaloids. Ipecac derived from the genus *Psychotria* is useful for induce vomiting. There are some genera which are used as ornamentals (e.g., *Gardenia*, *Hamelia*, *Pentas*, *Randia*, *Rondeletia*, *Serissa* and



*Hedyotis*).

### **1.1.1.2 Taxonomic history**

Rubiaceae were first published by Jussieu in 1789 and the classification of subfamilies and tribes in Rubiaceae remains problematic. Schumann in 1891 divided Rubiaceae into two subfamilies, Cinchonoideae and Coffeioideae, based on the character of ovule number per locule. Cinchonoideae have many seeds in each carpel while Coffeioideae only have one seed in each carpel. Bremekamp in 1952 rejected this classification and emphasized more characters for subfamily classification. Based on the features of testa structure, occurrence of albumen in seeds, raphides, and secondary pollen presentation, Bremekamp (1954, 1966) suggested eight subfamilies of Rubiaceae. Meanwhile, Verdcourt (1958) accepted only three of the eight subfamilies (Cinchonoideae, Rubioideae and Guttardoideae). Verdcourt also pointed out that the pollen-presentation mechanism could not be used as a subfamily feature, and he emphasized the importance of external hair type for subfamily classification. Later, Robbrecht (1988, 1994a, b) presented the new circumscription for the subfamily classification of Rubiaceae. Based on the abovementioned features and updated advances in researches of Rubiaceae, Robbrecht suggested the following four subfamilies: Antirheoideae, Cinchonoideae, Ixoroideae and Rubioideae (Robbrecht 1988; Robbrecht 1994a; Robbrecht 1994b).

In 1891, Schumann segregated the taxa of 'Rubioideae' into two subfamilies, Coffeioideae and Cinchonoideae. Actually, the concept of Rubioideae as a subfamily

originated from Bremekamp in 1952, and developed in 1954 and 1956. According to his suggestion, the tribes in Rubiaceae with raphides should belong to a new subfamily, Rubioideae. Although several raphidiate tribes' positions were still under investigation, this circumscription was accepted by Verdcourt (1958) and Robbrecht (1994a, b). All three of these researchers [ Verdcourt (1958), Bremekamp (1966) and Robbrecht (1988, 1994b)] have suggested 12 core tribes in Rubioideae as Anthospermeae, Argostemmateae, Coccocypseleae, Coussareeae, Hamelieae, Hedyotideae, Morindeae, Paederieae, Psychotrieae, Rubieae, Schradereae, and Spermacoceae. However, the tribal description for Rubioideae varied among different authors. Verdcourt (1958) gathered the Anthospermeae, Spermacoceae, and Hedyotideae as a group for their herbaceous or subshrub habitat and dry fruit trait (including the tribe with solitary ovules), whereas four tribes (Psychotrieae, Morindeae, Coussareeae, and Schradereae) with woody habitat and erect solitary ovule were grouped. For the tribes not possessing these key characters, they remained vaguely as in Verdcourt's classification. In 1966, Bremekamp made two different series for those tribes, one series with uniovulate carpels, and other series with multiovulate carpels. Robbrecht in 1988 had further described the concepts for the two series. He suggested that combining the Hedyotideae, Ophirrhizeae, Hamelieae, Argostemmateae, Hamelieae and Schradereae as a group was based on the character of multiovulate carpels, and Psychotrieae, Triainolepideae, Lathraeocarpeae, Morindeae, Coussareeae, Paederieae, Anthospermeae, Theligoneae, Spermacoceae, and Rubieae as the other group were based on solitary ovules.



However, none of the above classifications were based on phylogenetic analyses, and it is hard to evaluate which one provides the most accurate phylogenetic framework for the Rubiaceae and Rubioideae.

### **1.1.1.3 Molecular phylogenetic studies**

During recent years, both molecular and morphological data sets were applied to the phylogenetic construction of Rubiaceae (Bremer and Jansen 1991; Bremer 1992; Bremer and Eriksson 1992; Bremer and Struwe 1992; Manen, Natali et al. 1994; Bremer, Andreasen et al. 1995; Natali, Manen et al. 1995; Bremer 1996a; Bremer 1996b; Andersson and Rova 1999; Andreasen, Baldwin et al. 1999; Bremer, Jansen et al. 1999; Andreasen and Bremer 2000; Bremer and Manen 2000; Rova, Delprete et al. 2002; Motley, Wurdack et al. 2005; Robbrecht and Manen 2006; Karehed and Bremer 2007; Karehed, Groeninckx et al. 2008; Rydin, Razafimandimbisan et al. 2008; Groeninckx, Dessein et al. 2009; Lens, Groeninckx et al. 2009; Mouly, Razafimandimbison et al. 2009; Neupane, Dessin et al. 2009; Rydin, Razafimandimbison et al. 2009). With the increasing technological innovations, molecular data could be obtained more rapidly and easily, and thus our understanding of phylogenetic relationships within Rubiaceae has increased tremendously during the past decades. Many studies have proved the utility of molecular data in the analyses of intrafamilial, tribal, generic and interspecific relationships. Moreover, the molecular data could also provide phylogenetic construction with new perspectives. The classification of Rubiaceae is still being updated by newly available and relevant

data sets. The molecular phylogenetic studies for subfamily of Rubiaceae classification are summarized below.

Molecular phylogenetic study of Rubiaceae started in 1991. In the first study, 33 species in 33 genera from 18 tribes in Rubiaceae were studied based on data from restriction site mapping of plastid DNA (Bremer and Jansen 1991). The result revealed that Cinchonoideae were paraphyletic, and other three subfamilies were monophyletic. This result was inconsistent with the studies of Bremekamp (1952, 1954, 1966), Verdcourt (1958) and Robbrecht (1988, 1994a, b). In 1995, Bremer et al. sampled *rbcL* sequences of 56 species in 49 genera representing 23 tribes to investigate subfamily relationships. The phylogenetic result supported three subfamilies, Cinchonoideae (Cinchoneae, Chiococceae *s.l.*, Guettardeae, Hamelieae, Hillieae, Naucleae, and Rondeletieae), Ixoroideae (Coffeae, Gardenieae, Pavetteae and Vanguerieae), and Rubioideae (Rubiaceae, Anthospermeae, Hedyotideae, Morindeae, Ophiorrhizeae, Psychotrieae, and Theligoneae), with regard to the classification of Robbrecht (1988). However, there was still unclear indication for the position of some genera and tribes. One year later, Bremer (1996b) found that Rubiaceae was monophyletic, and the three-subfamily classification was supported by *rbcL* sequences. Bremer et al. (1999), based on *rbcL*, *ndhF* and morphological data, further confirmed the three-subfamily classification, and in the study of Andersson (1999), the data set from *rps16* region also confirmed the three-subfamily classification. Rova et al. (2002) published his phylogenetic study based on *trnL-F* sequences for 154 Rubiaceae species. The study found that there was no support for



subfamily Antirheoideae (Rova, Delprete et al. 2002). This had confirmed the conclusions of Bremer (1995, 1996b, 1999) and Andersson (1999). The above studies provided strong support for the three-subfamily classification, but the basalmost nodes for each subfamilies and the position of some tribal and genera were still uncertain or unsolved. As the largest subfamily with complicated and unclear tribal relationships, Rubioideae have attracted much attention. In earlier publications (Bremer & Jansen 1991; Bremer & Struwe, 1992, Bremer et al. 1995; Bremer, 1996a) of Rubiaceae, Rubioideae was monophyletic (except for excluding the tribe Hamelieae), but the deeper discussion of the tribal relationship in this subfamily was insufficient.

As mentioned above, the Rubioideae have 12 core tribes, which are Anthospermeae, Argostemmateae, Coccocypseleae, Coussareeae, Hamelieae, Hedyotideae, Morindeae, Paederieae, Psychotrieae, Rubieae, Schradereae, and Spermacoceae. In 1995, Natali suggested that the tribes of Hamelieae and Ophiorrhizeae should be excluded from Rubioideae. In the study of Bremer (1996b), Ophiorrhizeae, Pauridiantheae, and Knoxieae, which were not in the 12 core tribes, were included as a part of Rubioideae, and the study also asserted that Hamilieae should be transferred from Rubioideae to Cinchonoideae *s.s.*. In 1999, the *rps16* region was applied by Andersson and Rova to analyse 50 genera to investigate the phylogeny of Rubioideae (including tribes Anthospermeae, Coccocypseleae, Cruckshanksieae, Coussareeae, Gaertnereae, Hedyotideae, Knoxieae, Morindeae, Ophiorrhizeae, Paederieae, Pauridiantheae, Perameae, Psychotrieae, Rubieae, Spermacoceae, Theligoneae, and

Urophyllaeae), and the result confirmed the main groups suggested by Bremer (2000) and provided a more informative classification. In 2000, Bremer and Manen chose *rps16*, *rbcL* and *atpB-rbcL* spacer to infer the phylogenetic relationship of Rubioideae. The result showed that the basal clades of Rubioideae are Ophiorrhizeae, Urophyllaeae, Lasiantheae and Coussareese), and the subfamilies consisted of two major alliances (Psychotrieae alliance and Spermaceae alliance). The basal position of the four tribes was also confirmed by later studies (Robbrecht and Manen 2006; Rydin, Razafimandimbisan et al. 2008).

Although tribal classification of Rubioideae was well studied, there are still many problematic genera and species (Bremer and Eriksson 2009). Thus, in order to have a better understanding of Rubioideae, more sequences should be added.

## **1.1.2 Controversial taxonomic issues**

### **1.1.2.1 Spermaceae**

#### **1.1.2.1.1 Introduction**

The tribe of Spermaceae was recognized by Bentham and Hooker in 1873 as one of the 25 tribes in Rubiaceae based on the character of uniovulate locules. In 1891, Schumann published a similar system that included 21 tribes of Spermaceae. In his study, Spermaceae bore one-ovulate ovaries while Hedyotideae found with many-ovuled ovaries, and they were separated to different subfamilies. Verdcourt (1952) suggested a more closely relationship between the herbaceous tribes (e.g.



Hedyotideae and Spermaceae, etc.). In the same year, Bremekamp stated several important features for the subfamilies delimitation, and he rejected the character of ovule number per ovary. Although Hedyotideae possess numerous ovules per ovary, the presence of raphides in their tissues and valvate aestivation has led to their inclusion in the subfamily Rubioideae with Spermaceae. In a later study, Verdcourt (1958) confirmed Spermaceae and Hedyotideae as belonging to Rubioideae subfamily based on the characters of raphides present and albuminous seeds. Tribe of Spermaceae was delimited to include those species with raphides in tissue, seeds with adaxial groove and pluriaperturate pollen based on Robbrecht's (1988, 1993) findings.

However, based on the molecular studies (*rbcL* region) of Bremer (1995, 1996b) and Natali et al. (1995), Spermaceae were found to be deeply nested within Hedyotideae, making Hedyotideae paraphyletic. Bremer (1996b) proposed the solution of combining the two tribes. The united tribe was named as Spermaceae following the rules of priority (Greuter, McNeill et al. 2000). In 2000, Bremer and Manen described a much wide circumscription for Spermaceae *s.l.* which was comprised of Spermaceae *s.s.*, Hedyotideae, Manettieae, Knoxieae, and Triainolepideae. However, Andersson and Rova in 1999 pointed out that the emendation for combination was premature, and the sampling so far provided insufficient resolution. In their opinion, to include several genera of Hedyotideae based on the more precise way to interpret the molecular data was a more convenient solution. In later studies, Dessien (2003), Robbrecht and Manen (2006), Karehed and

Bremer (2007) suggested a new emendation of Spermaceae *s.l.* that excluded Knoxieae based on their molecular data. Dessien in 2003 stated that the Spermaceae should include Spermaceae *s.s.*, Manettieae, and some genera of Hedyotideae. In subsequent studies, Karehed in 2008 and Groeninckx in 2009 confirmed this classification by molecular studies with specimens of Spermaceae *s.l.* Thus, Spermaceae *s.l.*, was recognized as consisting of 61 genera and 1235 species in the world, characterized by 4-merous flowers and fimbriate stipules (Karehed et al. 2008; Groeninckx et al. 2009).

Continuous changes for generic delimitation in Spermaceae *s.l.* have caused the complicated taxonomic histories for this tribe (Lens, Groeninckx et al. 2009). Also, the controversy for the infratribal relationship of Spermaceae *s.l.* is a frequent research the subject. As an essential herbaceous lineage in Rubiaceae, Spermaceae *s.l.* could be identified as the following three groups based on morphological characters: *Hedyotis-Oldenlandia* group, Spermaceae *s.s.* group and the third group composed by two American genera, *Bouvardia* Salisb. and *Manettia* Mutis ex L. Bremekamp 1952 (Groeninckx, Dessein et al. 2009). Among them, the *Hedyotis-Oldenlandia* group, as one of the most species-rich groups in Rubiaceae, was traditionally included in tribe Hedyotideae Cham. & Schltdl. ex DC. However, based on the recently molecular studies, this group was moved to tribe Spermaceae *s.l.* The phylogenetic study for *Hedyotis-Oldenlandia* group and its closely relatives has been an attractive subject for many years.



## 1.1.2.2 *Hedyotis* – *Oldenlandia* Group

### 1.1.2.2.1 Taxonomical introduction

The genus *Hedyotis* L. (including *Oldenlandia*, *Hedyotis* and *Houstonia*) in the broad sense is a large genus with 699 taxa (including both species and varieties; Wang, 2001). *Oldenlandia* is a worldwide genus, mainly distributed in the tropics and subtropics, and occasionally in temperate regions. The genus *Hedyotis* is mainly distributed at Asian and Pacific regions. *Houstonia* is limited to South American, Canada, USA and Mexico.

*Oldenlandia* was first described by Plumier in 1703 and taken up by Linnaeus in 1737, while *Hedyotis* was first introduced by Dassow and accepted by Linnaeus in 1743. In 1753, *Hedyotis* (including three Asian species, *H. auricularia* L., *H. fruticosa* L. and *H. herbacea* L.), *Oldenlandia* (including four species, *O. biflora* L., *O. corymbosa* L., *O. umbellata* L. and *O. uniflora* L.), and *Houstonia* (including *H. caerulea* L. and *H. purpurea* L.) were published by Linnaeus. In later studies, *Hedyotis* was given nomenclatural priority over the other two genera as a unit (Lamarch, 1792, Wight and Arnott, 1843, Torrey and Gray, 19841). However, this combination is frequently the subject of debate.

The histories of *Hedyotis* and *Oldenlandia* are more complicated. In 1791, Lamark first united the two genera with some hesitation, and *Hedyotis* was chosen as the enlarged genus name. Willdenow in 1797 disagreed with this opinion and retained their separation. In the next study, Smith in 1811 used the name of *Hedyotis* for the

united genus, and used the name of *Oldenlandia* to describe a still different genus. However, the name of *Oldenlandia* was reused for the original species as suggested by Chamisso and Schlechtendal in 1829. Wight and Arnott in 1834 combined some closely related genera into *Hedyotis*, with these genera assigned to the rank of section. However, Meisner insisted on the generic classification for Wight and Arnott's sections. Later, Betham and Hooker in 1873 confirmed the distinction of *Hedyotis* and *Oldenlandia*. In 1891, Schumann united *Hedyotis* and *Oldenlandia* with other closely related genera. But, later researchers preferred to follow the opinion of Betham and Hooker, and regarded *Hedyotis* and *Oldenlandia* as distinct genera. In 1952, Bremekamp, in his study on African *Oldenlandia* species, separated *Oldenlandia* from *Hedyotis*. Lewis in 1962, still treated *Oldenlandia*, *Hedyotis* and *Houstonia* as subgenera of *Hedyotis* (Lewis 1962). In a more recent study, Halford (1992) distinguished *Hedyotis* from *Oldenlandia* by the capsule dehiscence type based on his study of Australian species (Halford 1992).

Genus *Kadua* was first introduced by Chamisso and Schlechtendal (1829) with five species. Endlicher (1836-1841) suggested that *Kadua* could be a section of *Hedyotis* based on the identities of *kadua* with *Hedyotis*. This view was followed by Steudel (1840), who moved *Kadua* to *Hedyotis*. A few years later, Walpers (1843) confirmed the position of *Kadua* as one of the sections in *Hedyotis*. Based on the Polynesian species of *Hedyotis*, Fosberg in 1943 treated *Kadua* as a subgenus of it. However, later studies did not support this modification (Gray, 1860; Mann, 1867; Wawar, 1874; Schumann, 1981, Kuntza, 1981). In their taxonomic concepts, *Kadua* remained



as a genus.

Another genus, *Kohautia*, was considered as one of the sections in either *Hedyotis* (Wight and Arnott 1834) or *Oldenlandia* (Hooker 1880; Schumann 1891). But the short-style character (stigma held well below the anthers) of *Kohautia* made it easily recognizable as distinct from *Hedyotis* and *Oldenlandia*.

In molecular studies, the results showed *Hedyotis* and *Oldenlandia* to be paraphyletic (Bremer 1996b; Bremer and Manen 2000) or polyphyletic (Andersson and Rova 1999; Andersson, Rova et al. 2002) in tribe of Spermaceae *s.l.* In the following molecular studies, Karehed et al. (2008) and Groeninckx *et. al.* (2009) confirmed the paraphyletic position of *Hedyotis* and *Oldenlandia*.

Data from morphological and molecular studies for the delimitation of *Hedyotis*, *Oldenlandia* and other closely related genera were far from consensus. This situation is the motivation for us to investigate more Chinese specimens in order to improve understanding of the relationships of this complicated lineage.

#### **1.1.2.2.2 *Hedyotis* in China**

*Hedyotis* L. in China is represented by 67 species which are distributed mainly in tropical and subtropical areas, and a few of them also expanded to the temperate region; In *Flora Republicae Popularis Sinicae*, Ko (1999) recorded 60 species in *Hedyotis*. In later studies, Wang (2003, 2007 and 2008) published five new species (*Hedyotis wangii* R. J. Wang, *Hedyotis yanzhouensis* F. W. Xing et R. J. Wang, *Hedyotis koana* R. J. Wang, *Hedyotis cheniana* R. J. Wang and *H. wuzhishanensis* R.

J. Wang) and Chen (2007, 2008) also found two new species (*Hedyotis shenzhenensis* and *Hedyotis shiuyingiae*) (Wang and Xing 2003; Chen 2007; Wang 2007; Chen 2008; Wang 2008). Referring to the *Flora of Hong Kong* (vol. 3, 2009) and *Flora of Macau* (vol. 2, 2007), the number of *Hedyotis* species in Hong Kong is 20 (3 species are endemic in Hong Kong) and 19 are in Macao.

*Hedyotis* in the south-China region is usually herbs or weak subshrubs. The stems are erect or half-scandent, terete or quadrangular. The leaves are opposite with stipules. The forms of stipules are various, entire or fringed with bristle-like subulate lobes. Flowers are borne in axillary clusters, terminal panicles, or pedunculate cymes. Stamens are exerted from the corolla tube, and their anthers are basifixed. The ovary bears 2-3 locules, and the stigma is bilobed.

Recently, both chemical constituents and taxonomy of this genus have attracted much research attention. Many of the species in this genus are important herbal medicines (Luo, 1999), such as *H. diffusa*, which is used widely in pharmaceutical researches for cancer treatment. For taxonomic classification of *Hedyotis*, Luo in 1999 classified *Hedyotis* to five sections based on the types of capsule dehiscence in *Flora Republicae Popularis Sinicae* (Ko 1999). The five sections are *Hedyotis*, *Diplophragma* Wight et Arn., *Dimetia* Wight et Arn., *Euoldenlandia* K. Schum. and *Gonotheca* Meissn. Section *Hedyotis* is characterized by indehiscent capsules, while *Diplophragma* and *Dimetia* are distinct by septicidally dehiscent capsules. The capsule of *Dimetia* protrudes at the apex. *Euoldenlandia* and *Gonotheca* are identified by loculicidally dehiscent capsules. Pictures of representative species of each



section are shown in Fig. 1- 2.

However, this classification places much emphasis on using the site of dehiscence in the fruits as a key character in subdividing the genus into various sections and subgroups. This character, however, can be variable and subject to adaptations to different dispersal strategies. Fosberg in 1954 pointed out that the difference between seeds and fruits was not appropriate for defining genera. Verdcourt in 1958 suggested that although the nature of dehiscence of the seed is of more importance than the hardness or softness of the pericarp, using the character of seed dehiscence or indehiscence for group classification had numerous exceptions, and it was one of the least predictive. In 1961 Lewis also stated that this seed character cannot clearly differentiate between *Hedyotis* species. A previous molecular study in our laboratory has shown discrepancy from classifications based on the site of dehiscence (Li, Jiang et al. 2010). Reliance upon only one key character for section division in *Hedyotis* is not appropriate, as confirmed by Li's et al. (2010).

The morphological characters largely correlated with geographical and environmental factors which were subjectively determined by taxonomists to some extent. Adding more molecular characters to infer the intrageneric structure of the *Hedyotis* in China could provide a new perspective for a better phylogeny inference. It could also provide more information for the revision of *Hedyotis* in *Flora Republicae Popularis Sinicae*.



**Figures 1- 2. Pictures of representative species in five sections (Hedyotis, Diplophragma, Dimetia, Euoldenlandia and Gonotheca) of Hedyotis.** Hedyotis section: A. *Hedyotis costata*; B. *Hedyotis auricularia*; C. *Hedyotis verticillata*. Diplophragma section: D. *Hedyotis vachellii*; E. *Hedyotis loganioides*; F. *Hedyotis bodinieri*; G. *Hedyotis bracteosa*; H. *Hedyotis consanguinea*; Dimetia section: I. *Hedyotis hedyotidea*; Euoldenlandia section: J. *Hedyotis herbacea*; K. *Hedyotis corymbosa*; Gonotheca section: L. *Hedyotis biflora*.







## **1.2 Traditional Chinese medicine (TCM)**

### **1.2.1 Introduction**

The usage of Traditional Chinese Medicine for disease treatment and health cultivation has a 3000-year history in China, and China is the only country in the world where western medicine and traditional medicine are practiced alongside each other at every level of the healthcare system (Hesketh and Zhu 1997). According to a report of the World Health Organization (2002), TCM accounts for more than 40% of all health care delivery in China. Over the past several decades, Traditional Chinese Medicine has drawn a lot of attention from the west, and it is practiced in one form or another by more than 300,000 practitioners in over 140 other countries as well (Scheid 1999).

At present, over 10,000 species of plants, animals and minerals are reported with medicinal values, and 10% of them are frequently used in TCM (Shaw, Ngan et al. 2002). Adulteration for TCM is very common problem in markets. The reasons include: crude materials from various regions as the sources (Chan 1995), and the conventional adoption of morphological features for herbal identification by the dealers. There is a wide variation in this practice, and mistakes have always existed.

For international application, TCM must meet international standards, and its safety and quality must be assured (But 2000). Thus, using standardized procedures for authentication and quality control of TCM natural products and proprietary products are very important in the practice, and many methods, like anatomical analysis,



chemical and chromatographic screening method and molecular analysis, could be applied in examination.

### 1.2.2 Baihuasheshecao

The traditional Chinese herb Baihuasheshecao comes from the whole plant of *Oldenlandia diffusa* according to the record in *Pharmacopoeia of the People's Republic of China* (2005). In *Flora Republicae Popularis Sinicae* (Ko 1999), *Oldenlandia diffusa* is regarded as a synonym of *Hedyotis diffusa*. Here, *H. diffusa* is applied for source species of Baihuasheshecao. A sample of Baihuasheshecao purchased from a drug store is shown in **Fig. 1- 3**.

**Figure 1- 3. Picture of Baihuasheshecao sample.**



*Hedyotis diffusa* is an annual diffuse herb, about 10-30 cm tall. The linear-lanceolate leaves are opposite and sessile, about 1-3 mm broad. Stipules united at the base are triangular, and the apexes are aristate. The 4-merous flower is solitary or in pairs with 2-5 mm pedicels. The corolla of *H. diffusa* is white, and the tube is 1.5-2 mm. The stamens with very short filaments insert in the throat of the tube. The ovary is

inferior. The capsule is ovoid, opening at the apex when mature. *H. diffusa* is distributed in southern China in low altitude regions with relatively high humidity.

Traditionally, Baihuasheshecao was used to relieve 'internal heat', detoxification, facilitation of diuresis and elimination of 'internal wetness'. And it has long been applied in China for the treatment of hepatitis, tonsillitis, sore throat, appendicitis, urethral infection and various forms of malignant tumors (Xu, Xu et al. 1997). According to researches in recent decades, *H. diffusa* was found to have antitumor, immunomodulatory (Chan 1995; Yoshida, Wang et al. 1997; Shan, Zhang et al. 2001; Chung, Jeong et al. 2002), anti-inflammatory, anti-bacterial, hepatoprotective (Lin, Ng et al. 2002; Wang, Zhang et al. 2005), anti-oxidative (Lu, Yang et al. 2000; Kim, Park et al. 2001; Kim, Lee et al. 2005) and neuroprotective (Kim, Park et al. 2001) activities. Notably, *H. diffusa* won much attention from its use in for cancer treatment.

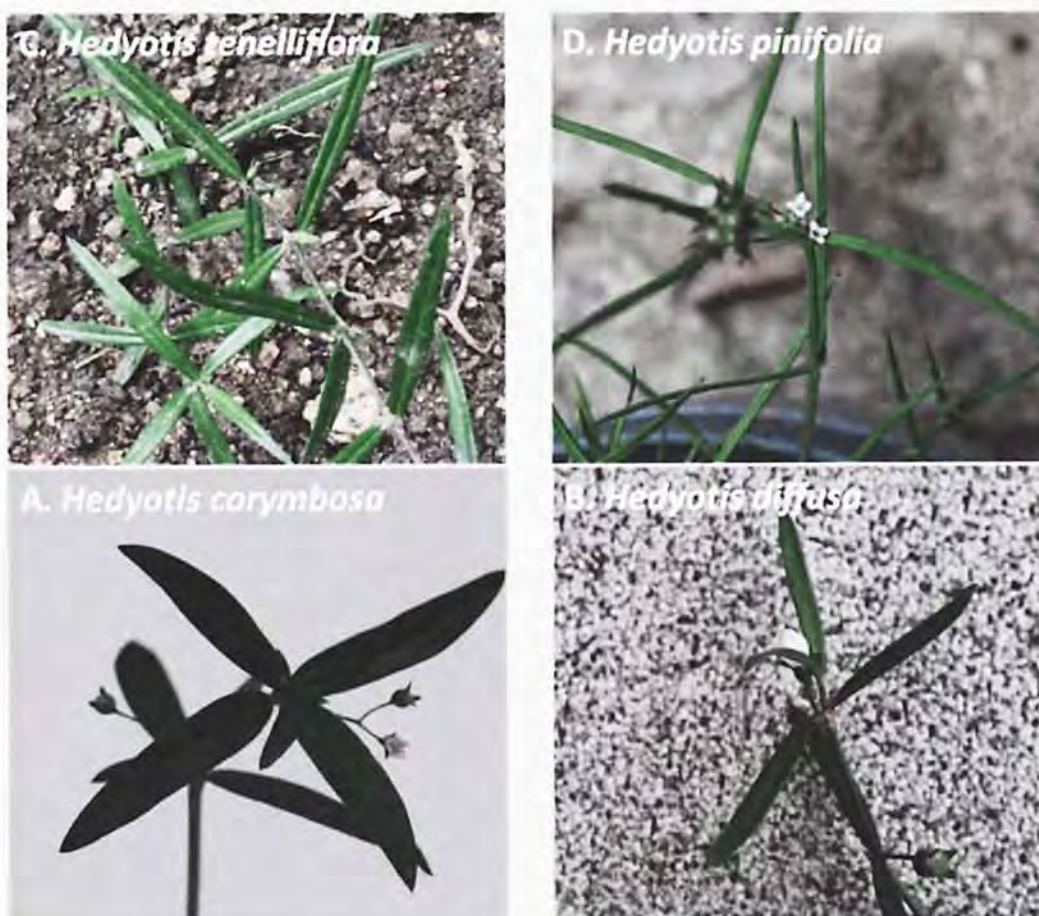
However, due to the overharvest of *H. diffusa*, the problem of substitution is very common in the market. The adulterants are commonly from the species in the same genus, as with *H. tenelliflora*, *H. corymbosa* and *H. pinifolia* (Yin and Hai 1989; Zheng 1990; Li, Ju et al. 1996), as well as species from different families (Caryophyllaceae, Molluginaceae and Loganiaceae etc.) The four species of *Hedyotis* (*H. diffusa*, *H. tenelliflora*, *H. corymbosa* and *H. pinifolia*) are similar in morphology, and the comparison of the morphological differentiation between the four species is shown in **Table1-1** and **Figure 1-4**.



**Table 1- 1. Morphological comparison of the four *Hedyotis* species.**

Species	Life form	Stem	Capsule	Flower
<i>H. diffusa</i>	annual herb	4-angled	Membranous, depressed globular, opening loculicidally at the apex	Solitary or in pairs, pedicel variable
<i>H. corymbosa</i>	annual herb	Sharpely 4-angled	Globular, with minute calyx-teeth in a ring around the apex	1-4 in axillary, terminal cymes, thin pedicel, 3-10 mm
<i>H. tenelliflora</i>	annual herb	Sharpely 4-angled	Ovoid or globular, crowned by linear-lanceolate calyx teeth	1-5 in axillary fascicles, no pedicel
<i>H. pinifolia</i>	perennial herb	Sharpely 4-angled	Sparsely hispidulous, opening at the apex	In axillary or terminal cymes, no pedicel

**Figure 1- 4. Pictures of the four *Hedyotis* species. A: *H. corymbosa*; B: *H. diffusa*; C. *H. pinifolia*; D. *H. tenelliflora*.**



### **1.2.3 Authentication of Baihuasheshecao**

Adulterants and Baihuasheshecao are different in their medical values, and the correct usage of genuine herbs is very important for the effectiveness of treatment.

Thus, application of effective authentication methods is very important.

Thin layer chromatography (TLC) (Xie, Zhang et al. 1997) and high performance liquid chromatography (HPLC) (Liang, Jiang et al. 2006) have been tried to authenticate genuine Baihuasheshecao species. Nuclear ribosomal DNA, called nrITS (internal transcribed spacer ITS1&2, 5.8S) region, have been widely applied in authentication and phylogenetic analysis of Chinese medicine (e.g., Lau, Shaw et al. 2001). DNA sequences from the ITS1 and 2 regions were found effective in differentiating Baihuasheshecao from its adulterants, especially, *H. corymbosa* (Hao, Liu et al. 2004; Liu, Hao et al. 2004; Liu and Hao 2005; Liu, Ding et al. 2006; Li, Jiang et al. 2010). However, cpDNA regions have not been used to distinguish the substitutes of Baihuasheshecao.

In this study, both nuclear and plastid regions will be used to differentiate the adulterants from genuine product.

## **1.3 Icefishes in the Hong Kong market**

### **1.3.1 Introduction**

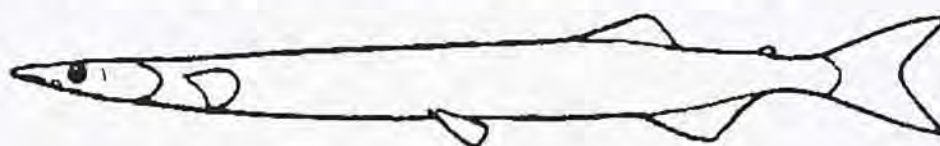
Salangids, commonly called icefish or noodlefish, belong to family Salangidae which is comprised of 17 species in six genera (1 species in *Protosalanx*, 2 species in



*Hemisalax*, 2 species in *Salanx*, 9 species in *Neosalanx*, 1 species in *Leucosoma* and 2 species in *Salangichthys*) (Nelson 1994; Zhang and Qiao 1994). Among them, 6 species (*N. tangkahkeii*, *N. taihuensis*, *N. oligodontis*, *N. pseudotaihuensis*, *N. brevirostris*, *Salanx cuvieri*) are endemic to China (Fang 1934; Zhang 1993; Xie and Xie 1997). Salangids are restricted to coastal seas of East Asia, ranging from Sakhalin, Vladivostok, Japan, the Korean Peninsula, China, to northern Vietnam (Cheng and Zheng 1987; Nelson 1994; Zhang and Qiao 1994; Sokolovskaya, Sokolovsky et al. 1998; Kottelat 2001; Kim and Park 2002). Salangids are all annual fishes, but their habitats are various. Most of them are anadromous, but some live in freshwater or are restricted to salt-water estuaries (Takita, Kawaguchi et al. 1988; Kawabata, Kubota et al. 1994; Takita 1995).

Family Salangidae belongs to superfamily Osmeroidea in order Osmeriformes and shares the character of fused palatine, pterygoid, quadrate and hyomandibular into hyopalatine (Nelson 1994). Salangids are characterized by neotenic features, a translucent body without scales, cartilaginous endoskeleton and notochord (Roberts 1984; Nelson 1994). The morphological and structural polymorphisms encountered in salangids handicap morphological-systematic analysis (Miya and Nishida 1996; Fu, Luo et al. 2005; Zhang, Li et al. 2007). The general appearance of salangids is shown in **Fig. 1- 5**. Moreover, species identification is further complicated by the neotenic features (Zhang, Li et al. 2007). Thus, the DNA markers would definitely help identification in salangids.

**Figure 1- 5. General appearance of salangids (Nelson 1994).**



Icefishes are high in nutritional value, especially their high protein content (12%-15%) (Xiong 1994). Referring to the '*Herbal Diet*' (Meng Xian, Dang dynasty, 612A.D. - 713 A.D.), icefishes provide the medical values, 'nourishment Ying, stomachic, benefits the lung and diuresis'. Not only has it been recommended as a nutritional food, but it has also been accepted as a delicious food in both China and other countries. For these reasons, icefishes are exported throughout the world with a production volume of 20,000 tons in 2003 in China (Li 2003). Usually, the large fresh product with milky and translucent white color bears a higher price.

This study carries out a FINS approach employing the mitochondrial 16S rDNA and 12S rDNA for clarifying the relationship between salangids and commercial icefishes products purchased in Hong Kong and Taiwan.

## **1.4 Molecular approach**

### **1.4.1 Introduction to molecular phylogeny**

Since Watson and Crick discovered the double helix structure of DNA 50 years ago, molecular technologies have deeply changed the ways to get the target data. For plant phylogenetic studies, using traditional ways to obtain the diagnostic morphological data for taxa delimitation is tedious and time-consuming, and even limited for the



numbers of representative type specimens. Molecular technologies have provided a much easier and faster way to access millions of characters. Moreover, among the various sources of comparative data currently applied to plant systematics (scientific study of all relationship and diversification among organisms through time; Simpsin, 1961), molecular data are the most intriguing, exciting, and conspicuous (Stuessy 2009). The debate regarding the utility and reliability of morphological and molecular methodologies has continued for decades. Actually, there is no evident difference in quality between morphological and molecular data, but DNA can produce more characters for phylogenetic construction (Bremer 1996a). Molecular approaches available for phylogentic study are various, including DNA–DNA hybridization, restriction fragment length polymorphisms, and polymerase chain reaction (PCR) based methods.

PCR was invented by Mullis (Mullis, Faloona et al. 1986; Mullis and Faloona 1987), and this technology has facilitated the development of automated DNA sequencing (McBride, Koepf et al. 1989; Swerdlow and Gesteland 1990; Tracy and Mulcahy 1991; Bock and Slightom 1995). The increasing sequences from different genes facilitate the employment of DNA regions for phylogeny estimation.

The nuclear ribosomal DNA or nrDNA region that codes for rRNA, is widely targeted for DNA sequencing. In plants, the ribosome submit region (5S, 5.8S, 18S, 26S/28S) of the nuclear genome has proven very useful at the higher levels of the hierarchy while the internal transcribed spacers (ITS1 and ITS2) between the subunits have also proven enormously resolvable at lower taxonomic levels (Soltis and Soltis 1998).

In animals, the 16S and 12S subunits have also been proved very useful in species identification. Moreover, the sequences from different plastid regions have been used frequently for phylogenetic study of plants (Olmstead and Franklin 1994).

#### **1.4.2 FINS (Forensically Informative Nucleotide Sequencing) for species identification**

Hampering species identification for food or herb is the lack of taxonomic knowledge and/or the morphological changes after processing the raw materials. Currently, as the advance of technologies, new approaches, like electrophoresis, immunological, chromatographic methods have been introduced to the tasks of identification for the resource species of food or herb. However, the protein-based approaches have limitations when dealing with the deep processing of raw materials for many types of commercial products (Bossier 1999; Teletcheaa 2005). The chemical method also has limitation in standardizing botanical preparations with variable sources and chemical complexity and in differentiation of closed species in the same genus (Joshi et al. 2004; Zhang et al. 2007).

The technology of FINS (Forensically Informative Nucleotide Sequencing) was first stated by Bartlett and Davidson (1992), who used it as an alternative technology for more accurate, easy and low cost approach. This approach involves four steps -- isolation DNA, PCR amplification of a target region, nucleotide sequencing and finally using a data base for phylogenetic analysis. Maximum parsimony analysis was been widely used for distinguishing different species as well as species



authentication for both TMC and food (Wen et al. 2010; Komatsu et al. 2001; Feng et al. 2010).

Moreover, with the development of automated nucleotide sequencing, the cost of sequencing was much cheaper. Practically, the identification of food or herb with FINS approach has been proved successful in many studies (Chapela et al. 2002; Guha and Kashyap 2006; Santaclara et al. 2007; Sahajpal and Goyal 2010). This method could apply to solve problems of adulterant, mis-labeling and illegal products for rights maintaining of customers, natural-resource protection and governments' criminal inspection (Blanco et al. 2008; Santaclara et al. 2007).

### **1.4.3 DNA sequence markers**

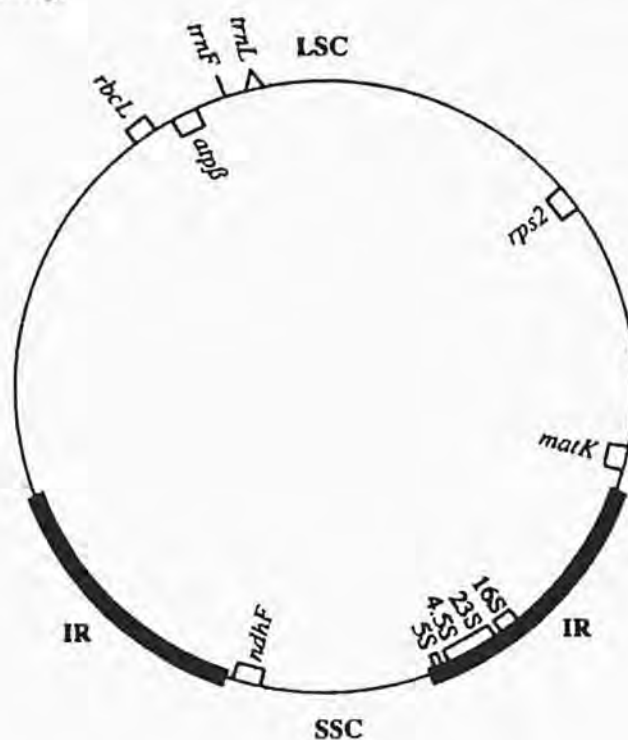
#### **1.4.3.1 Plastid DNA region**

The length of a typical plastid genome is usually from 135kb to 160kb in angiosperms (Stuessy 2009), and it is a circular molecule with two inverted repeats, a small single copy region and a large single-copy region. **Fig. 1- 6** shows the structure of plastid DNA (cpDNA).

Most genes in the plastid genome are single-copies, and they have a very slow evolution rate (Palmer 1985a; Palmer 1985b; Palmer 1986; Palmer 1991). The evolution rates of different parts of cpDNA are variable. As a result, applications of sequences from various parts of cpDNA regions for phylogenetic constructions vary from generic level to family level and even higher levels (Soltis and Soltis 1998). The protein-coding region of plastid DNA is widely used to infer the relationships of

angiosperms from the generic level to the familial level and even all lineages of land plants. Given the limitation of coding protein cpDNA for resolving relationships in lower taxonomic levels, comparative analysis of non-coding plastid sequences attracted much attention.

**Figure 1- 6. Schematic diagram of plastid genome (representative of most land plants).** IR: inverted repeat; SSC: small single-copy region; LSC: large single-copy. (Soltis and Soltis 1998).



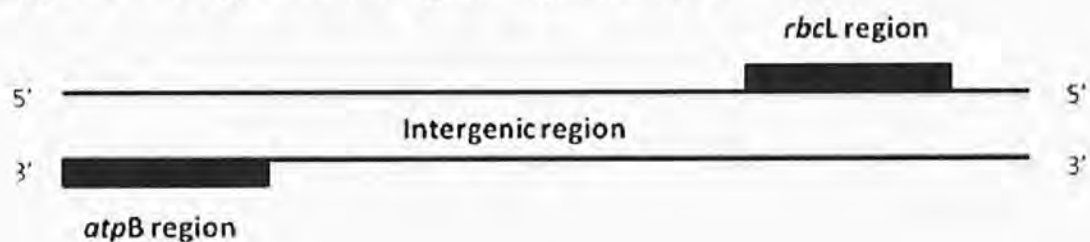
### ***rbcL* region**

The gene *rbcL* encodes the large subunit of the enzyme ribulose-1,5-bisphosphate carboxylase/ oxygenase (RuBisCO), and is located in the large single copy region of the cpDNA, with typical length of approximately 1430 bp (Fig. 1- 7). The region of *rbcL* was the first gene been sequenced in plants due to the key position of the



RUBISCO enzyme in the photosynthetic activity (Zurawski, Perrot et al. 1981; Zurawski, Whitfeld et al. 1986). The importance of the *rbcL* region in phylogenetic studies was emphasized by numerous times. It is an ideal gene applied in the phylogenetic studies of family-and-above levels (Soltis and Soltis 1998), with a wide taxonomic range in all lineages of land plants (Manhart 1994). Janssen and Bremer in 2004 stated that the conservative nature of *rbcL* makes it suitable for employment as a molecular clock and estimation of divergence of major groups (Janssen and Bremer 2004). The application of *rbcL* is limited in generic level analyses due to its conservative nature. As the evolution rate is various in different families, sequences of *rbcL* have also been applied to resolve generic or even specific levels (Soltis and Soltis 1998).

**Figure 1- 7. Diagram of *rbcL* gene in plastid DNA.**



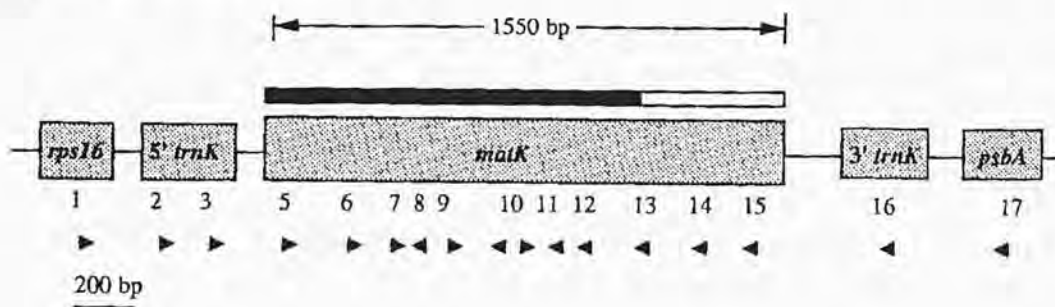
### ***matK* region**

The *matK* gene is located in the large single-copy region and is approximately 1550bp long. It is located within the intron of the transfer RNA gene for lysine, and encoding an intron-splicing maturease (Stuessy 2009) (Fig.1- 8). Among the protein-coding regions in plastid DNA, *matK* is one of the most rapidly evolving

(Wolfe 1991), as three times faster than the region of *rbcL*. The evolution rate of *matK* was proved suitable for the investigation of intergeneric or interspecific relationships (Soltis and Soltis 1998). Besides, it was also powerful in resolving the higher level of the hierarchy in angiosperm (Hilu, Borsch et al. 2003).

The *matK* gene could also be easily amplified by using highly conserved flanking coding regions that contain the *trnK* exons and the *rps16* and *psbA* genes (Soltis and Soltis 1998). The sequencing length of *matK* is usually about 1,000bp, which provides reliable resolution for phylogenetic studies within many families and genera of seed plants.

**Figure 1- 8. Location of *matK* in plastid DNA.** There is an approximately 1,000bp region of the black portion of bar above *matK* gene indicating the typical sequenced region (Soltis and Soltis 1998).



### ***rps16* intron region**

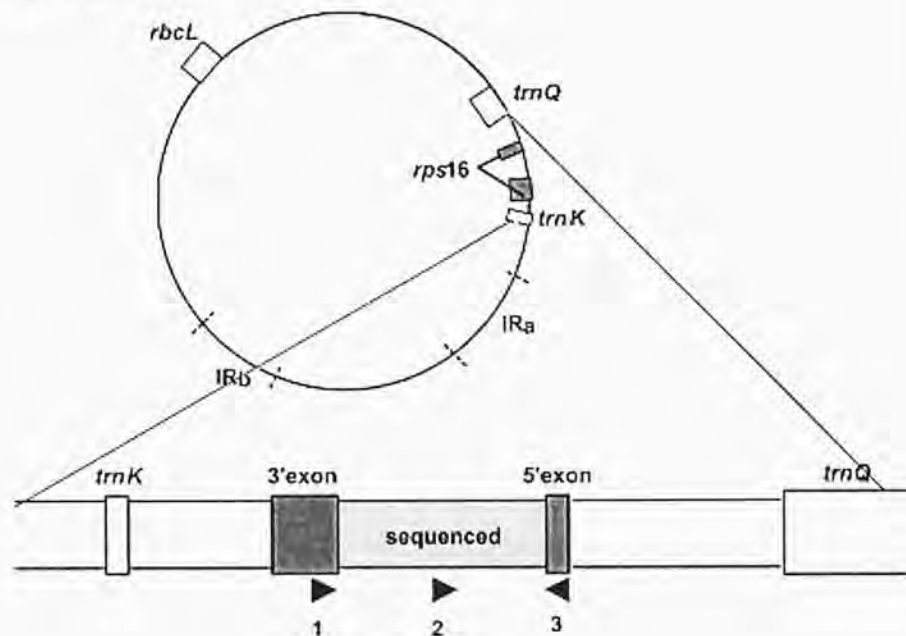
The plastid gene *rps16* intron region is located between *trnQ* and *trnK* within the large single-copy region in most flowering plants (Fig. 1- 9). According to several reports, this region is completely or partially missing in some members of many angiosperm families, including Linaceae, Malpighiaceae, Passifloraceae, Salicaceae,



Polygalaceae, Turneraceae, Violaceae, Connaraceae, Eucommiaceae, Fagaceae, and Leguminosae (Downie and Palmer 1992; Doyle, Doyle et al. 1995).

The length of the *rps16* region varies in different families. The evolutionary rate of *rps16* was two and three times lower than ITS region, which makes it suitable for the phylogenetic studies in generic to subfamilial level (Liden, Fukuhara et al. 1997; Oxelman, Liden et al. 1997).

**Figure 1- 9. Location of the *rps16* region in plastid DNA.** The *rps16* region is adjacent with both the *trnQ* and *trnK* genes. Positions of the primers are shown (Lee and Hymowitz 2001).

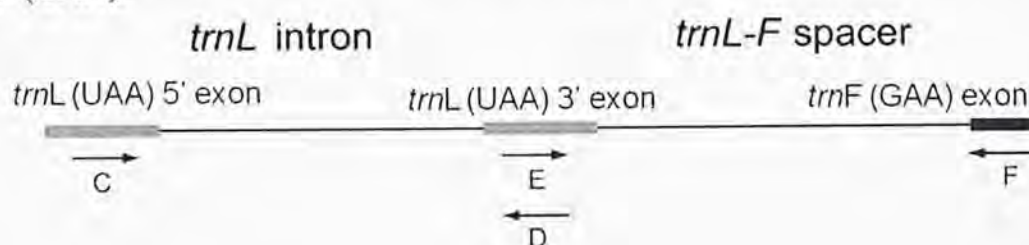


### ***trnL-F* region**

The *trnL-F* region is located in the large single-copy region, and is a non-coding protein region. The sequence of *trnL-F* is comprised of an intron in the transfer RNA gene *trnL* (UAA) and the intergenic spacers between *trnL* to *trnF* (GAA) (Stuessy

2009) (Fig. 1- 10). The *trnL* intron is recognized as the only Group I intron in plastids (Palmer 1991); it not only has a specific secondary structure but also has several highly conserved regions which can be found in all Group I introns (Westhof and Michel 1996; Stech, Quandt et al. 2003). The length of this region varies from different species, from only ~250bp in pteridophytes and bryophytes (Stech, Quandt et al. 2003) to as large as ~1400 bp in flowering plants (Bellstedt, Linder et al. 2001). Although the *trnL*-F intergenic spacer is shorter than the *trnL* intron, it shows a wide range of sizes in different plant groups from 100 bp in mosses and liverworts (Stech, Quandt et al. 2003) to more than 500 bp in angiosperms (Shaw et al 2005). With the publication of primers of ‘TabC’ and ‘TabF’ (Taberlet, Gielly et al. 1991), sequences of the *trnL* intron and *trnL*-F spacer were employed in numerous phylogenetic studies. This region is very popular in lower taxonomic levels of studies, particularly in inferring the relationship between genera and species. Because of the early publication and near-universal nature of the primers, this region has become one of the most widely used noncoding cpDNA sequences in plant systematics (Shaw, Lickey et al. 2005).

**Figure 1- 10. Diagram of *trnL*-F region.** This region is composed of an intron in transfer RNA gene *trnL* (UAA) and the intergenic spacer between *trnL* (UAA) and *trnF* (GAA).





### ***trnH – psbA* region**

The non-coding cpDNA, *trnH-psbA* intergenic spacer is located in the large single-copy region of most flowering plants. The length of this region is about 465 bp on average, and it ranges from 198 to 1077 bp in different genera (Shaw, Lickey et al. 2005). Sang in 1997 showed that the value of *trnH-psbA* intergenic region to plant systematics study by demonstrating that it is highly variable compared to *matK* and *trnL-trnF* (Sang, Crawford et al. 1997). The utility of this region in inferring the relationships in lower taxonomic levels is proved by many studies (e.g., Azuma, Thien et al. 1999; Hamilton 1999; Chandler, Bayer et al. 2001).

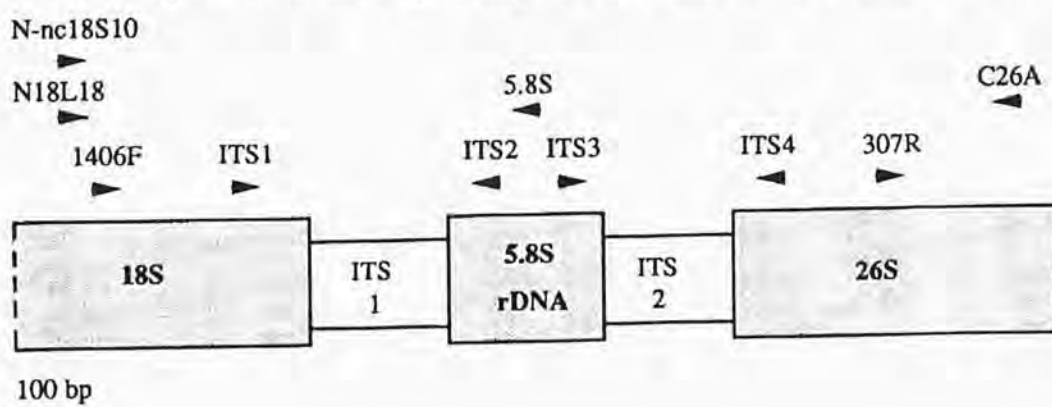
One advantage of this region is that it can be amplified and sequenced easily across all angiosperm lineages, as the universal primers for it were designed from the 75 bp length conserved adjacent region. The second advantage of this region is that it has a higher evolution rates than many non-coding regions, providing a very high percentage of variable characters (Hamilton, Braverman et al. 2003.). Although this region could provide enough variable characters, the small size limits the numbers of characters with which to build a well resolved phylogeny (Shaw, Lickey et al. 2005). So combining this locus with other regions is a good way for making up its limitation.

### **1.4.3.2 Nuclear DNA region**

The size of the nuclear genome is much larger than cpDNA, and the structure of a single unit of rDNA is shown in **Fig. 1- 11**. Each unit comprises an external

transcribed spacer (ETS), 18S subunit, an internal transcribed spacer (ITS1), 5.8S subunit, an second internal transcribed spacer (ITS2) and the 26S subunit, with thousands of repeats within most plant genomes. The lengths of each coding regions are different, but they are almost consistent in all plants: the 18S subunit is approximately 1800 bp (Nickrent and Soltis 1995), the 26S subunit is approximately 3,300 bp (Bult, Sweere et al. 1995), and the 5.8S subunit equals 160 bp (Takaiwa, Oono et al. 1985). The length of the IGS (intergenic spacer of Ribosomal DNA) and ITS1 and 2, however, are variable in different plants. With the tandem repeat structure and high copy number of rDNA (Rogers and Bendich 1987), attempts for amplifying or cloning this region is much easier than targeting a single-copy gene. Besides, rDNA exists in all organisms, and therefore has a much wider utility than the cpDNA region.

**Figure 1- 11. Diagram of a single unit of nrDNA.** In this diagram, the repeat unit of 18S and 26S nuclear ribosomal DNAs were separated by ITS1 and 2 regions and 5.8S nrDNA. Arrows represent the relative location of forward and backward primers available for amplification and sequencing of ITS-1 and ITS-2 (Palmer 1985a).



The nuclear region of ribosomal DNA cistron (rDNA) is the most frequently used



region for phylogenetic studies. The highly conserved regions of 18S and 26S rDNA are applied to the higher level of the hierarchy whereas the ITS and IGS regions are useful for constructing intraspecific and generic levels and even applied for the population-level studies.

### **nrITS 1&2**

In a repeat, ITS1 is located between the 18S subunit and 5.8S rDNA region, and ITS2 is located between the 5.8S region and 26S subunit. The length of the two regions vary from 565bp to 700bp in angiosperms (Liston, Robinson et al. 1996).

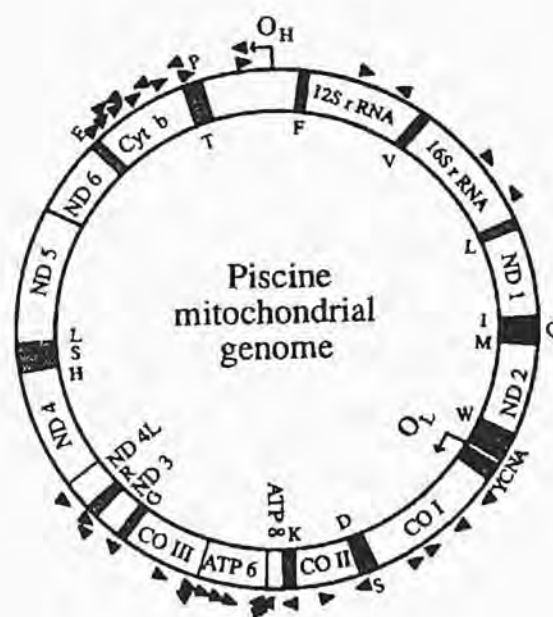
Amplification of the ITS region is easy in most angiosperms, due to the small size and 'universal primers' from highly conserved regions of the 18S and 26S subunits. Although amplification of ITS region is simple, sequencing of this region is hampered by the problem of multiple copies in genome (Alvarez and Wendel 2003) and possible occurrence of pseudogenes (Stuessy 2009). Sequencing two or more samples for each taxon is a good way to protect the reliability of the result, and another solution is to clone the ITS region to assure only one allele is sequenced. Despite the problem of sequencing of ITS, this region remains a good one for inferring phylogenetic analysis in intraspecific, intrageneric and even higher level of the hierarchy (Stuessy 2009).

### **1.4.3.3 Mitochondrial DNA region**

The length of the plant mitochondrial DNA genome is about ~16kb (Boore 1999) ,

and it is a single, double-stranded and circular molecule. The structure of the mitochondrial genome of fish is shown in Fig. 1.2, which contains 13 genes (seven subunits of NADH dehydrogenase, cytochrome *b*, three subunits of cytochrome *c* oxidase and two subunits of ATP synthetase) coding for proteins, two genes coding for ribosomal RNAs (12S and 16S rRNAs), and 22 genes coding for transfer RNAs (tRNAs) (Beaumont 1994). The evolutionary rate of mitochondrial DNA was proved 5-10 faster than single copy nuclear genes in animals (Brown, George et al. 1979; Perler, Efstratiadis et al. 1980), therefore providing more diverse sequences. After the introduction of 'universal primers', the sequences of mtDNA were heavily employed in phylogenetic studies (Kocher, Thomas et al. 1989).

**Figure 1- 12. Diagram of the Piscine mitochondrial genome.** The structure and gene order of mitochondrial genome in Piscine (Beaumont 1994).





## **16S and 12S Ribosomal RNA genes**

The ribosomal RNA gene in mitochondria is represented by two subunit, the 16S gene with 1571-1640 bp and 12S gene which is smaller, about 819-975 bp long in vertebrates (Beaumont 1994). 16S and 12S ribosomal RNA genes, with their important roles in protein synthesis, are relatively slowly evolving making them have the potential for resolving more ancient phylogenetic relationships (Mindell and Honeycutt 1999).

Both 16S and 12S have been found to provide a considerable number of informative characters and have been used frequently in many species-specific identification and phylogenetic studies for invertebrates (Chapela et al. 2002; Guha and Kashyap, 2006; Sahajpal and Goyal, 2010).

## **1.5 Objectives**

Because of the increasing reports of the mislabeling, adulterants and illegal products in the commercial market of China, food and drug authenticity have become a major concern of society. Although there are many analytical methods (morphological examination, protein technologies and chemical analyses) introduced to identify the genuine species for food and drugs, the limitation of these approaches have been discussed in many studies. As a fast, convenient, sensitive, accurate and stable analytical method for food and drug authenticity, molecular technology has increased steadily, and it has also been recognized as a good alternative to other technologies. The first aim of this study is to investigate the most suitable molecular markers with

the FINS analysis to identify the authority of two commodities, Icefish and Baihuasheshecao.

Icefish species are characterized by neotenic features, a translucent body without scales, cartilaginous endoskeleton and notochord (Roberts 1984; Nelson 1994), the traditional authentication approach like morphological and structural polymorphisms encountered in salangids handicap both species identification (Zhang et al. 2007) and systematics analysis (Miya and Nishida 1996; Fu et al. 2005; Zhang et al. 2007). This study carries out a FINS approach by employing mitochondrial 16S rDNA and 12S rDNA for clarifying the relationship between salangids and commercial icefishes products purchased in Hong Kong and Taiwan.

Baihuasheshecao (*Hedyotis diffusa*) is highly reputed as an herb for cancer treatment. However, many adulterants derived from *H. corymbosa*, *H. tenelliflora* and *H. pinifolia* of the same genus have been found on the market. The authentication of Baihuasheshecao is important to facilitate proper delivery of primary healthcare. In this study, plant samples of *Hedyotis* species and Baihuasheshecao commodities purchased from Hong Kong, mainland China and USA were collected for molecular authentication by one nuclear region (ITS1 and 2) and two plastid regions (*trnH-psbA* intergenic spacer region and the *trnL* intron and *trnL-F* intergenic spacer region) with FINS analysis. The efficiency of those three regions will be discussed in this study.

Additionally, the sequences from *Hedyotis* specimens used in this study could also be applied in solving a phylogeny problem in family Rubiaceae. As the position of



*Hedyotis* species from China in Spermacoceae has not been fully studied and the data of morphological and molecular studies for the position of *Hedyotis* is far from consensus, adding more molecular information of those species to inferring their phylogenetic relationships in this tribe could be very helpful to better understand the difficult *Hedyotis-Oldenlandia* group and their position in Rubiaceae. A further aim of this study is to use the sequences generated from six molecular regions to clarify the phylogenetic position of *Hedyotis* species in China and improve our understanding of the relationship of this complicated lineage.

## **CHAPTER 2**

# **MATERIALS AND METHODOLOGY**



## 2.1 Materials

Eight herb samples were brought from drug stores in Hong Kong, mainland China and USA, and they kept with silica gel in the Institute of Chinese Medicine (ICM) of CUHK. A total of 90 specimens representing 20 species in *Hedyotis* (Rubiaceae) were used in this study. They were collected from Hong Kong, Macao, mainland China and Taiwan with the help of staff from the Hong Kong Herbarium, Agricultural, Fisheries and Conservation Department, The Government of Hong Kong SAR and botanists, TAM Kai Yip, LAM Ying Wai, and TAM Kwok Kwong. To keep away from sunlight and wetness, the collected specimens were stored in the herbarium of The Chinese University of Hong Kong (CUHK). Total 20 icefishes samples brought from Hong Kong and Taiwan were stored at -20°C in the fridge at Food and Drug Authentication Laboratory, Department of Biology, The Chinese University of Hong Kong. Information on the Baihuasheshecao herbs, plants and icefishes is shown in **Tables 2-1- 2-3**.

Table 2- 1. List of herb samples used in this study is shown below.

Codes	Purchased places	Samples storage location
DS01	Zhejiang, Hangzhou 浙江杭州	Institute of Chinese Medicine Museum, CUHK
DS03	Tianjin 天津	Institute of Chinese Medicine Museum, CUHK
DS08	Hunan, Leng shuijiang 湖南冷水江	Institute of Chinese Medicine Museum, CUHK
DS10	WA USA 美國華盛頓州	Institute of Chinese Medicine Museum, CUHK
DS11	Guangzhou 廣州	Institute of Chinese Medicine Museum, CUHK
DS12	Hong Kong 香港	Institute of Chinese Medicine Museum, CUHK
DS13	Hong Kong 香港	Institute of Chinese Medicine Museum, CUHK
DS15	Boston, USA 美國波士頓	Institute of Chinese Medicine Museum, CUHK



Table 2- 2. List of field samples used in this study.

Species name	Codes	Collection location	Voucher specimen label	Specimen location/ Remarks
<b>Rubiaceae</b>				
<b>Group one: Sect. Hedyotis</b>				
<i>Hedyotis costata</i> (Roxb.) Kur	MA17	Trilho de Long Chao Kok, Macau	HED2009701	Herbarium Department of Biology CUHK
<i>Hedyotis costata</i> (Roxb.) Kur	MA18	Trilho de Long Chao Kok, Macau	HED2009702	Herbarium Department of Biology CUHK
<i>Hedyotis auricularia</i> L.	YU01	Saikung	HED2009703	Herbarium Department of Biology CUHK
<i>Hedyotis auricularia</i> L.	YU02	Saikung	HED2009704	Herbarium Department of Biology CUHK
<i>Hedyotis pinifolia</i> Wall. ex G. Don	MA11	Trilho do Nordeste de Coloane, Macau	HED2009705	Herbarium Department of Biology CUHK
<i>Hedyotis pinifolia</i> Wall. ex G. Don	MA12	Trilho do Nordeste de Coloane, Macau	HED2009706	Herbarium Department of Biology CUHK
<i>Hedyotis pinifolia</i> Wall. ex G. Don	MA13	Trilho do Nordeste de Coloane, Macau	HED2009707	Herbarium Department of Biology CUHK
<i>Hedyotis pinifolia</i> Wall. ex G. Don	YU19	Hong Kong	HED2009708	Herbarium Department of Biology CUHK
<i>Hedyotis tenelliflora</i> Blume	MA09	Trilho do Nordeste de Coloane, Macau	HED2009709	Herbarium Department of Biology CUHK
<i>Hedyotis tenelliflora</i> Blume	MA10	Trilho do Nordeste de Coloane, Macau	HED2009710	Herbarium Department of Biology CUHK
<i>Hedyotis tenelliflora</i> Blume	TM05	Taimoshan, Hong Kong	HED2009711	Herbarium Department of Biology CUHK
<i>Hedyotis tenelliflora</i> Blume	TM06	Taimoshan, Hong Kong	HED2009712	Herbarium Department of Biology CUHK
<i>Hedyotis verticillata</i> (L.) Lam.	MA14	Colina da Guia Municipal Park, Macau	HED2009713	Herbarium Department of Biology CUHK
<i>Hedyotis verticillata</i> (L.) Lam.	MA15	Colina da Guia Municipal Park, Macau	HED2009714	Herbarium Department of Biology CUHK
<i>Hedyotis verticillata</i> (L.) Lam.	MA16	Colina da Guia Municipal Park, Macau	HED2009715	Herbarium Department of Biology CUHK
<i>Hedyotis verticillata</i> (L.) Lam.	YU12	Lion Rock Country Park, Hong Kong	HED2009716	Herbarium Department of Biology CUHK
<i>Hedyotis costata</i> (Roxb.) Kurz	YU26	/	HK 36764	Herbarium of Hong Kong HK 36764
<b>Group two: Sect. Diplophragma</b>				
<b>Wight et Arn.</b>				

<i>Hedyotis acutangula</i> Champ. ex B enth.	HA01	Tailongsaiwan, Hong Kong	HED2009717	Herbarium Department of Biology CUHK
<i>Hedyotis acutangula</i> Champ. ex B enth.	HA02	Tailongsaiwan, Hong Kong	HED2009718	Herbarium Department of Biology CUHK
<i>Hedyotis acutangula</i> Champ. ex B enth.	BW21	Pat sin leng, Hong Kong	HED2009719	Herbarium Department of Biology CUHK
<i>Hedyotis assimilis</i> Tutch	BW07	Guangzhou	HED2009720	Herbarium Department of Biology CUHK
<i>Hedyotis assimilis</i> Tutch	BW09	Guangzhou	HED2009721	Herbarium Department of Biology CUHK
<i>Hedyotis bodinieri</i> H.Lév.	TM22	Taimoshan, Hong Kong	HED2009722	Herbarium Department of Biology CUHK
<i>Hedyotis bodinieri</i> H.Lév.	YU09	Taimoshan, Hong Kong	HED2009723	Herbarium Department of Biology CUHK
<i>Hedyotis bodinieri</i> H.Lév.	YU10	Taimoshan, Hong Kong	HED2009724	Herbarium Department of Biology CUHK
<i>Hedyotis bodinieri</i> H.Lév.	YU11	Taimoshan, Hong Kong	HED2009725	Herbarium Department of Biology CUHK
<i>Hedyotis bracteosa</i> Hance	MA01	Trilho do Nordeste de Coloane, Macau	HED2009726	Herbarium Department of Biology CUHK
<i>Hedyotis bracteosa</i> Hance	MA02	Trilho do Nordeste de Coloane, Macau	HED2009727	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	BX03	Pat sin leng, Hong Kong	HED2009728	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	BW20	Pat sin leng, Hong Kong	HED2009729	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	MA05	Statue of A-Ma Alto de Coloane, Macau	HED2009730	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	TM01	Taimoshan, Hong Kong	HED2009731	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	TM12	Taimoshan, Hong Kong	HED2009732	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	TM13	Taimoshan, Hong Kong	HED2009733	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	TM14	Taimoshan, Hong Kong	HED2009734	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	TM15	Taimoshan, Hong Kong	HED2009735	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	YU13	Taimoshan, Hong Kong	HED2009736	Herbarium Department of Biology CUHK
<i>Hedyotis consanguinea</i> Hance	YU15	Sai Kung, Hong Kong	HED2009737	Herbarium Department of Biology CUHK
<i>Hedyotis loganioides</i> Benth.	YU16	Ma On Shan, Hong Kong	HED2009738	Herbarium Department of Biology CUHK



<i>Hedyotis loganioides</i> Benth.	YU17	Ma On Shan, Hong Kong	HED2009739	Herbarium Department of Biology CUHK
<i>Hedyotis longiexserta</i> Merr. & F.P. Metcalf	BW06	Guangzhou	HED2009740	Herbarium Department of Biology CUHK
<i>Hedyotis shuiyingiae</i> T. Chen	TM09	Tai Mo Shan, Hong Kong	HED2009741	Herbarium Department of Biology CUHK
<i>Hedyotis shuiyingiae</i> T. Chen	TM10	Tai Mo Shan, Hong Kong	HED2009742	Herbarium Department of Biology CUHK
<i>Hedyotis shuiyingiae</i> T. Chen	TM11	Tai Mo Shan, Hong Kong	HED2009743	Herbarium Department of Biology CUHK
<i>Hedyotis shuiyingiae</i> T. Chen	TM19	Tai Mo Shan, Hong Kong	HED2009744	Herbarium Department of Biology CUHK
<i>Hedyotis shuiyingiae</i> T. Chen	TM20	Tai Mo Shan, Hong Kong	HED2009745	Herbarium Department of Biology CUHK
<i>Hedyotis uncinella</i> Hook. & Arn.	TM04	Tai Mo Shan, Hong Kong	HED2009746	Herbarium Department of Biology CUHK
<i>Hedyotis uncinella</i> Hook. & Arn.	TM16	Tai Mo Shan, Hong Kong	HED2009747	Herbarium Department of Biology CUHK
<i>Hedyotis uncinella</i> Hook. & Arn.	TM17	Tai Mo Shan, Hong Kong	HED2009748	Herbarium Department of Biology CUHK
<i>Hedyotis uncinella</i> Hook. & Arn.	TM18	Tai Mo Shan, Hong Kong	HED2009749	Herbarium Department of Biology CUHK
<i>Hedyotis uncinella</i> Hook. & Arn.	YU03	Tai Mo Shan, Hong Kong	HED2009750	Herbarium Department of Biology CUHK
<i>Hedyotis uncinella</i> Hook. & Arn.	YU04	Tai Mo Shan, Hong Kong	HED2009751	Herbarium Department of Biology CUHK
<i>Hedyotis vachellii</i> Hook. & Arn.	YU21	Lan Tau Peak, Hong Kong	HED2009752	Herbarium Department of Biology CUHK
<i>Hedyotis vachellii</i> Hook. & Arn.	YU22	Lan Tau Peak, Hong Kong	HED2009753	Herbarium Department of Biology CUHK
<i>Hedyotis vachellii</i> Hook. & Arn.	YU23	Lan Tau Peak, Hong Kong	HED2009754	Herbarium Department of Biology CUHK
<i>Hedyotis vachellii</i> Hook. & Arn.	YU24	Lan Tau Peak, Hong Kong	HED2009755	Herbarium Department of Biology CUHK
<b>Group Third: Sect. Dimetia</b>				
<b>Wight et Arn.</b>				
<i>Hedyotis hedyotidea</i> (DC.) Merr.	BW23	Pat Sin Leng, Hong Kong	HED2009756	Herbarium Department of Biology CUHK
<i>Hedyotis hedyotidea</i> (DC.) Merr.	KL02	Kei Leng, Hong Kong	HED2009757	Herbarium Department of Biology CUHK
<i>Hedyotis hedyotidea</i> (DC.) Merr.	KL03	Kei Leng, Hong Kong	HED2009758	Herbarium Department of Biology CUHK

<i>Hedyotis hedyotidea</i> (DC.) Merr.	MA03	Parque de Seac Pai Van (Seac Pai Van Country Park), Coloane, Macau	HED2009759	Herbarium Department of Biology CUHK
<i>Hedyotis hedyotidea</i> (DC.) Merr.	MA04	Parque de Seac Pai Van (Seac Pai Van Country Park), Coloane, Macau	HED2009760	Herbarium Department of Biology CUHK
<i>Hedyotis hedyotidea</i> (DC.) Merr.	TM21	Tai Mo Shan, Hong Kong	HED2009761	Herbarium Department of Biology CUHK
<b>Group Fourth:</b>				
<b>Sect. <i>Euoldenlandia</i> K.Schum.</b>				
<i>Hedyotis corymbosa</i> (L.) Lam.	BW22	Pat Sin Leng, Hong Kong	HED2009762	Herbarium Department of Biology CUHK
<i>Hedyotis corymbosa</i> (L.) Lam.	MA23	Lotus Bridge, Coloane, Macau	HED2009763	Herbarium Department of Biology CUHK
<i>Hedyotis corymbosa</i> (L.) Lam.	MA24	Lotus Bridge, Coloane, Macau	HED2009764	Herbarium Department of Biology CUHK
<i>Hedyotis corymbosa</i> (L.) Lam.	MA25	Arboreto, Encosta do Alinho de Coloane, Macau	HED2009765	Herbarium Department of Biology CUHK
<i>Hedyotis corymbosa</i> (L.) Lam.	MA26	Façade of the ruins of Church of Mater Dei, Macau	HED2009766	Herbarium Department of Biology CUHK
<i>Hedyotis corymbosa</i> (L.) Lam.	TA01	Taiwan	HED2009767	Herbarium Department of Biology CUHK
<i>Hedyotis corymbosa</i> (L.) Lam.	TB01	Taiwan	HED2009768	Herbarium Department of Biology CUHK
<i>Hedyotis corymbosa</i> (L.) Lam.	YU14	Sai Kung, Hong Kong	HED2009769	Herbarium Department of Biology CUHK
<i>Hedyotis diffusa</i> Willd.	BW16	Guangzhou	HED2009770	Herbarium Department of Biology CUHK
<i>Hedyotis diffusa</i> Willd.	BW17	Guangzhou	HED2009771	Herbarium Department of Biology CUHK
<i>Hedyotis diffusa</i> Willd.	BW18	Guangzhou	HED2009772	Herbarium Department of Biology CUHK
<i>Hedyotis diffusa</i> Willd.	MA19	Lotus Bridge, Coloane, Macau	HED2009773	Herbarium Department of Biology CUHK
<i>Hedyotis diffusa</i> Willd.	MA20	Lotus Bridge, Coloane, Macau	HED2009774	Herbarium Department of Biology CUHK
<i>Hedyotis diffusa</i> Willd.	MA21	Lotus Bridge, Coloane, Macau	HED2009775	Herbarium Department of Biology CUHK
<i>Hedyotis diffusa</i> Willd.	MA22	Lotus Bridge, Coloane, Macau	HED2009776	Herbarium Department of Biology CUHK



<i>Hedyotis diffusa</i> Willd.	MA27	Facade of the ruins of Church of Mater Dei, Macau	HED2009777	Herbarium Department of Biology CUHK
<i>Hedyotis diffusa</i> Willd.	TY01	Pai O, Lan Tau Island	HED2009778	Herbarium Department of Biology CUHK
<i>Hedyotis herbacea</i> L.	YU05	Kai Pik Shan, Sai Kung, Hong Kong	HED2009779	Herbarium Department of Biology CUHK
<i>Hedyotis herbacea</i> L.	YU06	Kai Pik Shan, Sai Kung, Hong Kong	HED2009780	Herbarium Department of Biology CUHK
<i>Hedyotis herbacea</i> L.	YU07	Kai Pik Shan, Sai Kung, Hong Kong	HED2009781	Herbarium Department of Biology CUHK
<b>Group Fifth</b>				
<b>Sect. <i>Gonothecha</i> Meissn.</b>				
<i>Hedyotis biflora</i> (L.) Lam.	MA06	Jardim Lou Lim Iok, Macau	HED2009782	Herbarium Department of Biology CUHK
<i>Hedyotis biflora</i> (L.) Lam.	MA07	Jardim Lou Lim Iok, Macau	HED2009783	Herbarium Department of Biology CUHK
<i>Hedyotis biflora</i> (L.) Lam.	MA08	Jardim Lou Lim Iok, Macau	HED2009784	Herbarium Department of Biology CUHK
<i>Hedyotis biflora</i> (L.) Lam.	YU20	Lantau island, Hong Kong	HED2009785	Herbarium Department of Biology CUHK
<b>Molluginaceae</b>				
<i>Mollugo stricta</i> L.	BW03	Guangzhou	HED2009786	Herbarium Department of Biology CUHK
<b>Loganiaceae</b>				
<i>Mitrasacme pygmaea</i> R. Brown	BW19A	Pat Sin Leng, Hong Kong	HED2009787	Herbarium Department of Biology CUHK
<i>Mitrasacme pygmaea</i> R. Brown	BW19	Pat Sin Leng, Hong Kong	HED2009788	Herbarium Department of Biology CUHK

**Table 2- 3. Codes for sample of icefishes are listed below.**

<b>Codes</b>	<b>Commodity names</b>	<b>Voucher specimen label</b>	<b>Collection Place</b>
D01	dry icefishes	SA200911	Shatin Market
D02	dry icefishes	SA200912	Shatin Market
D03	dry icefishes	SA200913	Sham Shui Po Market
D04	super dry icefishes	SA200914	Sham Shui Po Market
D05	Yunan dry icefishes	SA200915	Hong Kong
D06	Yunan big dry icefishes	SA200916	Hong Kong
F01	fresh icefishes	SA200917	Tai Po market
F02	fresh icefishes	SA200918	Tai Po market
F03	fresh icefishes	SA200919	Fanling
F04	fresh icefishes	SA200911	Shatin
F05	fresh icefishes	SA200912	Shatin Market
F06	fresh icefishes	SA200913	Sham Shui Po Market
F07	fresh icefishes	SA200914	Sham Shui Po Market
F08	fresh icefishes	SA200915	Sherk Wu Hui Market
F09	fresh icefishes	SA200916	Yuen Long
F10	fresh icefishes	SA200917	Sherk Wu Hui Market
F11	fresh icefishes	SA200918	Taiwai
F12	fresh icefishes	SA200919	Tai Po market
S01	pickled icefishes	SA200920	Shatin market pickled fish
S04	pickled icefishes	SA200921	Taiwai Cijin pickled fish



## 2.2 DNA extraction

Modified CTAB method was used for DNA extraction from fresh or dried leaf samples and commercial icefishes samples as stated in the **Tabs. 2.1-2.3**. First, experimental samples were washed by 75% ethanol and added with 400  $\mu$ l extraction buffer. They were then cut into small pieces by scissors, well mixed and incubated at 37°C for about 1 hr. Then, 400  $\mu$ l of 2X CTAB solution was added and gently mixed by pipetting up and down. An equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added to the mixture to extract the released protein, carbohydrates and liquid. The eppendorf tube was inverted several times before being put into the centrifuge. After the mixture was centrifuged for 10 min at 13,000 rpm, the aqueous layer was pipetted into a new sterile eppendorf. The same extraction procedure was repeated by using an equal volume of chloroform/isoamyl alcohol (24:1) for 15 min, for the removal of phenol from the aqueous phase. Following centrifuging, ~ 350  $\mu$ l of supernatant was removed to another sterile tube with adding 2/3 volume of isopropanol. The tube was then placed on ice for 20 min and centrifuged at the 13,000 rpm for another 10 min. Supernatant was discarded, and the pellet was washed with 75% ethanol twice. Next, the tube with pellet was put onto a 60°C heat block until the pellet was fully dried. The dried DNA pellet was dissolved in 50  $\mu$ l double distilled H<sub>2</sub>O and stored at -20°C. And the quality of DNA was checked by gel electrophoresis and OD absorbance at 260 and 280 nm.

**Table 2- 4. List of reagents and equipments used in modified CTAB method.**

<b>Reagents</b>	<b>Equipments</b>
<b>Extraction buffer</b>	
200 mM Tris-HCl (pH 8.0) (121 g/L Tris)	Sterilized scissors
200 mM NaCl (23.376 g/L NaCl)	
25 mM EDTA (pH 8.0) (186.1 g/L ethylenediaminetetra-acetate · 2H <sub>2</sub> O)	Incubator
0.5% SDS	Alcohol burner
<b>2X CTAB solution</b>	
2% CTAB (w/v)	Centrifuge
100 mM Tris-HCl (pH 8.0) (121 g/L Tris)	
20 mM EDTA (pH 8.0)	Shaker
1.4 M NaCl	
1% PVP (polyvinylpyrrolidone), MW 40,000	Heat block
<b>Phenol:chloroform:isoamyl alcohol (25:24:1)</b>	
<b>Chloroform:isoamyl alcohol (24:1)</b>	
<b>Isopropanol</b>	
<b>75% ethanol</b>	
<b>Double distilled H<sub>2</sub>O</b>	

### **2.3 Polymerase chain reaction (PCR) method**

17.3 µl double distilled water, 2.5 µl 10X PCR buffer [10 mM Tris-HCl (pH 8.3), 50 mM KCl, 25 mM MgCl<sub>2</sub>, 0.001% gelatin], 1 µl 2.5 mM dNTPs, 0.5 µl 10 mM forward and backward primer respectively, 0.2 µl *Taq* polymerase and diluted DNA 1.5 µl were added in sequence to a 200 µl PCR tube and gently mixed. The reaction mixture was put in a thermocycler (Eppendorf Mastercycler<sup>®</sup> gradient thermal cycler) for running reaction. The running profile is:

1. 95°C for 5 min
2. 95°C for 1 min



3. 5-10 number of cycles: 95°C for 1 min → 40°C for 1 min → 72°C for 1 min
4. 25-30 number of cycles: 95°C for 45 sec → 50°C-53°C for 1 min → 72°C for 1 min
5. 15°C forever (temporary storage)

The quality of PCR product was analyzed by gel electrophoresis.

Table 2-5. List of amplified regions, primer sequences, amplified size and their sources.

Genes	Primer sequences	Sources	Size (bp)
ITS (ITS1, 5.8S, ITS2)	ITS4 (5'-CACACCGCCCGTCGCTCCTACC GA -3')	White, 1990	~ 600
	RUT-ITSF1 (5'-TCGCGAGAAAGTCCACTGA -3')		
<i>trnL</i> intron and <i>trnL-trnF</i> intergenic spacer	TabC (5'-CGAAATCGGTAGACGCTACG -3')	Taberlet, 1991	~ 900
	TabF (5'-ATTTGAACTGGTGACACGAG -3')		
<i>psbA-trnH</i> intergenic spacer	<i>psbAF</i> (5'-GTTATGCAATGAACGTAATGCTC -3')	Sang et al., 1997	~ 400
	<i>trnHR</i> (5'-CGCGCATGGTGGATTCAAAATC -3')		
<i>matK</i> gene	<i>matK_2F</i> (5'-GTTCACTAATTGTGAAACGT -3')	Sang et al., 1997	~ 1,000
	<i>matK_3R</i> (5'-GATCCGCTGTGATAATGAGA -3')		
<i>rps16</i> gene	<i>rps16_F</i> (5'-GTGGTAGAAAGCAAACGTCGCACTT-3')	Oxelman, 1997	~ 800
	<i>rps16_R3</i> (5'-CGAATAGACGGCTCAT TGGGAT A -3')	Bremer, 2002	
<i>rbcL</i> gene	<i>rbcL_Z674F</i> (5'-TTTATAAATCACAAGCCGAAACTGGTG AAATC -3')	Bremer, 2002	~ 650
	<i>rbcL_Z1375R</i> (5'-AAT TTG ATC TCC TTCCAT ATTTCGCA -3')		
16S rRNA	16L1 (5'-CTGACCGTGCAAAGGTAGCGGTAATCACT- 3')	Hedges, 1994	~400
	16H1 (5'-CTCCGGTCTGAACACTCAGATCACGTAGG- 3')		
12S rRNA	12L1 (5'-AAAAAGCTTCAAACCTGGGATTAGATACCCCACTAT- 3')	Hedges, 1994	~400
	12H1 (5'-TGACTGCAGAGGGGTGACGGGGCGGTGTGT- 3')		



## **2.4 Gel electrophoresis**

Agarose gel was used to analyze DNA extraction product and PCR product. 1.4-1.6% agarose gel was prepared by heating 1.4-1.6 g agarose per 100 ml TAE [44.6 g/L Tris-base, 11.4 ml acetic acid, 20 ml/L 0.5 M EDTA pH8.0] in microwave for 4-6 min and 0.5 µg/ml ethidium bromide was added to the hot solution and gently mixed. Then, the solution was poured into the gel cassette for cooling down. After about 30 min later, the solidified agarose gel was put into a gel tank filled with 1X loading buffer. At the same time, loading buffer 6X was mildly mixed with either DNA extraction samples or PCR products, and the mixture was added into wells of gel. The 100 bp DNA marker was loaded at last. The gel was kept at constant voltage of 100-120 V for 20-40 min.

After gel running, it was examined under UV illumination. A photo of the gel was taken by BIORAD Gel Documentation System 1000 and printed.

## **2.5 PCR production purification**

Following the gel electrophoresis of PCR product, it was purified by using VIogene Gel-M<sup>®</sup> Gel Extraction System (Cat. No.: EG1002) and HOU-BIO DNA Gel Extraction Kit (Cat # H-026) following the manufacturers' instructions.

### **2.5.1 GEL-M<sup>TM</sup> GEL EXTRACTION SYSTEM**

The gel slice containing the interested DNA fragment was excised with a clean sharp scalpel by removing the extra gel under UV light. Then, the DNA fragment was put into a sterile tube with 500 µl GEX buffer and incubated at 60°C for 5 - 10 min until

the gel was completely dissolved. During the incubation, the tube was inverted several times in every 1-2 min. When it cooled down to room temperature, the mixture was loaded to the Gel-M™ column which had already placed in a collection tube. The column was centrifuged for 60 seconds, and the flow-through was discarded. The column was washed once with 500 µl WF buffer by centrifuging for 75 seconds at 12,100×g, and the flow-through was discarded. Then, 700 µl WS buffer was added to the column by centrifuging for 75 seconds at 13,000 rpm. The flow-through was discarded and the column was sent to centrifuge again for 3 min at 13,000 rpm. The column was then placed onto a sterile eppendorf, and 50 µl double distilled water was added. After standing the column for about 3 min, it was sent to centrifuge for DNA fragment collection. The quality of the DNA was analyzed by gel electrophoresis under UV light. Next, the extracted DNA sample was stored at -20°C.

### **2.5.2 HOU-BIO DNA Gel Extraction Kit**

The procedure of the HOU-BIO DNA Gel Extraction Kit is listed below.

The minimum amount of gel containing the target DNA band was carefully cut under UV light by a clean sharp scalpel, and the DNA band was removed to the sterile eppendorf with 600 µl binding buffer. The tube was kept in the incubator at 60°C for 5-10 min until the gel is completely dissolved. Meanwhile, the mixture was inverted several times during the incubation. The column was place into a collection tube, waiting for the mixture to cool down. Following that, the cooled mixture was added to a new column and centrifuged at 13,000 rpm for 75 seconds, and the flow-through was discarded. Then, the column was centrifuged at 13,000 rpm for 3 min for drying.



Next, the column was transferred to a new sterile eppendorf tube, and 50  $\mu$ l double distilled water was added. About 5 min later, the eppendorf was centrifuged at 12,100 rpm for 3 min of eluted DNA fragment collection. The quality of the DNA was analysis by gel electrophoresis under UV light, and then the DNA was stored at  $-20^{\circ}\text{C}$ .

## **2.6 Ligation and transformation and transformation of PCR product**

For PCR products that were contaminated by fungal DNA or had the problem of multiple copies, it is necessary to insert the PCR product to single copy plasmids for sequencing.

### **2.6.1 Ligation and transformation**

The kit from GEM<sup>®</sup>-T Easy Vector System I (Cat # A1360) was used for the cloning of PCR products. The procedure is followed by the protocol provided by the manufacturer with mild modification. The ampicillin solution was prepared by 50 mg/ml ampicillin in MilliQ water (filtered using 0.2 micron filter to sterilize), and then stored at  $-20^{\circ}\text{C}$ .

Before ligation and transformation, the LBA plate and LB were prepared. And then, the ligation reaction system was set up in a sterile 0.5 ml tube as described in Table 2.7. Before using the 2X rapid ligation buffer, it had to be shaken vigorously.

**Table 2- 6. List of reagents and equipment used for LBA plate and LB were prepared during the ligation and transformation.**

Reagent	Equipment
<b>LBA preparation</b>	Beaker
Luria agar 11.1 g	Autoclave machine
Double distilled water 300 volume	Plate
Ampicillin 300 $\mu$ l	
<b>LB preparation</b>	
Luria-Bertani powder 2.5 g	
Double distilled water 100 ml	

**Table 2- 7 List of ligation reaction system used in this study.**

Reagents	Standard ( $\mu$ l)	Control ( $\mu$ l)
2X Rapid Ligation Buffer	2.5	2.5
T4 DNA ligase pGEM <sup>®</sup> -T Vector	0.5	0.5
T4DNA ligase (3 Weiss units/ $\mu$ l)	0.5	0.5
PCR Product	1.75	/

The mixture was mildly mixed by pipetting up and down, and incubated at room temperature for 4 hours. Alternatively, the reaction was incubated at 25°C for 2 hr. The tubes with competent cells were placed on ice until just thawed. Cells were mixed by gently flicking the tube, and transferred to the ligation reaction mixture. The mixture was set on ice for 30 min and transferred to the incubator at 42°C for 1 min. After that, the tube was immediately returned to ice for one minute. Meanwhile, 100  $\mu$ l LB medium was put into incubator for preheating at 37°C. Then, the LB medium was added to the mixture for ligation reaction transformation. Next, the tube



was placed at 37°C for incubation. After 30 min incubation, 50 µl of 5% X-gal and 5 µl of 0.5 M IPTG mixture was added into the ligation reaction system. After spreading the mixture onto a LBA/ampicillin plate, the plate was incubated at 37°C for ~16 hr.

### **2.6.2 Cell culture**

After 16 hr incubation, 3 white colonies were picked from each LBA/ampicillin plate. About 50 ml LB median was mixed with 50 µg/ml ampicillin, and 1 ml of the mixture was added to each tube. Each colony was then transferred to the LB/ampicillin tube. The culture was put into the incubator at 37°C for shaking about 16 h. After incubation, the tube was centrifuged at 12,000 rpm for 10 min. The supernatant was discarded, and the cell culture pellet was left at the bottom of the tube. Before being sent to sequencing, the tube was stored at -20°C.

## **2.7 DNA sequencing and sequence analyses**

The purifying PCR products were sent to either Macogen or Beijing Institute of Genome (BIG) for sequencing.

The sequencing result came back from sequencing companies were edited manually by BioEdit 7.0.5.3 (Hall, 1999). Subsequently, all of the sequences have been uploaded to GenBank, and their accession numbers are listed in Tables 2.8-2.10.

**Table 2- 8. List of accession number of Baihuasheshecao samples is shown below.**

Code	ITS region	<i>trnH-psbA</i> intergenic spacer	<i>trnL</i> intron and <i>trnL-F</i> intergenic spacer
DS01	HQ148828	HQ148844	HQ148861
DS02	HQ148829	HQ148845	HQ148862
DS03	HQ148830	HQ148846	HQ148863
DS04	HQ148831	HQ148847	HQ148864
DS05	HQ148832	HQ148848	HQ148865
DS06	HQ148833	HQ148849	HQ148866
DS07	HQ148834	HQ148850	HQ148867
DS08	HQ148835	HQ148851	HQ148868
DS09	HQ148836	HQ148852	HQ148869
DS10	HQ148837	HQ148853	HQ148870
DS11	HQ148838	HQ148854	HQ148871
DS12	HQ148839	HQ148857	HQ148872
DS13	HQ148840	HQ148858	HQ148873
DS14	HQ148841	HQ148859	HQ148874
DS15	HQ148842	HQ148860	HQ148875

**Table 2- 9. List of accession number of field samples.**

Code	<i>trnL</i> gene and <i>trnL-F</i> intergenic spacer	<i>rps16</i> gene	<i>rbcL</i> gene	<i>matK</i> gene	ITS region	<i>trnH-psbA</i> intergenic spacer
BW06	HM752815	HM752900	HM752985	HM753070	HQ148742	HM640300
BW07	HM752816	HM752901	HM752986	HM753071	HQ148743	HM640301
BW09	HM752817	HM752902	HM752987	HM753072	HQ148744	HM640302
BW16	HM752818	HM752903	HM752988	HM753073	HQ148745	HM640303
BW17	HM752819	HM752904	HM752989	HM753074	HQ148746	HM640304
BW18	HM752820	HM752905	HM752990	HM753075	HQ148747	HM640305
BW20	HM752821	HM752906	HM752991	HM753076	HQ148748	HM640306
BW21	HM752822	HM752907	HM752992	HM753077	HQ148749	HM640307
BW22	HM752823	HM752908	HM752993	HM753078	HQ148750	HM640308
BW23	HM752824	HM752909	HM752994	HM753079	HQ148751	HM640309
BX03	HM752825	HM752910	HM752995	HM753080	HQ148752	HM640310
HA01	HM752826	HM752911	HM752996	HM753081	HQ148753	HM640311
HA02	HM752827	HM752912	HM752997	HM753082	HQ148754	HM640312
KL02	HM752828	HM752913	HM752998	HM753083	HQ148755	HM640313
KL03	HM752829	HM752914	HM752999	HM753084	HQ148756	HM640314
TA01	HM752830	HM752915	HM753000	HM753085	HQ148757	HM640315



TB01	HM752831	HM752916	HM753001	HM753086	HQ148758	HM640316
TY01	HM752832	HM752917	HM753002	HM753087	HQ148759	HM640317
TM01	HM752833	HM752918	HM753003	HM753088	HQ148760	HM640318
TM04	HM752834	HM752919	HM753004	HM753089	HQ148761	HM640319
TM05	HM752835	HM752920	HM753005	HM753090	HQ148762	HM640320
TM06	HM752836	HM752921	HM753006	HM753091	HQ148763	HM640321
TM09	HM752837	HM752922	HM753007	HM753092	HQ148764	HM640322
TM10	HM752838	HM752923	HM753008	HM753093	HQ148765	HM640323
TM11	HM752839	HM752924	HM753009	HM753094	HQ148766	HM640324
TM12	HM752840	HM752925	HM753010	HM753095	HQ148767	HM640325
TM13	HM752841	HM752926	HM753011	HM753096	HQ148768	HM640326
TM14	HM752842	HM752927	HM753012	HM753097	HQ148769	HM640327
TM15	HM752843	HM752928	HM753013	HM753098	HQ148770	HM640328
TM16	HM752844	HM752929	HM753014	HM753099	HQ148771	HM640329
TM17	HM752845	HM752930	HM753015	HM753100	HQ148772	HM640330
TM18	HM752846	HM752931	HM753016	HM753101	HQ148773	HM640331
TM19	HM752847	HM752932	HM753017	HM753102	HQ148774	HM640332
TM20	HM752848	HM752933	HM753018	HM753103	HQ148775	HM640333
TM21	HM752849	HM752934	HM753019	HM753104	HQ148776	HM640334
TM22	HM752850	HM752935	HM753020	HM753105	HQ148777	HM640335
MA01	HM752851	HM752936	HM753021	HM753106	HQ148778	HM640336
MA02	HM752852	HM752937	HM753022	HM753107	HQ148779	HM640337
MA03	HM752853	HM752938	HM753023	HM753108	HQ148780	HM640338
MA04	HM752854	HM752939	HM753024	HM753109	HQ148781	HM640339
MA05	HM752855	HM752940	HM753025	HM753110	HQ148782	HM640340
MA06	HM752856	HM752941	HM753026	HM753111	HQ148783	HM640341
MA07	HM752857	HM752942	HM753027	HM753112	HQ148784	HM640342
MA08	HM752858	HM752943	HM753028	HM753113	HQ148785	HM640343
MA09	HM752859	HM752944	HM753029	HM753114	HQ148786	HM640344
MA10	HM752860	HM752945	HM753030	HM753115	HQ148787	HM640345
MA11	HM752861	HM752946	HM753031	HM753116	HQ148788	HM640346
MA12	HM752862	HM752947	HM753032	HM753117	HQ148789	HM640347
MA13	HM752863	HM752948	HM753033	HM753118	HQ148790	HM640348
MA14	HM752864	HM752949	HM753034	HM753119	HQ148791	HM640349
MA15	HM752865	HM752950	HM753035	HM753120	HQ148792	HM640350
MA16	HM752866	HM752951	HM753036	HM753121	HQ148793	HM640351
MA17	HM752867	HM752952	HM753037	HM753122	HQ148794	HM640352
MA18	HM752868	HM752953	HM753038	HM753123	HQ148795	HM640353
MA19	HM752869	HM752954	HM753039	HM753124	HQ148796	HM640354
MA20	HM752870	HM752955	HM753040	HM753125	HQ148797	HM640355
MA21	HM752871	HM752956	HM753041	HM753126	HQ148798	HM640356
MA22	HM752872	HM752957	HM753042	HM753127	HQ148799	HM640357
MA23	HM752873	HM752958	HM753043	HM753128	HQ148800	HM640358
MA24	HM752874	HM752959	HM753044	HM753129	HQ148801	HM640359



MA25	HM752875	HM752960	HM753045	HM753130	HQ148802	HM640360
MA26	HM752876	HM752961	HM753046	HM753131	HQ148803	HM640361
MA27	HM752877	HM752962	HM753047	HM753132	HQ148804	HM640362
YU01	HM752878	HM752963	HM753048	HM753133	HQ148805	HM640363
YU02	HM752879	HM752964	HM753049	HM753134	HQ148806	HM640364
YU03	HM752880	HM752965	HM753050	HM753135	HQ148807	HM640365
YU04	HM752881	HM752966	HM753051	HM753136	HQ148808	HM640366
YU05	HM752882	HM752967	HM753052	HM753137	HQ148809	HM640367
YU06	HM752883	HM752968	HM753053	HM753138	HQ148810	HM640368
YU07	HM752884	HM752969	HM753054	HM753139	HQ148811	HM640369
YU09	HM752885	HM752970	HM753055	HM753140	HQ148812	HM640370
YU10	HM752886	HM752971	HM753056	HM753141	HQ148813	HM640371
YU11	HM752887	HM752972	HM753057	HM753142	HQ148814	HM640372
YU12	HM752888	HM752973	HM753058	HM753143	HQ148815	HM640373
YU13	HM752889	HM752974	HM753059	HM753144	HQ148816	HM640374
YU14	HM752890	HM752975	HM753060	HM753145	HQ148817	HM640375
YU15	HM752891	HM752976	HM753061	HM753146	HQ148818	HM640376
YU16	HM752892	HM752977	HM753062	HM753147	HQ148819	HM640377
YU17	HM752893	HM752978	HM753063	HM753148	HQ148820	HM640378
YU19	HM752894	HM752979	HM753064	HM753149	HQ148821	HM640379
YU20	HM752895	HM752980	HM753065	HM753150	HQ148822	HM640380
YU21	HM752896	HM752981	HM753066	HM753151	HQ148823	HM640381
YU22	HM752897	HM752982	HM753067	HM753152	HQ148824	HM640382
YU23	HM752898	HM752983	HM753068	HM753153	HQ148825	HM640383
YU24	HM752899	HM752984	HM753069	HM753154	HQ148826	HM640384
YU26	/	/	/	HQ225735	HQ148827	/
BW03	HQ148876	/	/	/	HQ148843	/
BW19A	HQ148877	/	/	/	HM622084	HQ225736
BW19	HQ148878	/	/	/	HM622085	HQ225737



**Table 2- 10. Accession No. of 12S and 16S regions of icefishes.**

Code	12S region	16S region
D01	HM622088	HM622115
D02	HM622089	HM622114
D03	HM622090	HM622113
D04	HM622077	HM622099
D05	HM622082	HM622100
D06	HM622083	HM622101
F01	HM622094	HM622112
F02	HM622095	HM622111
F03	HM622091	HM622110
F04	HM622092	HM622108
F05	HM622096	HM622109
F06	HM622080	HM622105
F07	HM622075	HM622106
F08	HM622078	HM622102
F09	HM622076	HM622107
F10	HM622093	HM622116
F11	HM622081	HM622103
F12	HM622079	HM622104
S01	HM622097	HM622118
S04	HM622098	HM622117

For the parts of Traditional Chinese Medicine and seafood authentication, FINS approach with maximum parsimony analysis was employed to differentiate genuine species. In this study, the maximum parsimony tree ( Nei and Kumar 2000) was conducted in MEGA version 5.0 (Tamura et al. 2011) with bootstrap method (Felsenstein 1985) of 1000 replicates. The evolutionary divergence between sequences was calculated by Mega version 5.0 (Tamura et al. 2011) with Kimura-two parameter (K2P) distance model (Kimura 1980). For the species identification, BLAST searches revealed the identities of samples, with GenBank used for the mitochondrial 16s rRNA gene.

The different sets of sequences for phylogenetic study of *Hedyotis* were aligned by

ClustalX version 2.0 (Larkin et al., 2007) with default parameters. And then the alignment result was put to the computer program of SeqState version 1.4 (Müller, 2005, 2006) for scoring of the gap characters by the simple indel coding method (Simmons and Ochoterena, 2000). Subsequently, parsimony analysis was done by PAUP\* 4.0b10 (Swofford, 2003), with heuristic searches (addition sequences search with 1000 replicates) and the tree-bisection-reconnection (TBR) branch swapping algorithm. The 1000 replicates of jackknife analysis (Farris et al., 1996) was also conducted by PAUP\* with an average proportion of 36.7879% of the characters deleted in each replicate (Farris et al., 1996), and the heuristic search option with one random addition sequence and TBR branch swapping algorithm. The cladograms were edited by TreeGraph version 2.0 (Stöver and Müller, 2010). The maximum likelihood trees were conducted by Mega 5.0 (Tamura et al. 2011) with 500 bootstrap replicates.



## **CHAPTER 3**

# **USING FORENSICALLY INFORMATIVE NUCLEOTIDE SEQUENCING (FINS) TECHNOLOGY FOR SPECIES IDENTIFICATION**

### 3.1 Authentication of Baihuasheshecao by FINS Analysis

FINS technology is a DNA sequence-based analysis method for the identification of unknown or suspicious biological samples. FINS analysis directly assesses genetic variations and phylogenetic relationships among reference species and testing samples in which the resolution of FINS analysis depends on the sequence variation of selected DNA regions. Therefore, the choice of DNA regions is crucial for identification using FINS approach. In this study, we have used three DNA regions from the nuclear and plastid genomes, namely nuclear ITS, plastid *trnH-psbA*, and plastid *trnL-F* regions for the authentication of Baihuasheshecao commodities using FINS approach. The sequences alignment matrices are shown in **Appendices 6- 1 – 6- 3**.

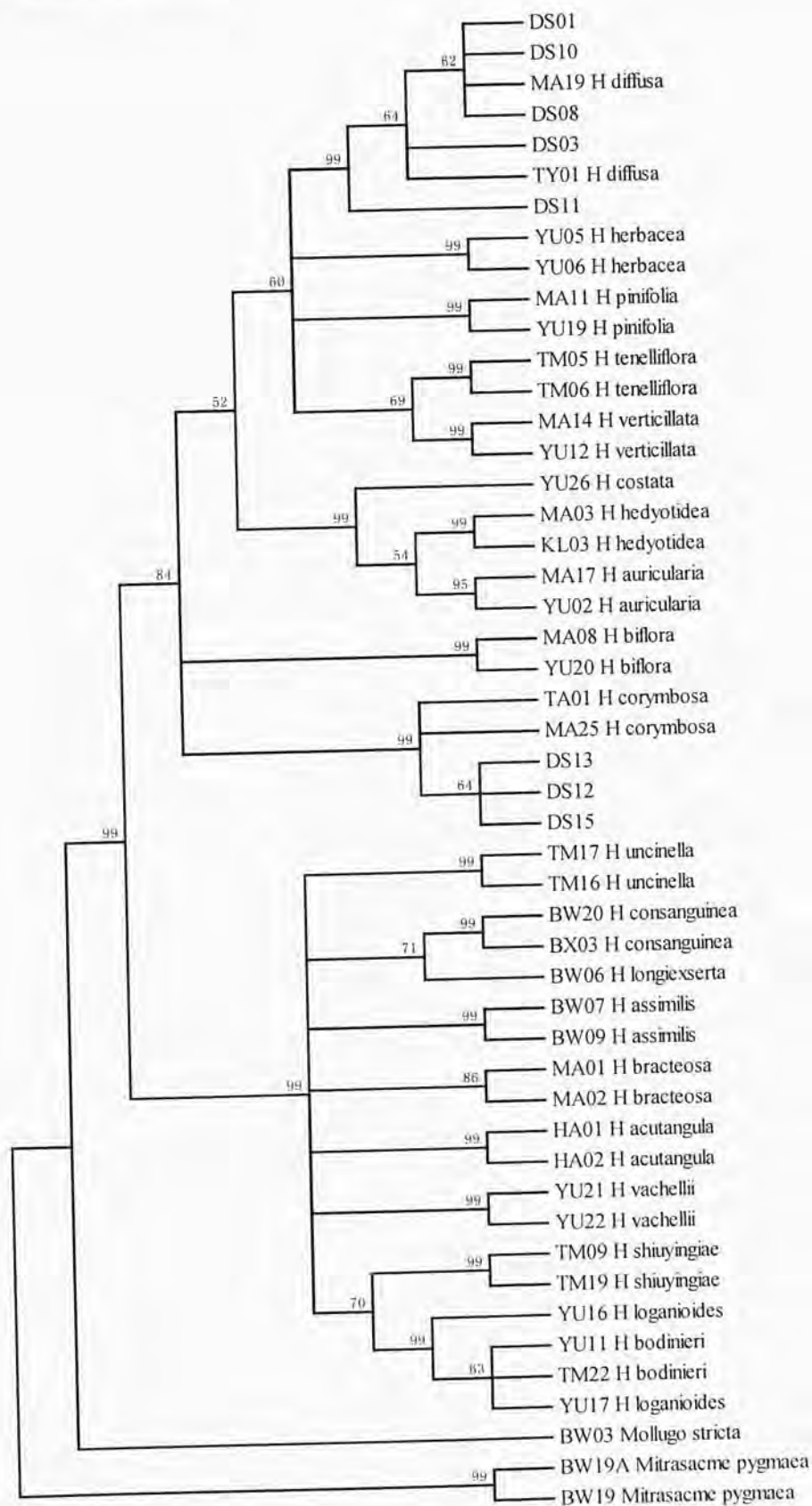
#### 3.1.1 Authentication using FINS technology

The consensus maximum parsimony tree constructed based on ITS region showed that *H. diffusa* (99%), *H. corymbosa* (99%), *H. pinifolia* (99%) and *H. tenelliflora* (99%) formed distinct clades with the support of high bootstrap frequencies (**Figure 3-1**). FINS analysis indicated that five Baihuasheshecao commodities collected from Zhejiang (DS01), Tianjin (DS03), Hunan (DS08), Washington State (DS10) and Guangdong (DS11) were genuine herb grouping with *H. diffusa*. In contrast, three commodities collected from Hong Kong (DS12 and DS13) and Boston (DS15) were adulterants derived from *H. corymbosa* (**Figure 3-1**). Similar conclusion was drawn



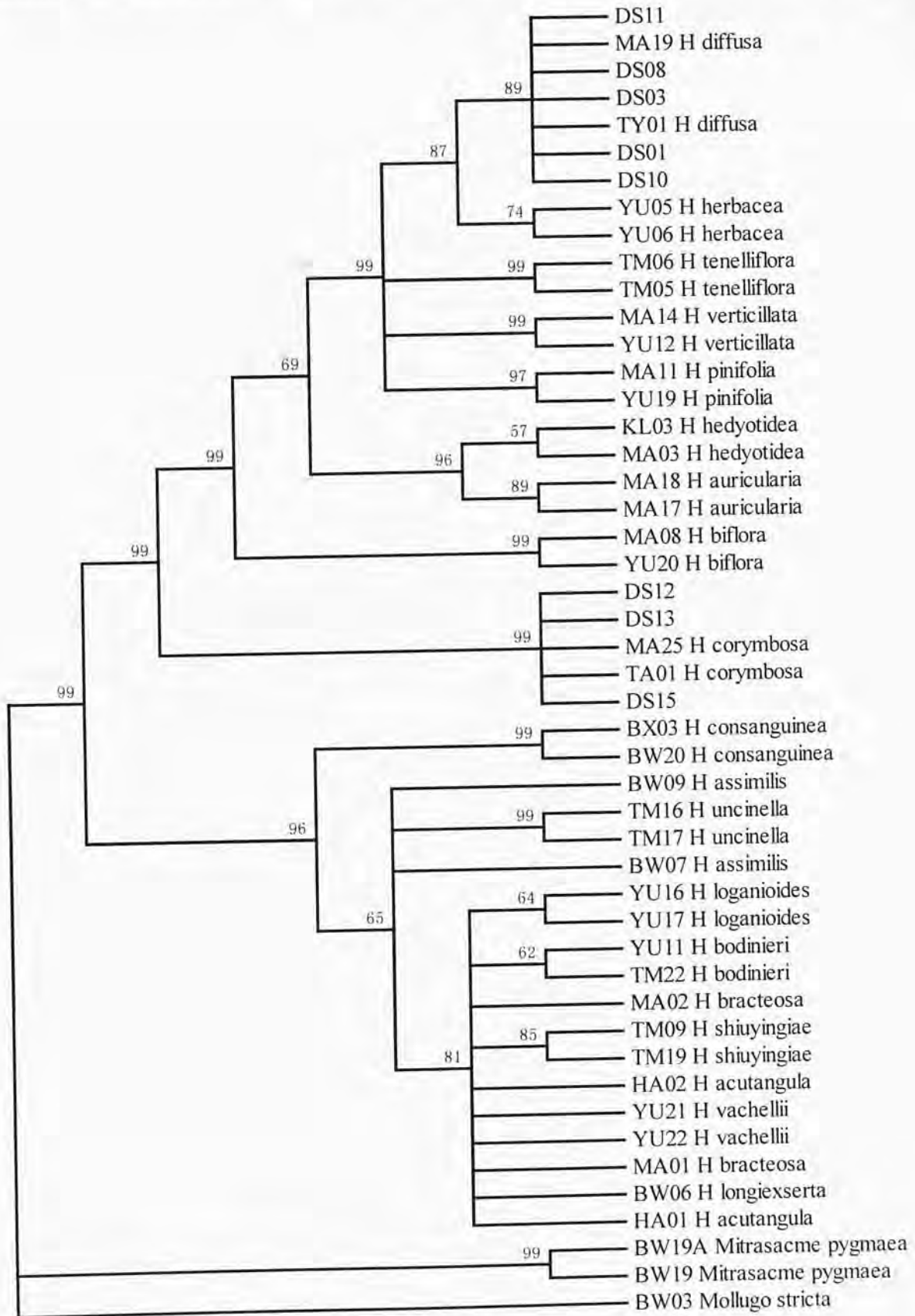
from the consensus maximum parsimony tree of *trnL-F* region that the distinct clades of *H. diffusa* (89%), *H. corymbosa* (99%), *H. pinifolia* (97%) and *H. tenelliflora* (99%) were well supported by high bootstrap support, and the Baihuasheshecao commodities grouped with *H. diffusa* (DS01, DS03, DS08, DS10 and DS 11) and *H. corymbosa* (DS12, DS13 and DS15), respectively (**Figure 3-2**). For the *trnH-psbA* region, *H. diffusa* (65%) formed a clade with weak support which resolved with *H. herbacea* forming a weakly supported clade (51%) (**Figure 3-3**). Although the clades of *H. corymbosa* (99%) and *H. pinifolia* (98%) are strongly supported by bootstrap assessment, the clade of *H. tenelliflora* was weakly supported with bootstrap frequency of 57% (**Figure 3-3**). FINS analysis based on *trnH-psbA* region indicated that three Baihuasheshecao commodities (DS12, DS13 and DS15) are adulterants derived from *H. corymbosa*. However, the identities of five Baihuasheshecao commodities (DS01, DS03, DS08, DS10 and DS 11) were less defined as these samples might be derived from *H. diffusa* or *H. herbacea* (**Figure 3-3**). Despite this unsolved gathering of *H. diffusa* or *H. herbacea*, the possibility of *H. herbacea* being used as an adulterant of Baihuasheshecao is low because of the scarce and rare distribution of *H. herbacea* in the field. *H. herbacea* is morphologically similar to *H. diffusa*, which differs only by sparse soft hair on the pedicel in the former species. Our data suggested that these two species can be discriminated by both ITS and *trnL-F* regions independently of each other.

**Figure 3- 1. Phylogenetic tree based on nuclear ITS region constructed using maximum parsimony method.** A total of 49 taxa comprised with 38 plant samples of 20 *Hedyotis* species, 3 plant samples of 2 outgroup species, and 8 herb samples of Baihuasheshecao commodity were included. Fifty-percent majority-rule consensus represented the maximum parsimony tree constructed with 595 steps based on 276 informative sites in 750 characters based on 276 informative sites (ensemble consistency index of 0.66 and ensemble retention index of 0.90) was accessed by bootstrap frequencies (1000 replications) with cut-off value of 50% are indicated above branches.

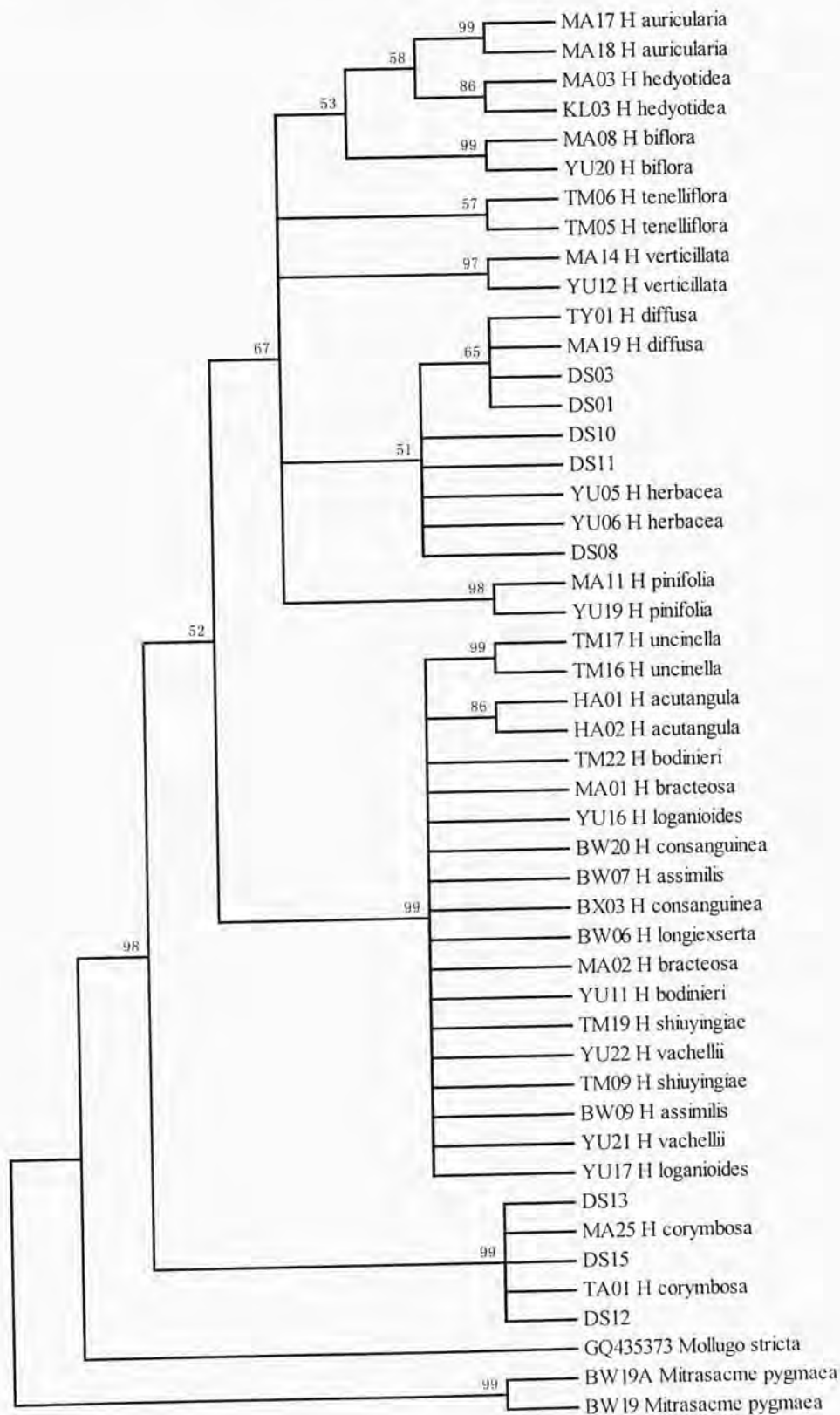




**Figure 3- 2. Phylogenetic tree based on plastid *trnL-F* region constructed using maximum parsimony method.** A total of 48 taxa comprised with 37 plant samples of 19 *Hedyotis* species, 3 plant samples of 2 outgroup species, and 8 herb samples of Baihuasheshecao commodity were included. Fifty-percent majority-rule consensus represented the maximum parsimony tree constructed with 487 steps based on 335 informative sites in 1068 characters (ensemble consistency index of 0.89 and ensemble retention index of 0.97). The bootstrap frequencies (1000 replications) with cut-off value of 50% are indicated above branches.



**Figure 3- 3. Phylogenetic tree based on plastid *trnH-psbA* region constructed using maximum parsimony method.** A total of 48 taxa comprised with 37 plant samples of 19 *Hedyotis* species, 2 plant samples of 1 outgroup species, 1 downloaded sequence of 1 outgroup species, and 8 herb samples of Baihuasheshecao commodity were included. Fifty-percent majority-rule consensus represented the maximum parsimony tree constructed with 190 steps based on 299 informative sites in 531 characters (ensemble consistency index of 0.82 and ensemble retention index of 0.94). The bootstrap frequencies (1000 replications) with cut-off value of 50% are indicated above branches.





### 3.1.2 Relative effectiveness of DNA regions for FINS analysis

The most crucial factor for successful FINS analysis is the choice of suitable DNA region, which should have sufficient discrimination power and should be easily amplified, sequenced and analyzed.

In this study, each species needs to be resolved as a monophyletic group separate from all other species in any of three MP trees in order to be discriminated. In three MP trees, FINS analysis showed that ITS region was able to discriminate 18 *Hedyotis* species (90.0%), while the *trnL-F* and *trnH-psbA* regions were able to discriminate 15 (78.9%) and 10 (52.6%) *Hedyotis* species respectively (**Figures 3- 1 - 3- 3**). ITS and *trnH-psbA* regions were selected as complementary DNA barcodes for land plants due to their discrimination power (CBOL Plant Working Group, 2009; Kress et al., 2005). Our data indicate that the ITS region showed the highest discrimination power for most *Hedyotis* species (excepted *H. bodinieri* and *H. longanioides*) among the DNA regions studied (**Figure 3- 1**). In contrast, *trnH-psbA* region discriminated only 10 *Hedyotis* species while *H. assimilis*, *H. herbacea*, *H. bodinieri*, *H. bracteosa*, *H. consanguinea*, *H. herbacea*, *H. longiexserta*, *H. shiuyingiae* and *H. vachellii* were not discriminated (**Figure 3.3**). The *trnL-F* region is another locus frequently used for identification of medicinal plants (Li et al., 2011). It showed moderate discrimination rate and failed to discriminate *H. acutangula*, *H. bracteosa*, *H. longiexserta* and *H. vachellii* (**Figure 3- 2**). Our data also revealed that *H. herbacea*, *H. consanguinea* and *H. vachellii* could only be discriminated by ITS

region and *H. bodinieri* and *H. longanioides* could only be discriminated by *trnL-F* region. Although ITS regions showed the highest discrimination power, *trnL-F* region resolved the phylogenetic relationship among most *Hedyotis* species (Figure 3- 2). Previous report showed that combined use of DNA regions could provide a better discrimination rate (Kress and Erickson, 2007). It may be necessary to integrate ITS and *trnL-F* regions for a better discrimination of all the *Hedyotis* species.

Amplification of the three regions was generally easy for most of the samples in this study. The amplicon sizes were about 700 bp, 1000 bp, and 400 bp for ITS, *trnL-F*, and *trnH-psbA*, respectively. For herbal commodities, DNA degradation may lead to insufficient DNA quality and quantity for subsequent molecular identification. Therefore, DNA region of large size, such as *trnL-F*, may lead to unsuccessful amplification in commodity samples. Internal PCR primers may be necessary to generate the complete *trnL-F* region in this case (Taberlet et al., 1991). Besides, fungal contamination is frequently found in herbal commodity which is improperly processed and stored that may generate false positive amplification results for ITS region. In this case, plant-specific primers, such as RUT-ITSF1, should be used. For DNA sequencing, *trnL-F* region usually generates high quality sequencing results. The presence of multiple copies, pseudogenes, and secondary structure of ITS region sometimes lead to poor quality of sequence data (Baldwin et al., 1995; Alvarez & Wendel, 2003; Stuessy, 2009). The *trnH-psbA* region sometimes does not generate bidirectional unambiguous sequences (CBOL Plant Working Group, 2009). The



presence of a poly-A/T structure lowers the successful rate of DNA sequencing (Zhu et al., 2010). Although molecular cloning of *trnH-psbA* region prior to DNA sequencing can improve the DNA sequencing results, the presence of large amount of indels makes sequence alignment and data analysis difficult. In balance of discrimination power and generation of quality DNA sequence, *trnL-F* region is the most suitable DNA region for identification of Baihuasheshecao using FINS approach.

### **3.1.3 Phylogenetic interpretation**

For taxonomic classification of *Hedyotis*, Luo classified *Hedyotis* to five sections based on the types of capsule dehiscence in *Flora Republicae Popularis Sinicae* (Luo, 1999). The five sections are sect. *Hedyotis*, sect. *Diplophragma* Wight et Arn., sect. *Dimetia* Wight et Arn., sect. *Euoldenlandia* K. Schum. and sect. *Gonothecha* Meissn (Luo, 1999). The representative species of each section are given in Figure 1.2. Section *Hedyotis* is characterized by indehiscent capsule, while *Diplophragma* and *Dimetia* are distinct by septicidal dehiscent capsule. The capsule of *Dimetia* protrudes at apex. *Euoldenlandia* and *Gonothecha* are specified by loculicidal dehiscent capsule. This classification places much emphasis on using the site of dehiscence in the fruits as a key character in subdividing the genus into various sections and subgroups. The dehiscence position of seed, however, can be varied and subjected to adaptations to the living environment such as dispersal strategies. In 1954, Fosberg pointed out that the differences in seeds and fruits were not

appropriate for defining genera (Fosberg and Terrell, 1985). In 1958, Verdcourt suggested that although the nature of dehiscence of the seed is of more important than the hardness or softness of the pericarp, using the character of seed dehiscence or indehiscence for group classification had numerous exceptions, and it was one of the most misleading characters (Verdcourt, 1958).

In 1961, Lewis also stated that the seed character cannot be a diagnostic character in clearly differentiating *Hedyotis* species (Lewis, 1961). In order to provide more information for the revision of *Hedyotis* in *Flora of China* and understanding the evolution of *Hedyotis* in China, molecular data are necessary for the inference of the phylogenetic relationship of the *Hedyotis*. In this study, ten species in *Diplophragma* formed a strongly supported clade (96%-99% bootstrap support based on three single-locus trees). However, the resolution of these trees was insufficient to resolve the phylogenetic relationship of all the species within this section. Only *trnL-F* region partially resolve the relationship among *H. assimilis*, *H. consanguinea* and *H. uncinella* (Figure 3.2). All the maximum parsimony trees based on the three DNA regions showed that the species in section *Hedyotis* with indehiscent capsule (*H. auricularia* and *H. costata*) clustered with the species in section *Hedyotis* with loculicidal dehiscent capsule (*H. teneflora*, *H. pinifolia* and *H. verticillata*), and species in section *Dimetia* with septicidal capsule with protrude at apex (*H. hedyotideae*), together with those in section *Euoldenlandia* with loculicidal dehiscent capsule (*H. diffusa* and *H. herbacea*). This finding was in agreement with our previous report that nine *Hedyotis* species from various sections cluster together



forming a monophyletic group independent to classification pattern (Li et al., 2010). Our data suggested that there is a need to re-assess the morphological characters for taxonomic classification of *Hedyoits* in China.

## **3.2 Authentication of Salangids (Icefishes) by FINS Analysis**

### **3.2.1 Analysis based on mitochondrial ribosome DNA region**

PCR amplification and DNA sequencing of two mitochondrial ribosome genes were successful for all 20 samples. The sequences of 12s rDNA and 16s rDNA were about 330 bp and 410 bp, respectively, and they were all uploaded to GenBank with accession numbers were shown in **Table 2-10**.

#### **3.2.1.1 Blast analysis with 16S region**

After homology search with the sample 16s rDNAs, S01 and S04 were of Family Engraulidae in Order Clupeoidei. S04 sequence was 99% similar, the top BLAST hit with that of *Encrasicholina heteroloba* (AB246183), and S01 only has 92% similarity. 16s rDNA sequences of D01, D02, D03, F01, F02, F03, F04, F05 and F10 perfectly matched with those of *Neosalanx taihuensis* (AY279364) and *Neosalanx jordani* (AY279365), but only 98% (highest match rate) matched with *Neosalanx tangkahkeii* (EU419753). F07 and F09 samples were 99% (highest match rate) matched with *Hemisanx brachyrostralis* (AY279363). D04 and D05, D06, F08, F11 and F12 were 100% matched with *Protosalanx chinensis* (AY279361), while F06 showed slightly lower 99% similarity.

### 3.2.1.2 Analysis based on mitochondrial 12S rRNA

The sequences of 12s region were aligned by CLUSTALX (2.0) software and the sequences alignment matrix was shown in **Appendix Table 6- 4**.

Differences between sequences were also calculated using a statistical distance method (Kimura 2-parameter) by software MEGA version 4.1. The distance matrix of 12s region is shown in **Table 3- 1**. The sequence distance of S01 and S04 to icefishes ranged from 0.232 to 0.276. The 12s rDNA sequences distance between F07 and F09 was 0.006, and those of D01, D02, D03, F01, F02, F03, F04, F05 and F10 were zero. Distance among the samples D04, F06, F08, F11 and F12 ranged from zero to 0.003 within the intraspecific level (0–0.004) whereas the distance of D05 and D06 to D04, F06, F08, F11 and F12 ranged from 0.003 to 0.009. The distance of *Neosalanx* group to *Protosalanx* group and *Hemisanx* group was 0.047-0.050 and 0.050-0.053, respectively. The distance between *Protosalanx* group and *Hemisanx* group ranged from 0.048 to 0.051.



**Table 3- 1. Estimates of evolutionary divergence of the 12s rRNA sequences.** All results are based on the pairwise analysis of 20 sequences and other sequences from GenBank (S\_m: *Salangichthys microdo*; H\_v: *Hyperlophus vittatus*). Analyses were conducted using the Kimura 2-parameter method in MEGA5.0. Details of samples were listed in **Figure 3- 4.**

	F07	F09	D04	F08	F12	F06	F11	D05	D06	D01	D02	D03	F03	F04	F10	F01	F02	F05	S_m	S01	S04	H_v	
F07																							
F09	0.006																						
D04	0.057	0.057																					
F08	0.057	0.057	0.000																				
F12	0.057	0.057	0.000	0.000																			
F06	0.060	0.060	0.003	0.003	0.003																		
F11	0.053	0.053	0.003	0.003	0.003	0.006																	
D05	0.057	0.057	0.006	0.006	0.006	0.009	0.009																
D06	0.053	0.053	0.003	0.003	0.003	0.006	0.006	0.003															
D01	0.053	0.050	0.047	0.047	0.047	0.050	0.050	0.047	0.044														
D02	0.053	0.050	0.047	0.047	0.047	0.050	0.050	0.047	0.044	0.000													
D03	0.053	0.050	0.047	0.047	0.047	0.050	0.050	0.047	0.044	0.000	0.000												
F03	0.053	0.050	0.047	0.047	0.047	0.050	0.050	0.047	0.044	0.000	0.000	0.000											
F04	0.053	0.050	0.047	0.047	0.047	0.050	0.050	0.047	0.044	0.000	0.000	0.000	0.000										
F10	0.053	0.050	0.047	0.047	0.047	0.050	0.050	0.047	0.044	0.000	0.000	0.000	0.000	0.000									
F01	0.053	0.050	0.047	0.047	0.047	0.050	0.050	0.047	0.044	0.000	0.000	0.000	0.000	0.000	0.000								
F02	0.053	0.050	0.047	0.047	0.047	0.050	0.050	0.047	0.044	0.000	0.000	0.000	0.000	0.000	0.000	0.000							
F05	0.053	0.050	0.047	0.047	0.047	0.050	0.050	0.047	0.044	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000						
S_m	0.060	0.066	0.076	0.076	0.076	0.080	0.080	0.070	0.073	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090	0.090
S01	0.247	0.252	0.265	0.265	0.265	0.260	0.260	0.270	0.270	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.276	0.231
S04	0.245	0.236	0.244	0.244	0.244	0.239	0.239	0.248	0.248	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.244	0.252
H_v	0.213	0.213	0.217	0.217	0.217	0.212	0.212	0.208	0.212	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.208	0.156

### 3.2.2 Analysis based on mitochondrial 16S rRNA

The sequences from 16s region were aligned in ClustalX (2.0) software with the sequence alignment matrix shown in **Appendix Table 6- 5**.

Differences between sequences were also calculated using a statistical distance method (Kimura 2-parameter) by software MEGA version 4.1. The distance matrix made by 16s region was shown in **Table 3- 2**.

The sequence distance of S01 and S04 to other icefishes ranged from 0.275 to 0.330. In contrast, S04 sequence was distanced from that of *Encrasicholina heteroloba* by 0.002 while S01 to S04 showed greater distances of 0.078 and 0.081, respectively. The sequence distance of *Hemisanx brachyrostralis* to F07 and F09 was 0.005, within the intraspecific variation (0–0.004). As with 12s rDNAs, the 16s rDNA sequences of D01, D02, D03, F01, F02, F03, F04, F05, F10, *Neosalanx taihuensis* and *Neosalanx jordani* were identical with zero distances (**Tables 3- 2 – 3- 3**). *Neosalanx tangkahkeii* was 0.015 different from them. Sequence distances among the samples of D04, D05, D06, F08, F11, F12 and *Protosalanx chinensis* were zero. The sequence distance of F06 to D04, D05, D06, F08, F11, F12 and *Protosalanx chinensis* was 0.002 within the intraspecific level. Thus the results support the genus groupings of *Neosalanx*, *Protosalanx* and *Hemisanx*. The distance of *Neosalanx* group to *Protosalanx* group and *Hemisanx* group was 0.033 and 0.059, respectively. The distance between *Protosalanx* group and *Hemisanx* group ranged from 0.044 to 0.046.



**Table 3- 2. Estimates of evolutionary divergence of 16s rRNA sequences.** All results are based on the pairwise analysis of 20 sequences and other sequences from GenBank (H\_b: *Hemisalix brachystralis*; N\_tai: *Neosalix taihuensis*; N\_tan: *Neosalix tangkahkeii*; N\_j: *Neosalix jordani*; P\_c: *Protosalix chinensis*; E\_h: *Encrasicholina heteroloba*). Analyses were conducted using the Kimura 2-parameter method in MEGA 5.0. Details of samples were listed in **Figure 3- 5.**

	F07	F09	H_b	D04	D05	D06	F08	F11	F12	P_c	F06	N_tai	N_tan	N_j	F05	F04	F03	F02	F01	D03	D02	F10	D01	S04	E_h	S01	
F07																											
F09	0.000																										
H_b	0.005	0.005																									
D04	0.044	0.044	0.038																								
D05	0.044	0.044	0.038	0.000																							
D06	0.044	0.044	0.038	0.000	0.000																						
F08	0.044	0.044	0.038	0.000	0.000	0.000																					
F11	0.044	0.044	0.038	0.000	0.000	0.000	0.000																				
F12	0.044	0.044	0.038	0.000	0.000	0.000	0.000	0.000																			
P_c	0.044	0.044	0.038	0.000	0.000	0.000	0.000	0.000	0.000																		
F06	0.046	0.046	0.041	0.002	0.002	0.002	0.002	0.002	0.002	0.002																	
N_tai	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033																
N_tan	0.076	0.076	0.073	0.051	0.051	0.051	0.051	0.051	0.051	0.051	0.049	0.015															
N_j	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015														
F05	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015	0.000													
F04	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015	0.000	0.000												
F03	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015	0.000	0.000	0.000											
F02	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015	0.000	0.000	0.000	0.000										
F01	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015	0.000	0.000	0.000	0.000	0.000									
D03	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000								
D02	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000							
F10	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000						
D01	0.059	0.059	0.057	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.033	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
S04	0.283	0.283	0.291	0.306	0.306	0.306	0.306	0.306	0.306	0.306	0.302	0.306	0.330	0.306	0.306	0.306	0.306	0.306	0.306	0.306	0.306	0.306	0.306	0.306	0.306	0.306	0.306
E_h	0.279	0.279	0.287	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.298	0.302	0.325	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.302	0.002
S01	0.290	0.290	0.298	0.304	0.304	0.304	0.304	0.304	0.304	0.304	0.300	0.299	0.322	0.299	0.299	0.299	0.299	0.299	0.299	0.299	0.299	0.299	0.299	0.299	0.299	0.078	0.081

### 3.2.3 Analysis based on combined regions

Combined datasets' sequences were aligned by ClustalX (2.0) software and the sequences alignment matrix is shown in **Appendix Table 6- 6**. Differences between sequences were also calculated using a statistical distance method (Kimura 2-parameter) by software MEGA version 4.1., and the distance matrix made by the combined two regions is shown in **Table 3- 3**.

The sequence distance of S01 and S04 to icefishes ranged from 0.275 to 0.293. The sequence distance between F07 and F09 was 0.003, within the intraspecific variation (0–0.004). Samples D01, D02, D03, F01, F02, F03, F04, F05 and F10 were of the same species as their combined sequence distances were zero. Samples D04, D05, D06, F06, F08, F11, F12 were conspecific with combined sequence distances ranged from zero to 0.005. The distances from *Neosalanx* group to *Protosalanx* group and *Hemisanx* group were 0.039-0.042 and 0.056, respectively. The distance between *Protosalanx* group and *Hemisanx* group ranged from 0.048 to 0.051.



**Table 3- 3. Estimates of evolutionary divergence of the 12s rRNA and 16s rRNA sequences.** All results are based on the pairwise analysis of 20 sequences. Analyses were conducted using the Kimura 2-parameter method in MEGA5.0 Details of samples were listed in **Figure 3-4 – 3-6.**

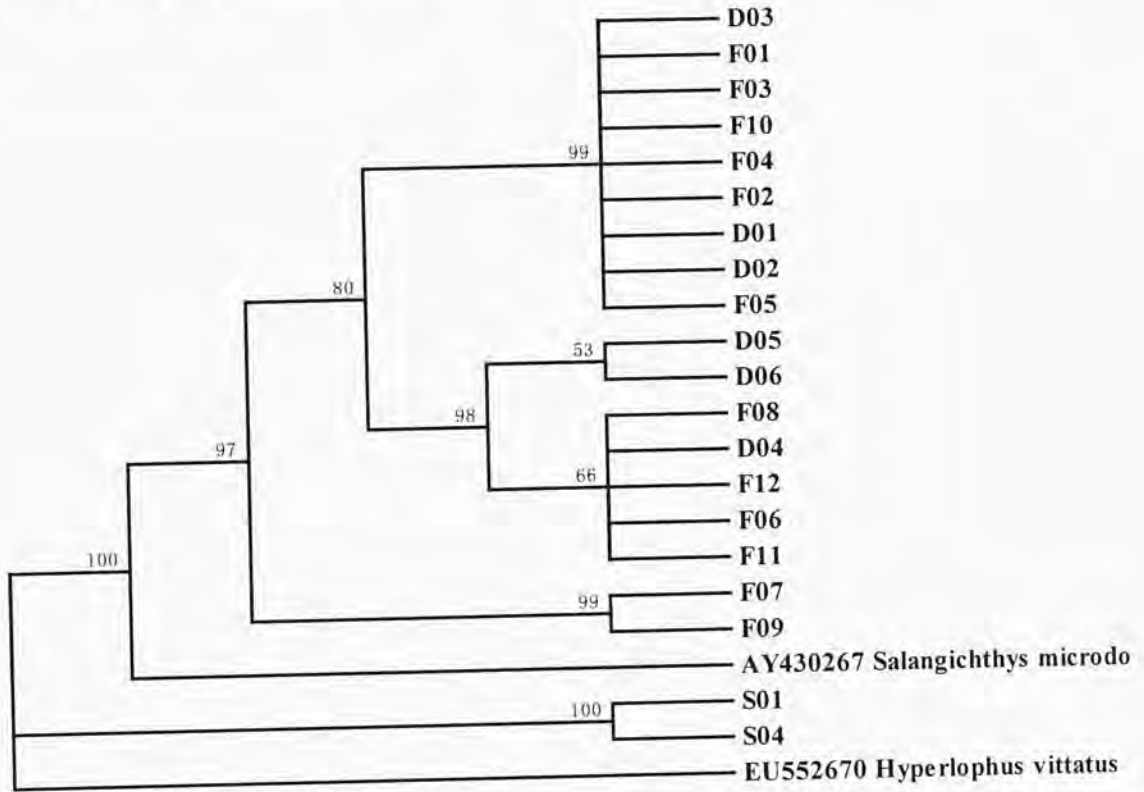
	F07	F09	D04	F08	F12	F11	D05	D06	F06	D01	D02	F04	F05	F03	F02	F01	F10	D03	S01	S04
F07																				
F09	0.003																			
D04	0.048	0.048																		
F08	0.048	0.048	0.000																	
F12	0.048	0.048	0.000	0.000																
F11	0.046	0.046	0.001	0.001	0.001															
D05	0.048	0.048	0.003	0.003	0.003	0.004														
D06	0.046	0.046	0.001	0.001	0.001	0.003	0.001													
F06	0.051	0.051	0.003	0.003	0.003	0.004	0.005	0.004												
D01	0.056	0.055	0.040	0.040	0.040	0.042	0.040	0.039	0.040											
D02	0.056	0.055	0.040	0.040	0.040	0.042	0.040	0.039	0.040	0.000										
F04	0.056	0.055	0.040	0.040	0.040	0.042	0.040	0.039	0.040	0.000	0.000									
F05	0.056	0.055	0.040	0.040	0.040	0.042	0.040	0.039	0.040	0.000	0.000	0.000								
F03	0.056	0.055	0.040	0.040	0.040	0.042	0.040	0.039	0.040	0.000	0.000	0.000	0.000							
F02	0.056	0.055	0.040	0.040	0.040	0.042	0.040	0.039	0.040	0.000	0.000	0.000	0.000	0.000						
F01	0.056	0.055	0.040	0.040	0.040	0.042	0.040	0.039	0.040	0.000	0.000	0.000	0.000	0.000	0.000					
F10	0.056	0.055	0.040	0.040	0.040	0.042	0.040	0.039	0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
D03	0.056	0.055	0.040	0.040	0.040	0.042	0.040	0.039	0.040	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
S01	0.275	0.277	0.289	0.289	0.289	0.287	0.292	0.292	0.285	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.288
S04	0.279	0.275	0.290	0.290	0.290	0.288	0.293	0.293	0.286	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.288	0.081

### **3.2.4 Phylogenetic analysis**

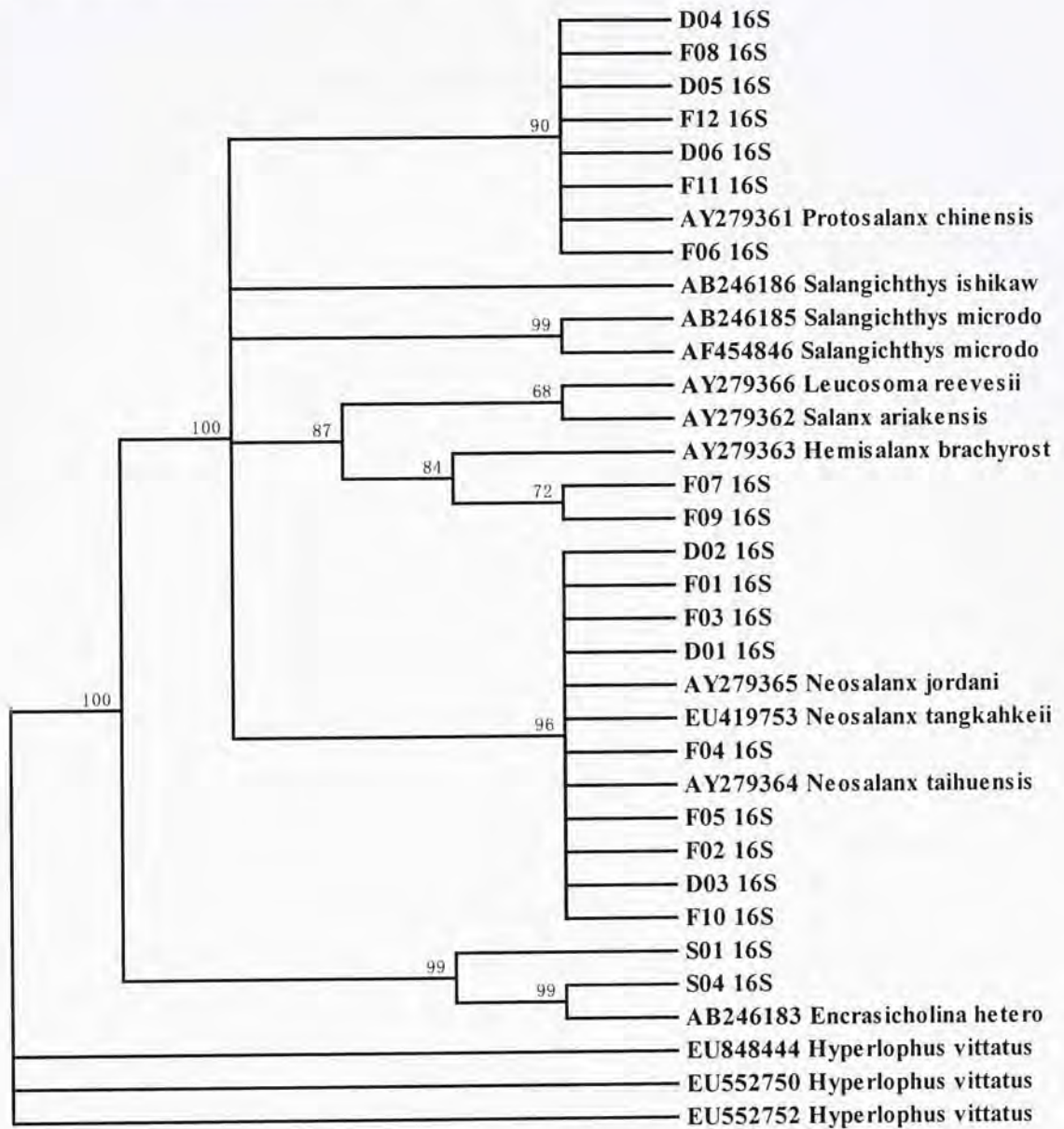
Sequences of 12s rRNA region and 16s rRNA regions and their combined sequences were used to construct three maximum parsimony (MP) trees (**Figures 3-4 – 3-6**) with 1000 bootstrap replicates. The MP trees based on two single-region MP trees and one combined 12s and 16s trees clearly reveals that some samples (D01, D02, D03, F01, F02, F03, F04, F05 and F10) were clustered together with strong bootstrap support (99%-100% bootstrap value), some samples (F07, F09) belonged to another group with 99%-100% bootstrap support, some samples (D04, D05, D06, F06, F08, F11, F12) were gathered with strong bootstrap support and the S01 and S04 samples which united together forming another clade.



**Figure 3- 4. Phylogenetic tree based on 12s rRNA sequences constructed using maximum parsimony method.** The sequences download from GenBank were specified in this figure. The bootstrap values were indicated below branches (50% cut-off value for consensus tree).

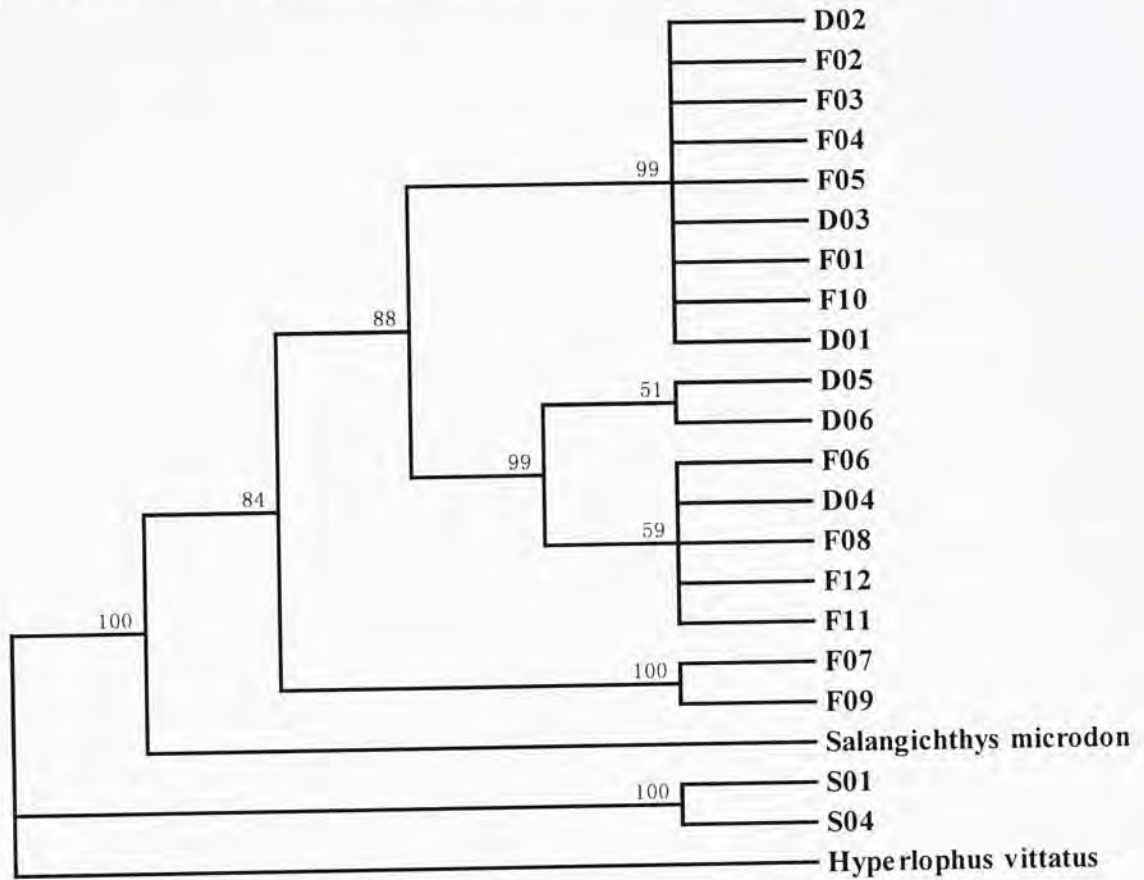


**Figure 3- 5. Phylogenetic tree based on 16s rRNA sequences constructed using maximum parsimony method.** The sequences download from GenBank were specified in this figure. The bootstrap values were indicated below branches (50% cut-off value for consensus tree).





**Figure 3- 6. Phylogenetic tree based on combined 12s rRNA and 16s rRNA sequences constructed using maximum parsimony method from 20 fish samples. The sequences download from GenBank were specified in this figure. The bootstrap values were indicated below branches (50% cut-off value for consensus tree).**



### 3.2.5 Discussion

#### 3.2.5.1 16s rDNAs and 12s rDNAs

In this study, the 16s rRNA and 12s rRNA region have a lower power in inferring the relationship within genus *Neosalanx*, with 100% bootstrap value and zero interspecific genetic distance for all the sequences of D01, D02, D03, F01, F02, F03, F04, F05, F10, *Neosalanx taihuensis* and *Neosalanx jordani*. This may be because the two genes are too conserved not suitable for species differentiation or proliferation of names is assigned to the same evolutionary lineage (Miya and Nishida 1996; Fu et al. 2005; Zhang et al. 2007). Similar results were obtained by Zhang and his colleagues (2007) whose phylogenetic analysis based on cytochrome *b* sequences showed that *Neosalanx tangkahkeii*, *Neosalanx taihuensis* and *Neosalanx pseudotaihuensis* were always clustered together with 100% support, with 0.004–0.006 genetic distance. Fu and his colleagues (2005) used partial DNA sequences of cytochrome *b* and 16s mitochondrial genes to determine the phylogenetic placement of salangid fishes and the generic relationships within the salangids. The result found that the sequence divergence of *Neosalanx taihuensis* and *Neosalanx jordani* is zero (Fu and others 2005). In later study, sequences of cytochrome *b* from 354 individuals of *N. taihuensis* revealed that low nucleotide diversity ( $\pi = 0.0022 \pm 0.0001$ ) between them, and their relationships was further unveiled by unrooted ML tree which showed no distinct clade formed based on the individuals from different populations (Zhao and others 2008). Attempts have been



made to study the long cytochrome *b* (1,141 bp) for these samples but amplification was not always possible. Nevertheless, in this study, the 16s rRNA sequence distance of D01, D02, D03, F01, F02, F03, F04, F05 and F10 sample group differs significantly from *Neosalanx tangkahkeii* by 0.015 which is above the interspecific genetic distance, and they were clustered together with 99% bootstrap support. Actually morphological examination of the sample group showed homogenous features: short body length with flat and small head while literature description of these 9 species in *Neosalanx* does not differ significantly (Zhang 1987). In order to have a more thorough and thoughtful perspective of interspecific differentiation of *Neosalanx*, more molecular regions as COI and COII, more species and specimens of Salangidae are required for analysis.

### **3.2.5.2 Mislabeled**

The missing or unclear labeling for the fish commodities is very common in the market. The pickled fish samples, S01 and S04 sold as salangids belong to Engraulidae although the latter is also edible. The other samples all belong to Salangidae. Some of them were sold as super-icefish or big-icefish for a higher price than others. The super or big ones are not differentiated in quality, and they just belong to different species. Actually, the fresher icefish has the better quality whereas the quality of dry icefish should be based on the color of the product. Usually, the ones with white color were air-dried well while the orange yellow ones showed signs of microbial spoilage.

As a very important export product, the resources for different species of salangids are different. For the better protecting and sustainable usage of the natural resource, the proper labeling with their correct species name as one of the regulations is necessary but the highest priority to handle (Ogden 2008).

### **3.2.5.3 FINS approach and BLAST analysis**

The phylogenetic analysis approach for FINS technology was first suggested by its founders, Barlett and Davidson. In this study, the FINS technology was employed to differentiate the samples to their families or species. Both the sequence distance matrices and the sequence based phylogenetic trees were made for the species identification, and two distance matrices and three MP trees with 1000 bootstrap replicates were made in this study. According to the analyses, all of the samples were proved to belong to two families and four genera.

The results showed that the samples of D01, D02, D03, F01, F02, F03, F04, F05 and F10 were supported as a species of genus *Neosalanx*. F07 and F09 were *Hemisanx brachyrostralis*. D04, D05, D06, F06, F08, F11 and F12 were *Protosalanx chinensis*.

The pickled samples, S01 and S04, sold as salangids actually belonged to the family Engraulidae in Clupeiformes. Due to the limitation of the molecular markers used and the limited availability of corresponding sequences in public domain, the species identification for S01 and group D01, D02, D03, F01, F02, F03, F04, F05 and F10 remain to be resolved. Nevertheless, utilization of similar morpho-forms is commonly practiced in Chinese food culture.



### **3.3 Conclusions**

FINS approach was effective in authentication of genuine traditional medicine from adulterants as well as in authentication of food species, like fish species, with conducting on the phylogenetic approach. However, insufficient resource of sequence databases of traditional medicine and food resources could limit the effectiveness of FINS approach. Thus, building the specified GenBank for both of traditional medicine and food resources was critical for species authentication.

Moreover, the sequences of the reference plants here were also useful for cladistic analyses and revealed that major revision is needed for classification of sections in *Hedyotis*.

## **CHAPTER 4**

# **PHYLOGENETIC STUDY OF HEDYOTIS IN CHINA AND THEIR POSITION IN SPERMACOCEAE**



## **4.1 Phylogentic study of *Hedyotis* species in Chinese**

### **4.1.1 Nuclear ITS region**

#### **4.1.1.1 Sequence alignment**

All of the sequences of nuclear ITS region of *Hedyotis* species were aligned in ClustalX (2.0) software. The sequence length of ITS intergenic spacer region ranged from 604 bp to 669 bp, and the sequence alignment matrix is shown in **Appendix Table 6- 7**. The sequences of outgroup species were obtained from Genbank of the National Center for Biotechnology Information (NCBI) listed below: *Batopedina pulvinellata* (AM266989) and *Carphalea madagascariensis* (AM266995).

#### **4.1.1.2 Phylogenetic analysis**

In this study, 88 sequences of ITS region were involved and their aligned matrix was comprised of 774 characters, in which 392 of the characters were constant, 85 were variable but uninformative and 297 were parsimony-informative. Gaps were coded by simple indel coding method. Based on 9042 trees, Jackknife 50% majority-rule consensus tree with 1000 replicates of ITS region was conducted in PAUP\* 4.0b10 with heuristic searches and the tree-bisection-reconnection (TBR) branch swapping algorithm presented in **Figure 4- 1**. The individual rooted tree was 910 steps long, with consistency index (CI) of 0.6319, homoplasy index (HI) of 0.3681, retention index (RI) of 0.9338, and rescaled consistent index (RC) of 0.5900. CI excluding

uninformative characters was 0.5879 while HI excluding uninformative characters was 0.4121.

There were two clades (one with 100 jackknife support and the other with 60 jackknife support) split at node one. The first group formed by section *Diplophragma* with well supported (100%). In the second clade, *H. corymbosa* in section *Euoldenlandia* formed a basal clade with strongly jackknife support (100%) to rest of the numbers in this clade. *H. bioflora*, in the section of *Gonothea* formed a sister group to the other species in *Hedyotis*, *Euoldenlandia* (excluding *H. corymbosa*) and *Dimetia* sections was a monophyletic group with 69% jackknife support. The remaining terminals were separated to two clades, the former one with 97% jackknife support was made up by the two sister groups from section *Hedyotis* and *Euoldenlandia*, the later one with 100% jackknife support comprised by two species from *Hedyotis* and *Dimetia* sections, respectively.

The maximum likelihood (ML) tree of ITS region with 500 bootstrap replicates was conducted by Nearest-Neighbor-Interchange (NNI) heuristic method, and it was presented in **Figure 4- 2**. Gaps in here were treated as complete deletion. In this ML tree, two clades were split at node one. The first group was formed by *Diplophragma* and *H. corymbosa*, and the second group was formed by the species in other four sections of *Hedyotis*. *Hedyotis corymbosa* formed a basal clade to the remaining terminals of section *Diplophragma* with 99% bootstrap support. Compared to Jackknife tree, ML tree of ITS region presented weaker resolution power in inferring the inner phylogenetic relationship in section *Diplophragma*.



Figure 4- 1. Jackknife 50% majority-rule strict consensus tree of the *Hedyotis* based on ITS region. The jackknife value was shown above the branch.

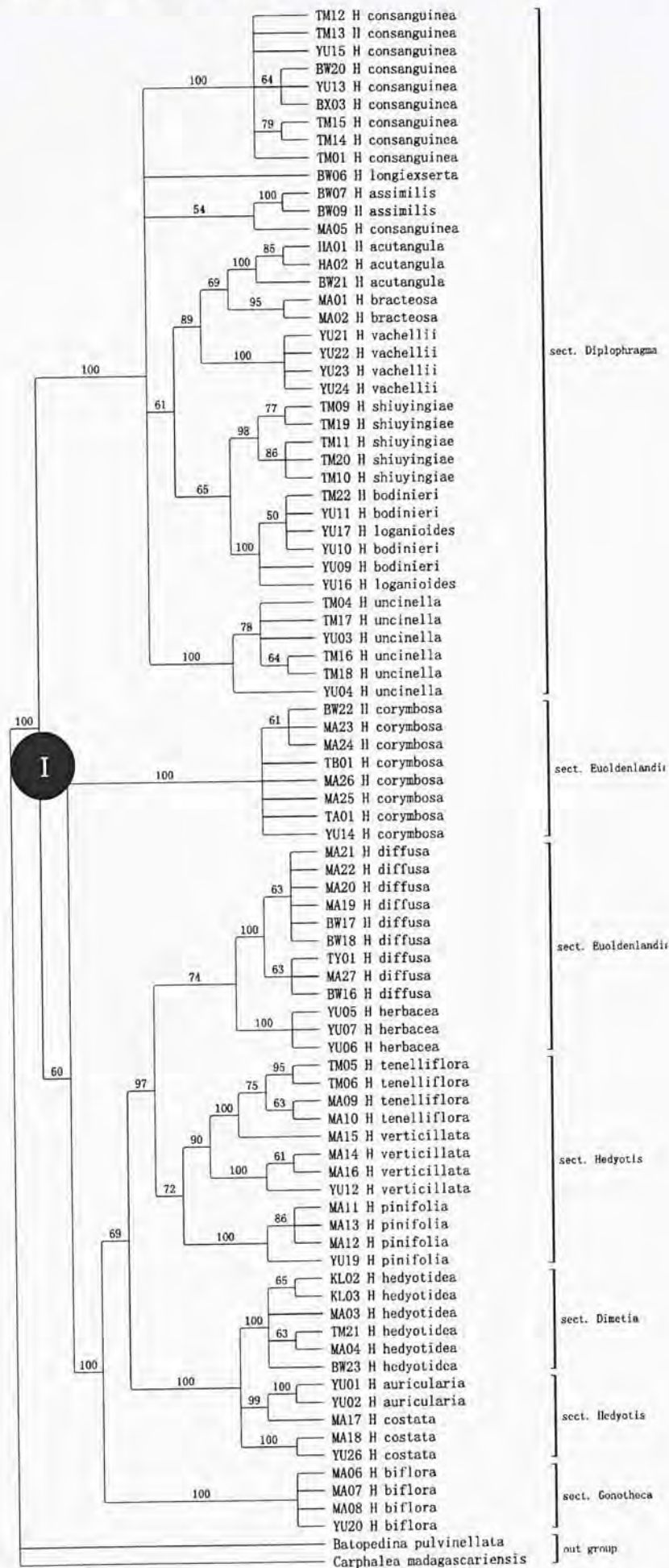
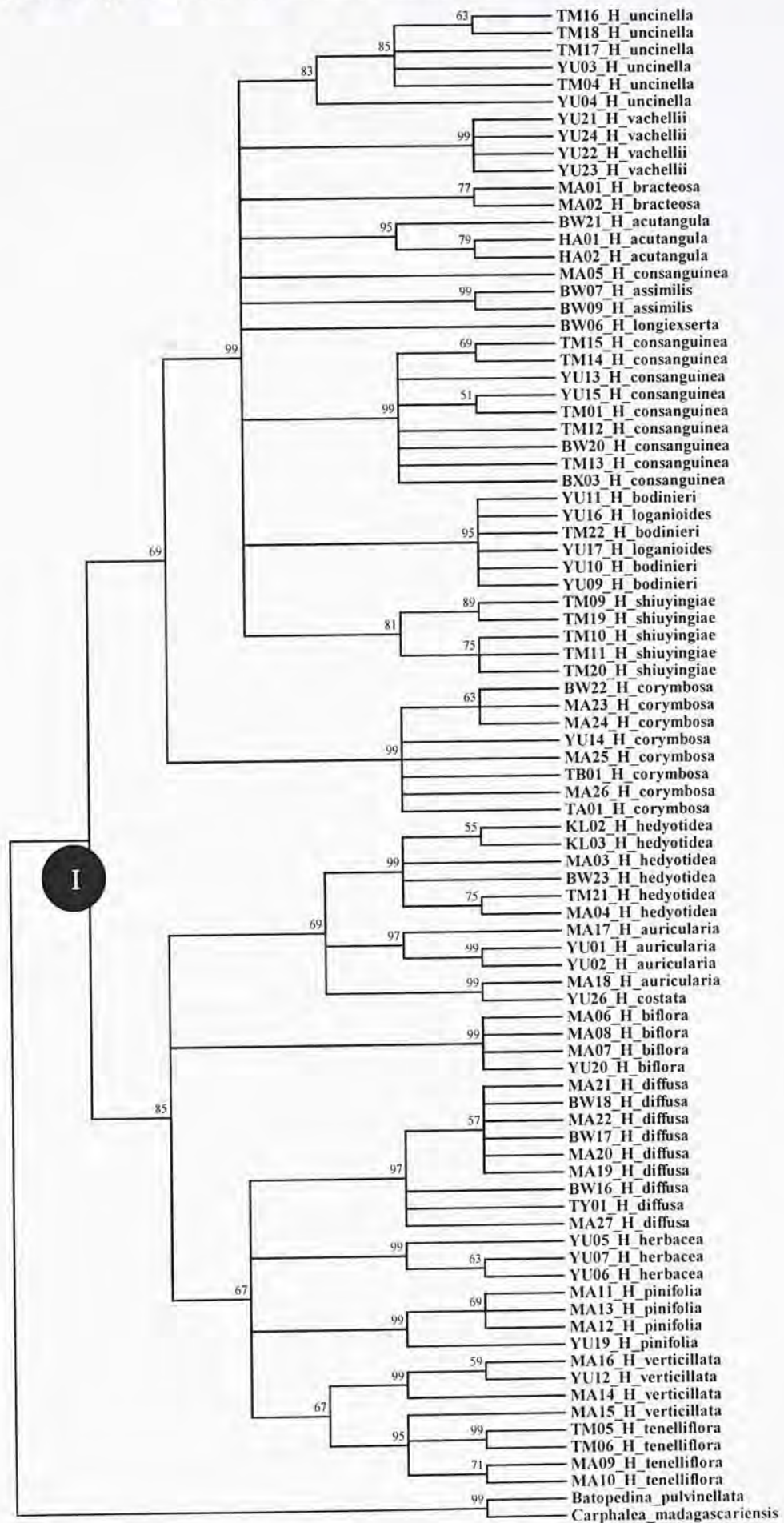


Figure 4- 2. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of the *Hedyotis* based on ITS region. The bootstrap value was shown above the branch.





## **4.1.2 Plastid *trnL* intron and *trnL-F* intergenic spacer region**

### **4.1.2.1 Sequence alignment**

The sequences of chloroplast *trnL* intron and *trnL-F* intergenic spacer region of 89 taxa of *Hedyotis* species were conducted by ClustalX (2.0) software for multiple sequence alignment, and length of this region was about 853 bp to 957 bp. The sequences alignment matrix is shown in **Appendix Table 6- 8**. The species for outgroups obtained from Genbank of the National Center for Biotechnology Information (NCBI) are *Dunnia sinensis* (EU145585) and *Ophiorrhiza elmeri* (EU145564).

### **4.1.2.2 Phylogenetic analysis**

Gaps were coded by simple indel coding method in this study. The aligned matrix was comprised of 1161 characters, 749 of which were constant, 138 of which were variable but uninformative and 274 of which were parsimony-informative. Being conducted in PAUP\* 4.0b10 with heuristic searches and the tree-bisection-reconnection (TBR) branch swapping algorithm, a jackknife 50% majority-rule consensus tree (1000 jackknife replicates) is presented in **Figure 4- 3** with 590 steps long was got from 86700 trees. A consistency index (CI) of 0.7763, homoplasy index (HI) of 0.2237, retention index (RI) of 0.9722, and rescaled consistent index (RC) of 0.7547 were borne by the region. CI excluding uninformative characters was 0.7034

while HI excluding uninformative characters was 0.2966.

There were two clades formed in the phylogenetic tree of *trnL* intron and *trnL-F* intergenic spacer: the upper clade was made up by the species in section *Diplophragma* with 100% jackknife support, and the later one was consisted of other four sections numbers with 100% jackknife support. In the second clade, *H. corymbosa*, a species in section *Euoldenlandia*, which formed a monophyletic group with 100% jackknife support, was separated from other species as a basal clade. *H. biflora* in *Gonotheca* section formed a well supported clade (99%) that was sister to the rest species. The remaining data were comprised of two well supported clades. The former group was comprised of *H. hedyotideae* (section *Dimetia*), *H. auricularia* and *H. costata* (section *Hedyotis*) with 100% jackknife support, and the later was made up by the species from *Hedyotis* and *Euoldenlandia* sections.

The maximum likelihood (ML) tree of *trnL* intron and *trnL-F* intergenic spacer with 500 bootstrap replicates was conducted by Nearest-Neighbor-Interchange (NNI) heuristic method, and it was presented in **Figure 4- 4**. Gaps were treated as complete deletion. In this ML tree, two clades with 99% bootstrap supported were formed at node one. The upper group was formed by the species in four sections (*Dimetia*, *Hedyotis*, *Euoldenlandia* and *Gonotheca*), while the other group consisted of species in *Diplophragma* section. Compared to Jackknife tree, ML tree of *trnL* intron and *trnL-F* intergenic spacer can't separate *H. herbacea* from *H. diffusa*. It can't resolve the inner phylogenetic relationship in *Diplophragma*.



Figure 4- 3. Jackknife 50% majority-rule strict consensus tree of the *Hedyotis* based on *trnL* intron and *trnL-F* intergenic spacer region. The jackknife value is shown above the branch.

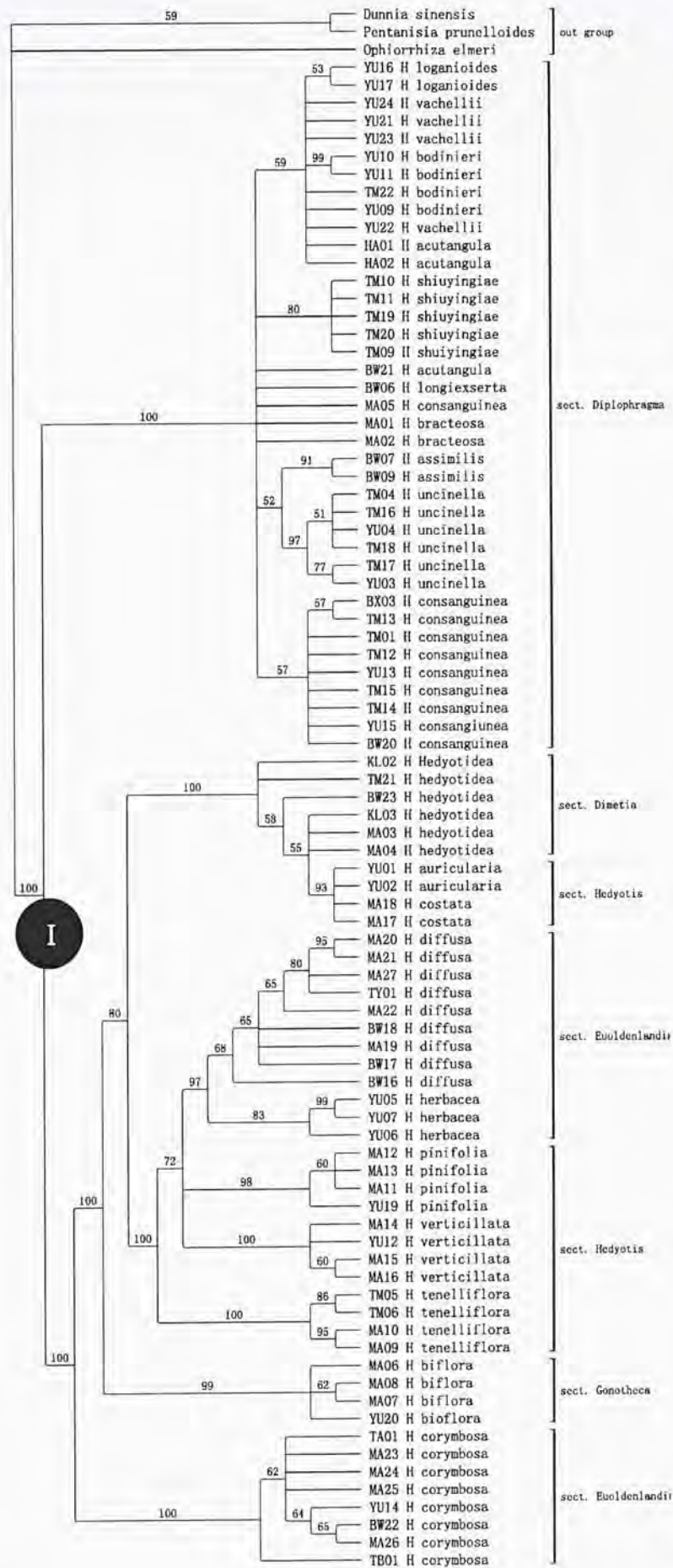
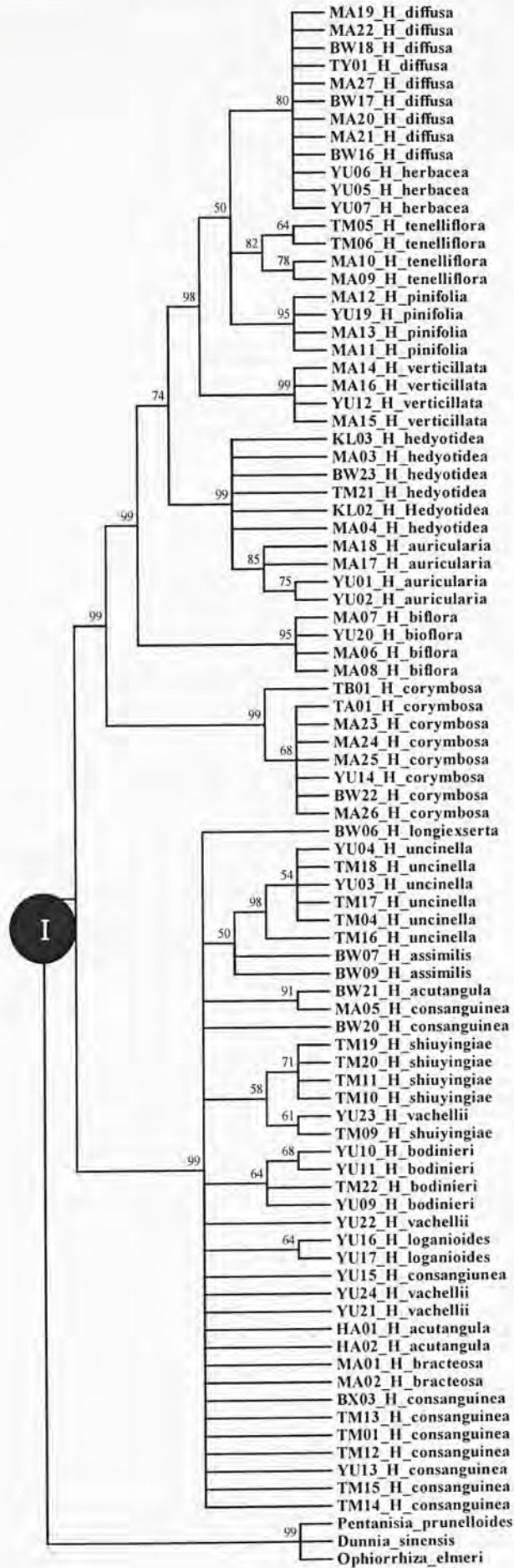


Figure 4- 4. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of *trnL* intron and *trnL*-F intergenic spacer. The bootstrap value is shown above the branch.





### 4.1.3 Plastid *trnH-psbA* intergenic spacer region

#### 4.1.3.1 Sequence alignment,

Plastid *trnH-psbA* intergenic spacer of *Hedyotis* species was successfully sequenced and then conducted by ClustalX (2.0) software for multiple sequence alignment. The sequence length of *trnH-psbA* intergenic spacer region ranged from 265 bp to 475 bp, and the sequences alignment matrix is shown in **Appendix Table 6- 9**. The outgroup species obtained from Genbank of the National Center for Biotechnology Information (NCBI) were *Kirengeshoma palmate* (EU701049) and *Rubia tinctorum* (EU531723).

#### 4.1.3.2 Phylogenetic analysis

The sequences of *trnH-psbA* intergenic spacer region from 88 specimens were included in this study. The aligned matrix was comprised of 686 characters, 258 of which were constant, 117 of which were variables but uninformative and 311 of which are parsimony-informative. Gaps were coded by simple indel coding method. Jackknife 50% majority-rule consensus tree with 1000 replicates for *trnH-psbA* intergenic spacer region of *Hedyotis* conducted in PAUP\* 4.0b10 with heuristic searches and the tree-bisection-reconnection (TBR) branch swapping algorithm is presented in **Figure 4- 5**. Most parsimony trees got 0.7475 consistency index, 0.2525 homoplasy index (HI), 0.9470 retention index (RI), 0.7079 rescaled consistent index (RC). CI excluding uninformative characters was 0.6920 while HI excluding

uninformative characters was 0.3080.

Samples of *H. corymbosa* in Euoldenlandia section were clustered with *Kirengeshoma palmate* (outgroup species) at 63% jackknife support. Species in section Diplophragma formed a well supported (100%) clade, while the other four sections were clustered together with 98% jackknife support. Within the second clade, the two major groups were separated: one was consisted of the *H. pinifolia*, *H. verticillata* and *H. piniflora* in Hedyotis section with 82% jackknife support and the other one was comprised of the species in Euoldenlandia (excluding *H. corymbosa*), Dimetia, Gonotheca, and Hedyotis (*H. auricularia* and *H. costata*) sections with 62% jackknife support.

Sequences of *trnH-psbA* intergenic spacer were used to construct the bootstrap maximum likelihood tree (500 replicates) with Nearest-Neighbor- Interchange (NNI) heuristic method. The ML tree of *trnH-psbA* intergenic spacer was presented in **Figure 4- 6**. Gaps were treated as complete deletion here. In this ML tree, *H. corymbosa*, *Kirengeshoma palmate* and *Rubia tinctorum* formed the basal group with 56% bootstrap support. Species in section Diplophragma formed a moderate jackknife supported (85%) clade. Compared to Jackknife tree, ML tree of *trnH-psbA* intergenic spacer showed a poor resolution power to investigate the inner phylogenetic relationship in *Hedyotis*.



Figure 4- 5. Jackknife 50% majority-rule strict consensus tree of the *Hedyotis* based on *trnH-psbA* intergenic spacer. The jackknife value is shown above the branch.

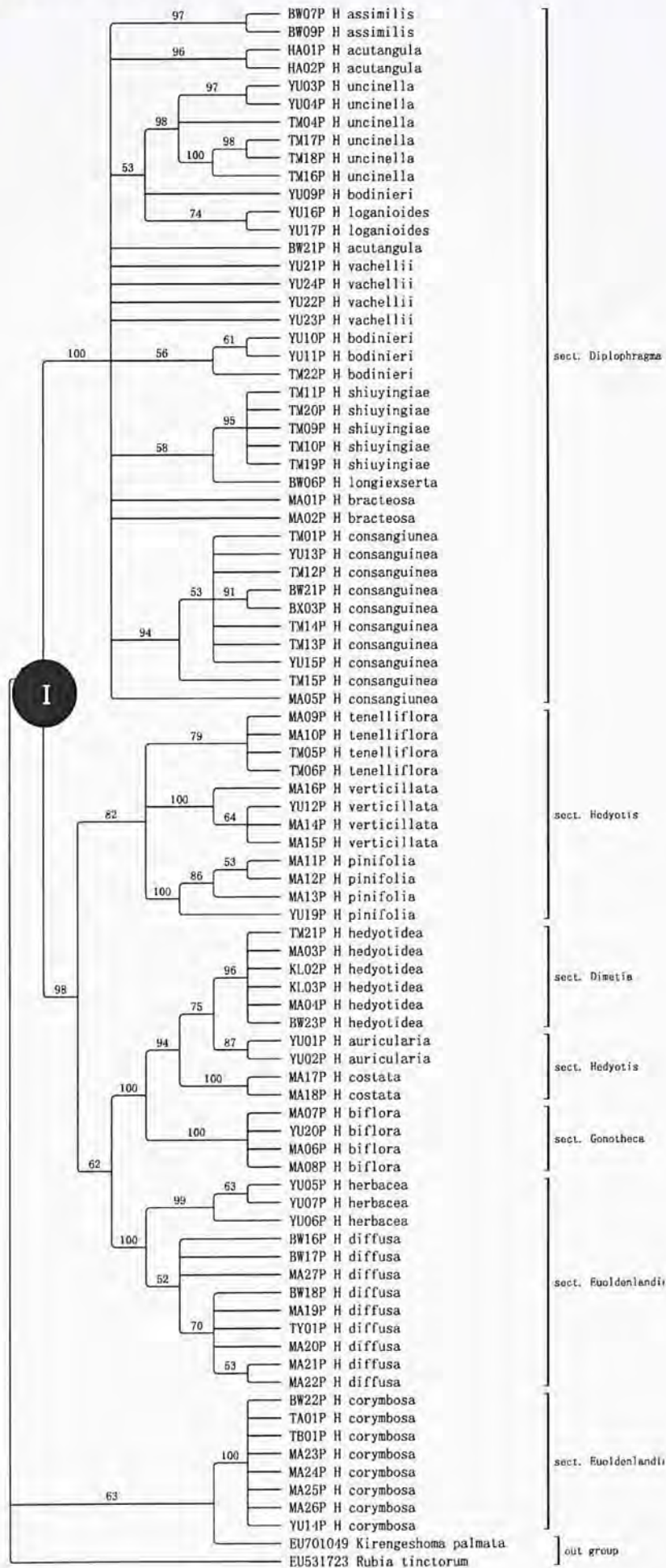
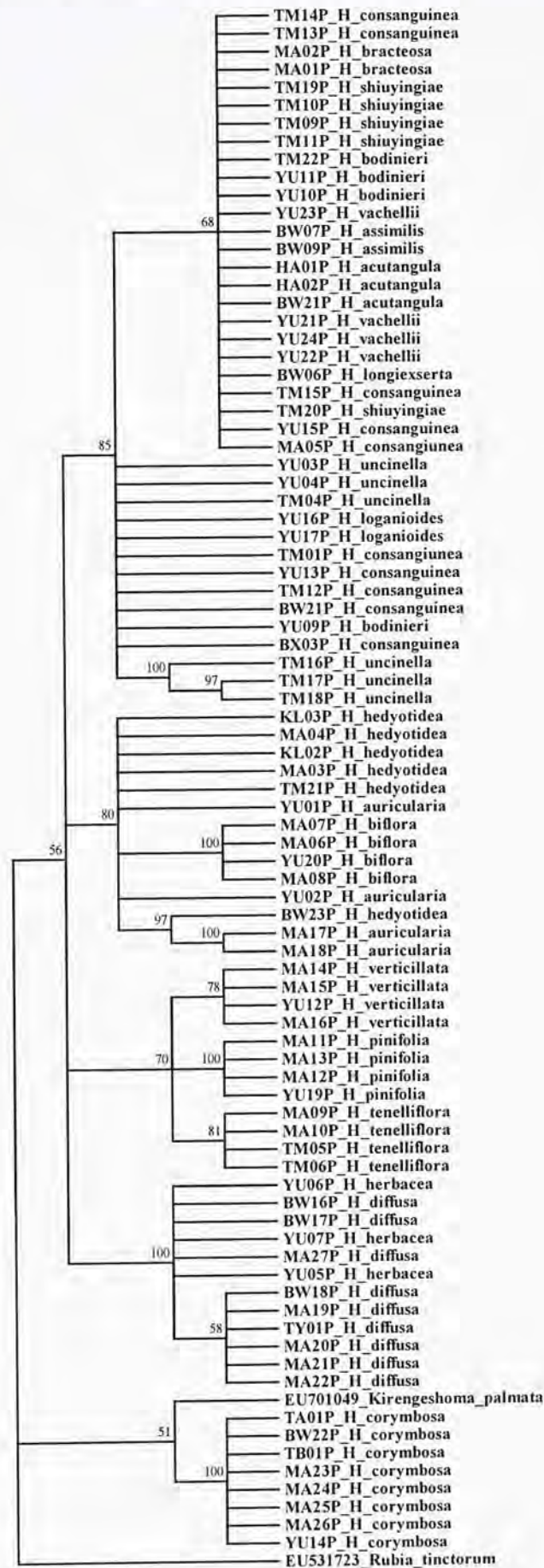


Figure 4- 6. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of *trnH-psbA* intergenic spacer. The bootstrap value is shown above the branch.





## 4.1.4 Plastid *rbcL* region

### 4.1.4.1 Sequence alignment

CLUSTALX (2.0) software was employed for multiple sequence alignment of plastid *rbcL* region for 89 *Hedyotis* species, and the alignment matrix is shown in **Appendix Table 6- 10**. The sequence length of *rbcL* region was about 670 bp to 671 bp, and sequences of outgroup species obtained from Genbank of the National Center for Biotechnology Information (NCBI) were *Pentanisia prunelloides* (AM117255), *Dunnia sinensis* (EU145467) and *Ophiorrhiza elmeri* (EU145464).

### 4.1.4.2 Phylogenetic analysis

Gaps were coded by simple indel coding method in this study. The aligned matrix was comprised of 675 characters, 565 of which were constant, 50 of which were variables but uninformative and 60 of which were parsimony-informative. 326 trees of *rbcL* region were got from PAUP\* 4.0b10 with heuristic searches and the tree-bisection-reconnection (TBR) branch swapping algorithm. Jackknife 50% majority-rule consensus tree measured by 1000 jackknife replicates is presented in **Figure 4- 7**. The consistency index (CI) is 0.6932, homoplasy index (HI) is 0.3068, retention index (RI) is 0.9146, and rescaled consistent index (RC) is 0.6340. CI excluding uninformative characters was 0.5645 while HI excluding uninformative characters was 0.4355.

There were three clades with jackknife support (97%, 93%, 71%, respectively) split

at node one. The first well supported (97%) clade formed by section *Diplophragma*. In the second strongly support (93%) clade, *H. corymbosa*, a species in section *Euoldenlandia*, formed a monophyletic group. The third clade with 71% jackknife support was well separated from the other two clades. One clade consisted of the species in *Dimetia* and *Hedyotis* sections with 85% jackknife support, and the other with 52% support comprised of two sister groups, one of which was consisted of *H. bioflora* in *Gonotheca* section with 100% jackknife support, and the second one formed by species in *Hedyotis* and *Euoldenlandia* (excluding *Hedyotis corymbosa*) sections with 68% jackknife support.

Sequences of *rbcL* region were used to construct the bootstrap maximum likelihood tree (500 replicates) with Nearest-Neighbor- Interchange (NNI) heuristic method, and this ML tree was presented in **Figure 4- 8**. Gaps were treated as complete deletion here. This MP tree also formed three clades at node one with 63%, 96%, 98% bootstrap support, respectively. The group numbers in each clade of MP tree were the same as in jackknife tree.



Figure 4- 7. Jackknife 50% majority-rule strict consensus tree of the *Hedyotis* based on *rbcl* region. The jackknife value is shown above the branch.

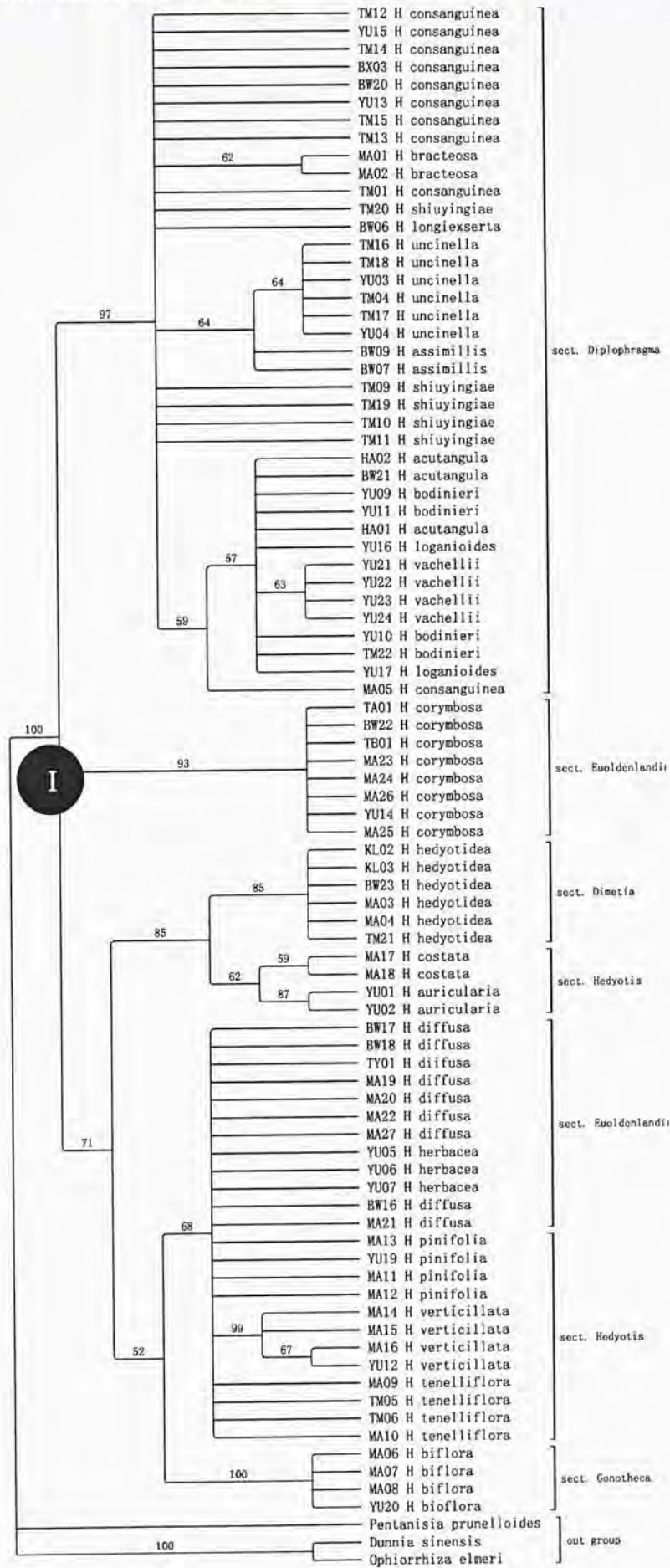
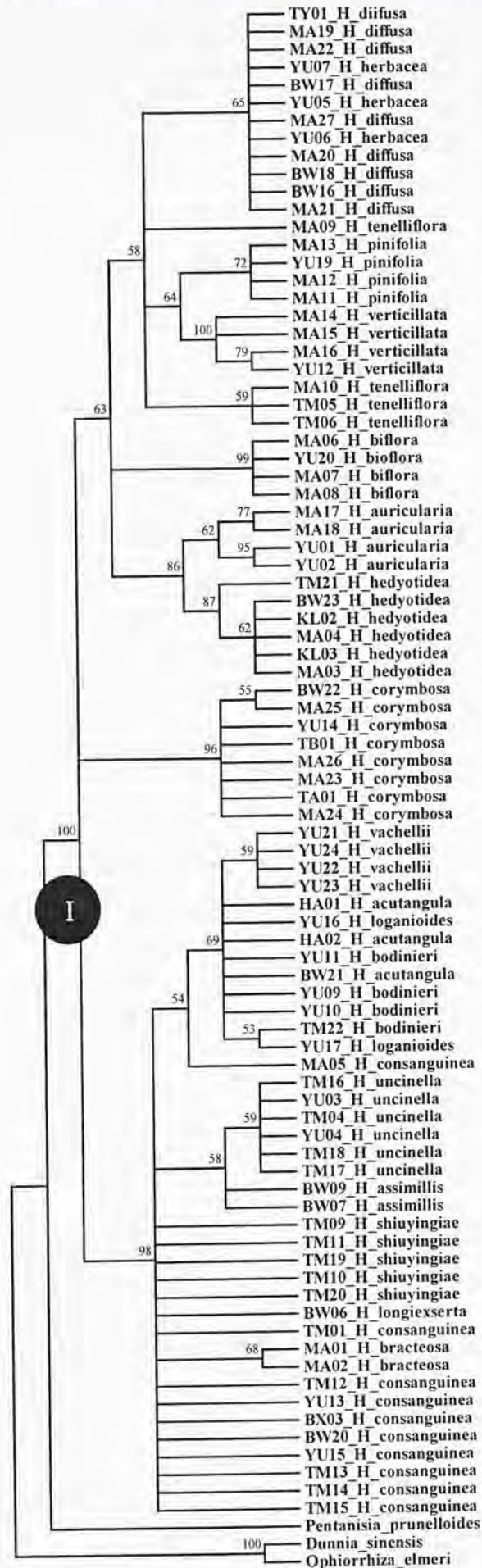


Figure 4- 8. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of *rbcl* region. The bootstrap value is shown above the branch.





## 4.1.5 Plastid *matK* region

### 4.1.5.1 Sequence alignment

After being successfully sequenced, plastid *matK* region of *Hedyotis* species were aligned by ClustalX (2.0) software. The sequences length of *matK* region ranged from 868 bp to 875 bp, and the sequences alignment matrix could be found in **Appendix Table 6- 11**. Two species of *Psychotria kirkii* (AY538413) and *Ophiorrhiza pumila* (AB247150) obtained from Genbank of the National Center for Biotechnology Information (NCBI) were selected as outgroup.

### 4.1.5.2 Phylogenetic analysis

The aligned matrix was comprised by 88 taxa of 906 characters, in which 647 characters were constant, 99 were variables but uninformative and 160 of which are parsimony-informative. In this study, gaps were coded by simple indel coding method. 22980 trees were got here. Jackknife 50% majority-rule consensus tree (1000 jackknife replicates) of *matK* region of *Hedyotis* conducted in PAUP\* 4.0b10 with heuristic searches and the tree-bisection-reconnection (TBR) branch swapping algorithm is presented in **Figure 4- 9**. The most parsimony trees were 384 steps long, with consistency index (CI) of 0.7552, homoplasy index (HI) of 0.2448, retention index (RI) of 0.9546, and rescaled consistent index (RC) of 0.7209. CI excluding uninformative characters was 0.6655 while HI excluding uninformative characters was 0.3345.

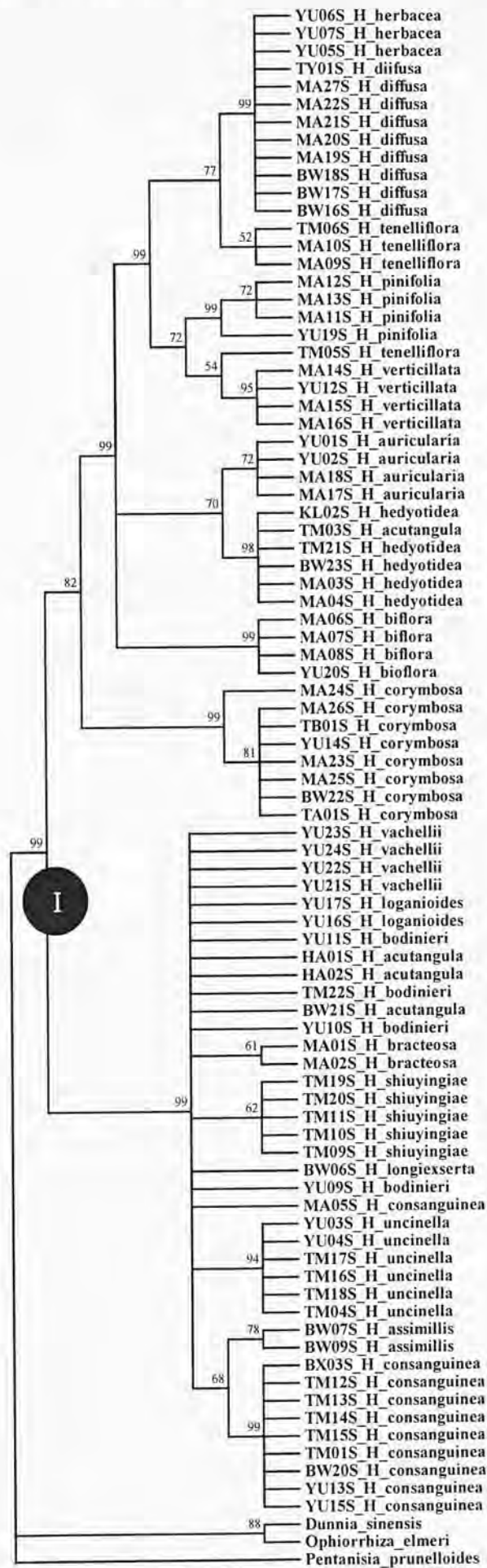
There were two clades with strongly jackknife support (100%) split at node one. The first monophyletic group formed by *Diplophragma* with well jackknife supported (100%) and the second clade formed by four other sections with 92% jackknife support. Within the second clade, *H. corymbosa* in *Euoldenlandia* section formed a basal clade with strongly support (100%) to rest of the numbers which were separated from two clades. The former with 60% jackknife support was made up by the three sister groups from sections of *Hedyotis*, *Dimetia* and *Gonotheca*, respectively. The later with 98% jackknife support was comprised of two sister groups from sections of *Hedyotis* and *Euoldenlandia*.

Sequences of *matK* region were used to construct the bootstrap (500 replicates) maximum likelihood tree with Nearest-Neighbor-Interchange (NNI) heuristic method, and the tree was presented in **Figure 4- 10**. Gaps were treated as complete deletion here. In this ML tree, two clades were formed at node one with 82% and 99% bootstrap support, respectively. The group numbers in each clade of MP tree were the same as in jackknife tree. The inner phylogenetic relationship of the second clade was similar in two phylogenetic trees.





Figure 4- 10. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of *matK* region. The bootstrap value is shown above the branch.





## 4.1.6 Plastid *rps16* region

### 4.1.6.1 Sequence alignment

The sequences ranged from 635 bp to 673 bp of plastid *rps16* region were conducted in ClustalX (2.0) software for multiple sequence alignment, and the sequences alignment matrix is shown in **Appendix Table 6- 12**. *Ophiorrhiza elmeri* (EU145510) and *Pentanisia prunelloides* (AM266860) from Genbank of the National Center for Biotechnology Information (NCBI) were chosen as outgroup in this study.

### 4.1.6.2 Phylogenetic analysis

Eighty-nine sequences of *rps16* region were included in this study. The aligned matrix comprised of 752 characters, 546 of which were constant, 73 of which were variables but uninformative and 133 of which were parsimony-informative. Gaps were coded by simple indel coding method. Jackknife 50% majority-rule consensus tree of *rps16* region with heuristic searches and the tree-bisection-reconnection (TBR) branch swapping algorithm was assessed by 1000 jackknife replicates. The tree is presented in **Figure 4- 11**. Most parsimony trees were 274 steps long, with consistency index (CI) of 0.8394, homoplasy index (HI) of 0.1606, retention index (RI) of 0.9777, and rescaled consistent index (RC) of 0.8207. Additionally, CI excluding uninformative characters was 0.7800, and HI excluding uninformative characters was 0.2200 accordingly.

There were two clades with strongly jackknife support (99%) split at node one. The first clade formed by section *Diplophragma* with well supported (100%) while the second clade formed by other four sections with 99% jackknife support. In the second clade, *H. corymbosa* in *Euoldenlandia* section formed a basal clade with strongly support (100%) to rest of the numbers. The remaining data were separated from the three clades. The first one with 85% jackknife support was made up by two separate groups from *Hedyotis* and *Dimetia* sections, *H. biflora* in section *Gonotheca* formed the second clade with strongly jackknife support and the last one was consisted of two sister groups from section *Hedyotis* and section *Euoldenlandia*.

The maximum likelihood (ML) tree of *rps16* region with 500 bootstrap replicates was conducted by Nearest-Neighbor-Interchange (NNI) heuristic method, and it was presented in **Figure 4- 12**. Gaps were treated as complete deletion here. In this ML tree, two clades with 82% and 99% bootstrap supported were formed at node one. The upper group was formed by species in *Hedyotis*, *Euoldenlandia*, *Gonotheca* and *Dimetia* with *H. corymbosa* as the basal clade, and the other group consisted of *Diplophragma*. Compared to Jackknife tree, ML tree of *rps16* region can't separate *H. herbacea* from *H. diffusa*. Both the Jackknife tree and ML tree showed poor resolution power in inferring the inner relationship in *Hedyotis*.



Figure 4- 11. Jackknife 50% majority-rule strict consensus tree of the *Hedyotis* based on *rps16* region. The jackknife value is shown above the branch.

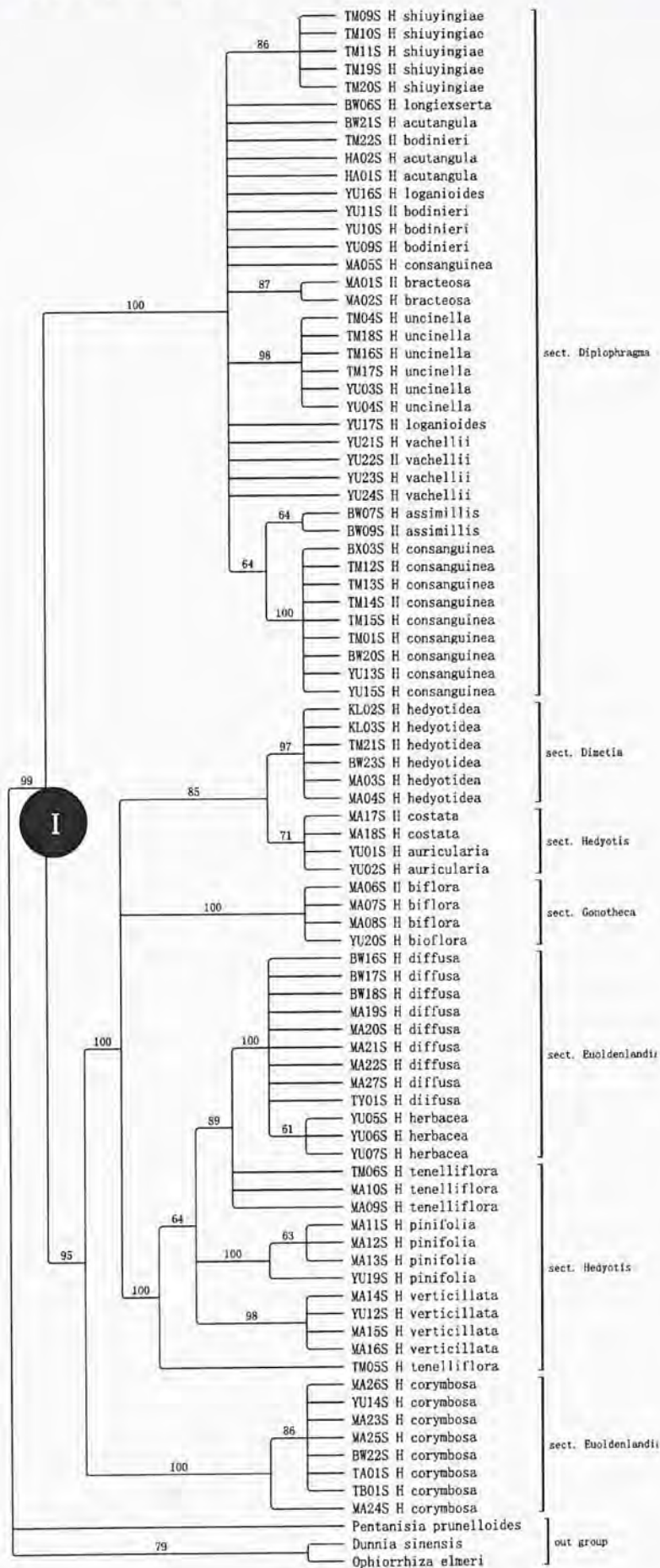
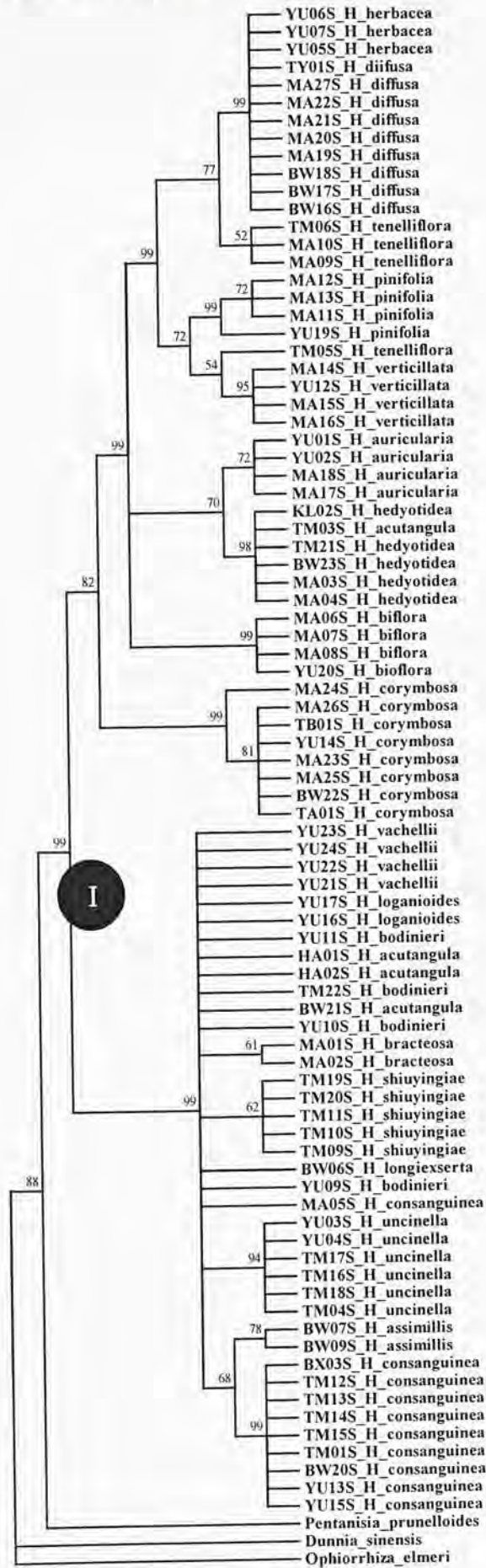


Figure 4- 12. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of *rps16* region. The bootstrap value is shown above the branch.





## 4.1.7 Combined analysis

### 4.1.7.1 Sequence alignment

The combined dataset of six molecular regions of *Hedyotis* was conducted by ClustalX (2.0) software for multiple sequence alignment. The sequences length of combined region ranged from 3966 bp to 4247 bp, and the sequences alignment matrix is shown in **Appendix Table 6- 13**. Outgroup species were selected from Genbank of the National Center for Biotechnology Information (NCBI) *Psychotria\_kirkii* (AF072038, AY538413, X83663, AY538469, AF410728 and FJ208657).

### 4.1.7.2 Phylogenetic analysis

Gaps were coded by simple indel coding method. Combined sequences dataset generated from 86 taxa was included in this study. The dataset was comprised of 4911 characters, 3211 of which were constant, 565 of which were variables but uninformative and 1135 of which were parsimony-informative. A jackknife 50% majority-rule consensus tree with 1000 replicates for *Hedyotis* conducted in PAUP\* 4.0b10 with heuristic searches was generated and the tree-bisection-reconnection (TBR) branch swapping algorithm is presented in **Figure 4- 13**. The most parsimony trees were 2846 steps long, with consistency index (CI) of 0.7294, homoplasy index (HI) of 0.2706, retention index (RI) of 0.9567, and rescaled consistent index (RC) of 0.6979. CI excluding uninformative characters was 0.6590 while HI excluding

uninformative characters was 0.3410.

The outgroup selection was very hard for the combined datasets, as the regions of *trnH-psbA* and *matK* regions were rarely used for phylogenetic analyses in Rubioideae subfamily. *Psychotria kirkii* of Psychotria with six regions' sequences found in GenBank, was qualified as outgroup species. The first clade with 100% jackknife supported was consisted of species in Diplophragma section showed in this cladogram. MA05 sample collected from Macao was separated from the *H. consangiunea* group in combined datasets analysis as well as in single-region analyses. In the second clade, *H. corymbosa* of section Euoldenlandia, formed a sister group with strong support (100%) to the rest of the species in the section. The remaining data were separated to two clades. The former with 100% jackknife support was consisted of the two sister groups from the sections of Hedyotis and Euoldenlandia and the later with 94% jackknife support was comprised of the species from Hedyotis (*H. auricularia* and *H. costata*), Dimetia, and Gonotheca.

The maximum likelihood tree of combined region with 500 bootstrap replicates was conducted by Nearest-Neighbor-Interchange (NNI) heuristic method, and it was presented in **Figure 4- 14**. Gaps were treated as complete deletion here. In this ML tree, three major groups were formed. The first group was formed by species in Hedyotis, Euoldenlandia, Gonotheca and Dimetia with *H. corymbosa* as the basal clade, the second group was consisted of Diplophragma, and the third group was made up of outgroup species. The combined regions of ML tree showed the best work in resolving the inner relationship of Diplophragma.



#### 4.1.8 Comparison of phylogenetic utility of the six DNA regions

In order to compare the utility of six different data sets involved in this study, the relative indices are listed in **Table 4.1**. For single-region analyses, ITS, *trnH-psbA* and *trnL* intron and *trnL-F* intergenic spacer were the most informative regions, proving about 2 times as many informative characters as *matK* and *rps16* regions, and about 6 times as many informative characters as *rbcL* region. The region of *trnH-psbA* had the highest percent variability (0.453) while the region of *rbcL* represented the least percent variability (0.089) in this study. When comparing homoplasy levels of the trees from each study (excluding uninformative characters), the consistency index (excluding uninformative characters) of *rbcL* was only 0.5645, and the homoplasy index (excluding uninformative characters) was 0.4355. The *rps16* region showed 0.78 of the best consistency index (excluding uninformative characters) and 0.22 of homoplasy index (excluding uninformative characters). Compared to other indices, retention index (RI) was much insensitive to the large amount of taxa and characters and it ranged from 0.9146 to 0.9777 (*rps16*) for six single region. **Table 4- 1** shows the variable percentage for six molecular regions, the *trnH-psbA* region of which showed the highest proportion of parsimony-informative characters in total characters. Not surprisingly, the fraction of PIC/TC for *rbcL* region was the lowest.

Five molecular regions (excluding the *rbcL* region) presented good performance in inferring the relationship of Hedyotis, Dimetia, Euoldenlandia and Gonotheca

sections. ITS region resolved the highest proportion of inter nodes within Diplophragma clade, and the result was also supported by the combined datasets analysis. However, the other five regions showed very poor resolution in inferring the phylogenetic relationship of section Diplophragma.



Figure 4- 13. Jackknife 50% majority-rule strict consensus tree of the *Hedyotis* based on the combined datasets. The jackknife value is shown above the branch.

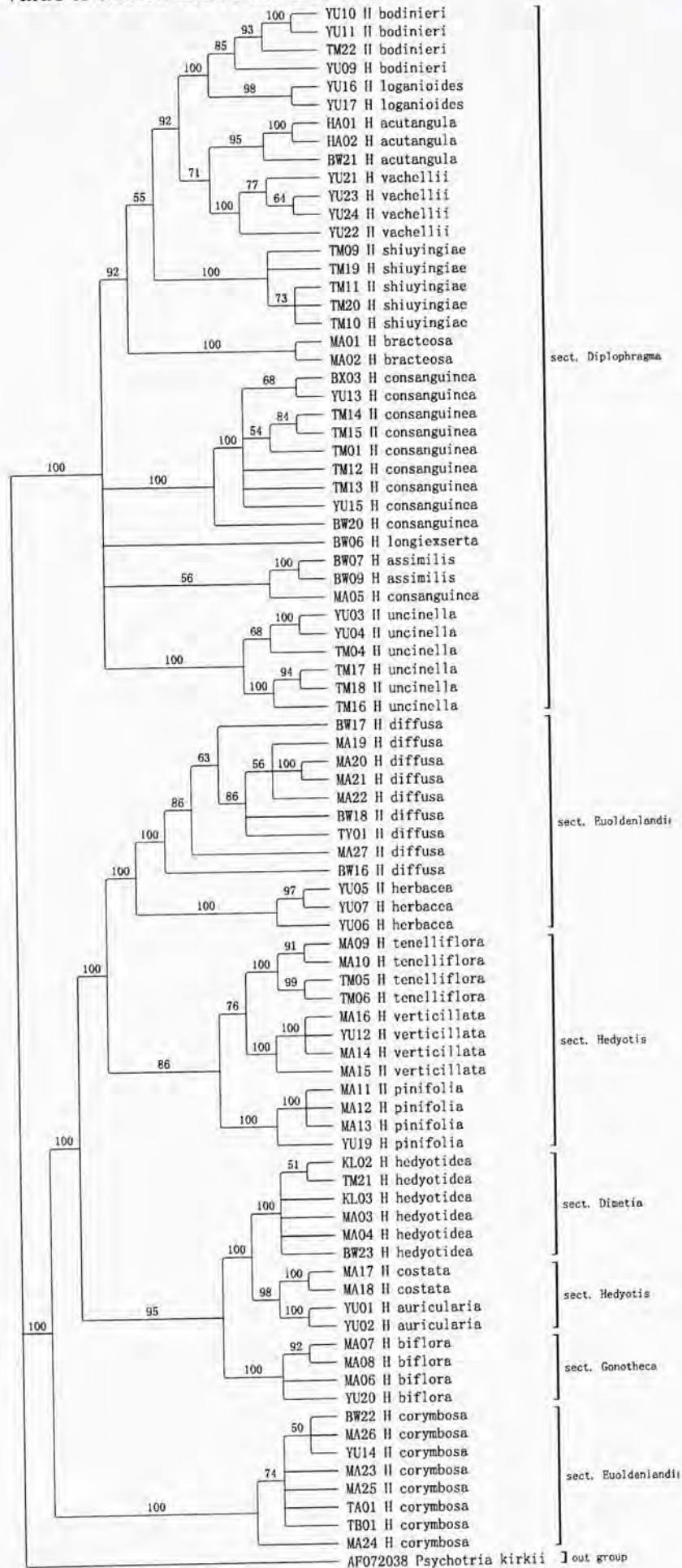
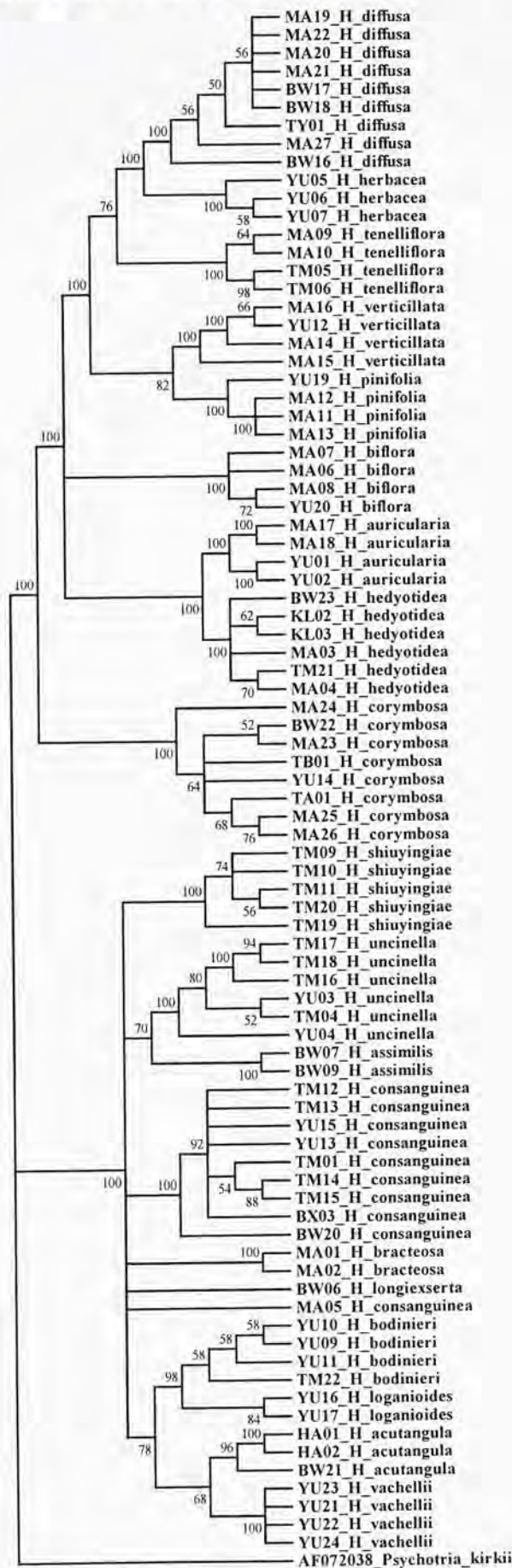


Figure 4- 14. Bootstrap 50% majority-rule strict consensus maximum likelihood tree of combined region. The bootstrap value is shown above the branch.





**Table 4- 1. Total taxa (TT), total characters (TC), constant characters (CC), parsimony-uninformative variable character (PUIC), parsimony-informative characters (PIC), tree length (TL), consistency index (CI), homoplasy index (HI), CI excluding uninformative characters (CIE), HI excluding uninformative characters (HIE), retention index (RI), rescaled consistency index (RC) for different datasets.**

<b>Dataset</b>	<b>TT</b>	<b>TC</b>	<b>CC</b>	<b>PUIC</b>	<b>PIC</b>	<b>TL</b>	<b>CI</b>	<b>HI</b>	<b>CIE</b>	<b>HIE</b>	<b>RI</b>	<b>RC</b>
<b>ITS</b>	88	774	392	85	297	910	0.6319	0.3681	0.5879	0.4121	0.9338	0.5921
<b><i>trnL</i> and <i>trnL-F</i></b>	88	1161	749	138	274	590	0.7763	0.2237	0.7034	0.2966	0.9722	0.7547
<b><i>trnH-psbA</i></b>	87	686	258	117	311	705	0.7475	0.2525	0.692	0.308	0.947	0.7079
<b><i>rbcL</i></b>	88	675	565	50	60	176	0.6932	0.3068	0.5645	0.4355	0.9146	0.634
<b><i>matK</i></b>	89	906	647	99	160	384	0.7552	0.2448	0.6655	0.3345	0.9546	0.7209
<b><i>rps16</i></b>	88	752	546	73	133	274	0.8394	0.1606	0.78	0.22	0.9777	0.8207
<b>Combined</b>	86	4911	3211	565	1135	2846	0.7294	0.2706	0.6590	0.3410	0.9576	0.6979

## 4.2 The phylogenetic position of *Hedyotis* (species in China) in the tribe of Spermaceae s. l.

### 4.2.1 Plastid *trnL*-F intergenic spacer region

#### 4.2.1.1 Sequence alignment

Plastid *trnL*-F intergenic spacer region of Rubiaceae species was employed to test the phylogenetic position of *Hedyotis* (species in China). The sequence length of *trnL*-F intergenic spacer in this study ranged from 326 to 449 bp. After conducting sequence alignment by ClustalX (2.0) software, the alignment matrix of 186 samples was presented in **Appendix Table 6- 14**. Sequences of species in Spermaceae obtained from GenBank of the National Center for Biotechnology Information (NCBI) are listed below.

**Table 4- 2. List of sequences downloaded from GenBank of the National Center for Biotechnology Information (NCBI).**

Genebank accession number	Species
AF333349	<i>Arcytophyllum aristatum</i>
AF333351	<i>Arcytophyllum ciliolatum</i>
AF333355	<i>Arcytophyllum lavarum</i>
AF333357	<i>Arcytophyllum macbridei</i>
AF333358	<i>Arcytophyllum muticum</i>
AF333361	<i>Arcytophyllum nitidum</i>
AF333365	<i>Arcytophyllum setosum</i>
AF333367	<i>Arcytophyllum thymifolium</i>
AF333369	<i>Arcytophyllum vernicosum</i>
DQ359165	<i>Bouvardia glaberrima</i>
EU642537	<i>Bouvardia ternifolia</i>
EU543085	<i>Conostomium natalense</i>
EU543086	<i>Conostomium quadrangulare</i>
EU543087	<i>Conostomium zoutpansbergense</i>



---

EU543088	<i>Crusea calocephala</i>
EU543089	<i>Crusea megalocarpa</i>
EU543090	<i>Dentella dioeca</i>
EU543091	<i>Dentella repens</i>
EU145574	<i>Dibrachionostylus kaessneri</i>
EU543092	<i>Diodia aulacosperma</i>
EU543093	<i>Diodia spicata</i>
EU145585	<i>Dunnia sinensis</i> (outgroup)
EU543094	<i>Emmeorrhiza umbellata</i>
EU543095	<i>Ernodea littoralis</i>
EU543096	<i>Galianthe brasiliensis</i>
EU543097	<i>Galianthe eupatorioides</i>
EU543098	<i>Hedyotis fruticosa</i>
EU543099	<i>Hedyotis korrorensis</i>
EU543101	<i>Hedyotis lessertiana</i>
EU543100	<i>Hedyotis lessertiana</i> var <i>lessertiana</i>
EU543102	<i>Hedyotis macrostegia</i>
EU543103	<i>Hedyotis quinquenervia</i>
EU543104	<i>Hedyotis rhinophylla</i>
AF333378	<i>Hedyotis serpens</i>
EU543105	<i>Hedyotis swertioides</i>
EU543107	<i>Hedythyrus spermacocinus</i>
EU543109	<i>Houstonia caerulea</i>
AF333380	<i>Houstonia longifolia</i>
EU543110	<i>Kadua acuminata</i>
EU642538	<i>Kadua affinis</i>
EU543111	<i>Kadua centranthoides</i>
EU543112	<i>Kadua cordata</i>
EU642539	<i>Kadua coriacea</i>
EU543113	<i>Kadua degeneri</i>
EU642540	<i>Kadua elatior</i>
EU642541	<i>Kadua fluviatilis</i>
EU642542	<i>Kadua flynnii</i>
EU543114	<i>Kadua foggiana</i>
EU642543	<i>Kadua fosbergii</i>
EU642544	<i>Kadua laxiflora</i>
EU543116	<i>Kadua parvula</i>
EU642545	<i>Kadua rapensis</i>
EU543117	<i>Kohautia amatymbica</i>
EU543118	<i>Kohautia caespitosa</i>
EU543119	<i>Kohautia coccinea</i>
EU543120	<i>Kohautia cynanchica</i>
EU543122	<i>Kohautia obtusiloba</i>
EU642546	<i>Kohautia senegalensis</i>

---

---

EU543123	<i>Kohautia subverticillata</i>
EU543124	<i>Kohautia virgata</i>
EU543125	<i>Lelya osteocarpa</i>
EU543126	<i>Manettia lygistum</i>
EU543127	<i>Manostachya ternifolia</i>
EU543128	<i>Mitracarpus frigidus</i>
EU543129	<i>Mitrasacmopsis quadrivalvis</i>
EU543130	<i>Oldenlandia affinis</i>
EU543131	<i>Oldenlandia angolensis</i>
EU543132	<i>Oldenlandia biflora</i>
EU543133	<i>Oldenlandia capensis var capensis</i>
EU543134	<i>Oldenlandia capensis var pleiosepala</i>
EU543137	<i>Oldenlandia echinulosa var pellucida</i>
EU543138	<i>Oldenlandia fastigiata</i>
EU543139	<i>Oldenlandia galioides</i>
EU543140	<i>Oldenlandia geophila</i>
EU543143	<i>Oldenlandia herbacea</i>
EU543142	<i>Oldenlandia herbacea var goetzei</i>
EU543143	<i>Oldenlandia herbacea var herbacea</i>
EU543144	<i>Oldenlandia lancifolia</i>
EU543145	<i>Oldenlandia microtheca</i>
EU543147	<i>Oldenlandia robinsonii</i>
EU567467	<i>Oldenlandia rosulata</i>
EU543149	<i>Oldenlandia taborensis</i>
EU543106	<i>Oldenlandia tenelliflora</i>
EU543150	<i>Oldenlandia uniflora</i>
EU543151	<i>Oldenlandia wiedemannii</i>
EU145564	<i>Ophiorrhiza elmeri</i> (outgroup)
AF152617	<i>Pentanisia prunelloides</i> (outgroup)
EU543154	<i>Pentodon pentandrus</i>
EU543156	<i>Richardia scabra</i>
EU543157	<i>Richardia stellaris</i>
EU543158	<i>Spermacoce capitata</i>
EU543159	<i>Spermacoce erosa</i>
EU543160	<i>Spermacoce filituba</i>
EU543161	<i>Spermacoce flagelliformis</i>
EU543162	<i>Spermacoce hispida</i>
EU543108	<i>Spermacoce ocymifolia</i>
EU543163	<i>Spermacoce prostrata</i>
EU543164	<i>Spermacoce remota</i>
EU543165	<i>Spermacoce ruelliae</i>
EU543166	<i>Stenaria nigricans</i>
EU543168	<i>Thecorchus wauensis</i>

---



#### 4.2.1.2 Phylogenetic analysis

The aligned matrix was consisted by 893 characters, 355 of which were constant, 183 of which were variables but uninformative and 355 of which were parsimony-informative. Gaps were coded by simple indel coding method. A fifty percent majority-rule consensus tree (1000 jackknife replicates) based on 82600 trees was obtained by heuristic searches and the tree-bisection-reconnection (TBR) branch swapping algorithm (**Figure 4- 15**). The most parsimony trees were 1121 steps long, with consistency index (CI) of 0.6343, homoplasy index (HI) of 0.3657, retention index (RI) of 0.9405, rescaled consistent index (RC) of 0.5965, CI excluding uninformative characters of 0.5596, and HI excluding uninformative characters of 0.4404.

In clade one, *H. biflora* (synonymous with *Oldenlandis biflora*) was clustered to *Kadua* with 85% jackknife support. This clade was then placed as a sister group to *H. heydotis* – *H. auricularia* – *H. costata* group with weak jackknife support. In the clade two (100% jackknife support), the *H. diffusa* and *H. herbacea* in *Euoldenlandia* section formed a clade (87% jackknife support) with *Oldenlandia galioides* and *Oldenlandia lancifolia*. *H. tenelliflora*, *H. pinifolia* and *H. verticillata* also included in this clade. *H. corymbosa*, several *Oldenlandia* species and two species of *Kohautia* formed a sister clade (100% jackknife support) to *Spermacoce* group. As the *Diplophragma* was a monophyletic group showed in previous analyses, it formed a well supported group with other seven numbers in *Hedyotis* s.s. Among them, the type species, *H. fruticosa* of this section, was also included.

The maximum likelihood (ML) tree of *trnL*-F intergenic spacer with 500 bootstrap replicates was conducted by Nearest-Neighbor-Interchange (NNI) heuristic method, and it was presented in **Figure 4- 19**. Gaps were treated as complete deletion here. The ML tree of *trnL*-F intergenic spacer presented very poor resolution power in inferring the relationship between genera in Spermacoceae.





Figure 4- 16. Part one of the magnified Figure 4-15.

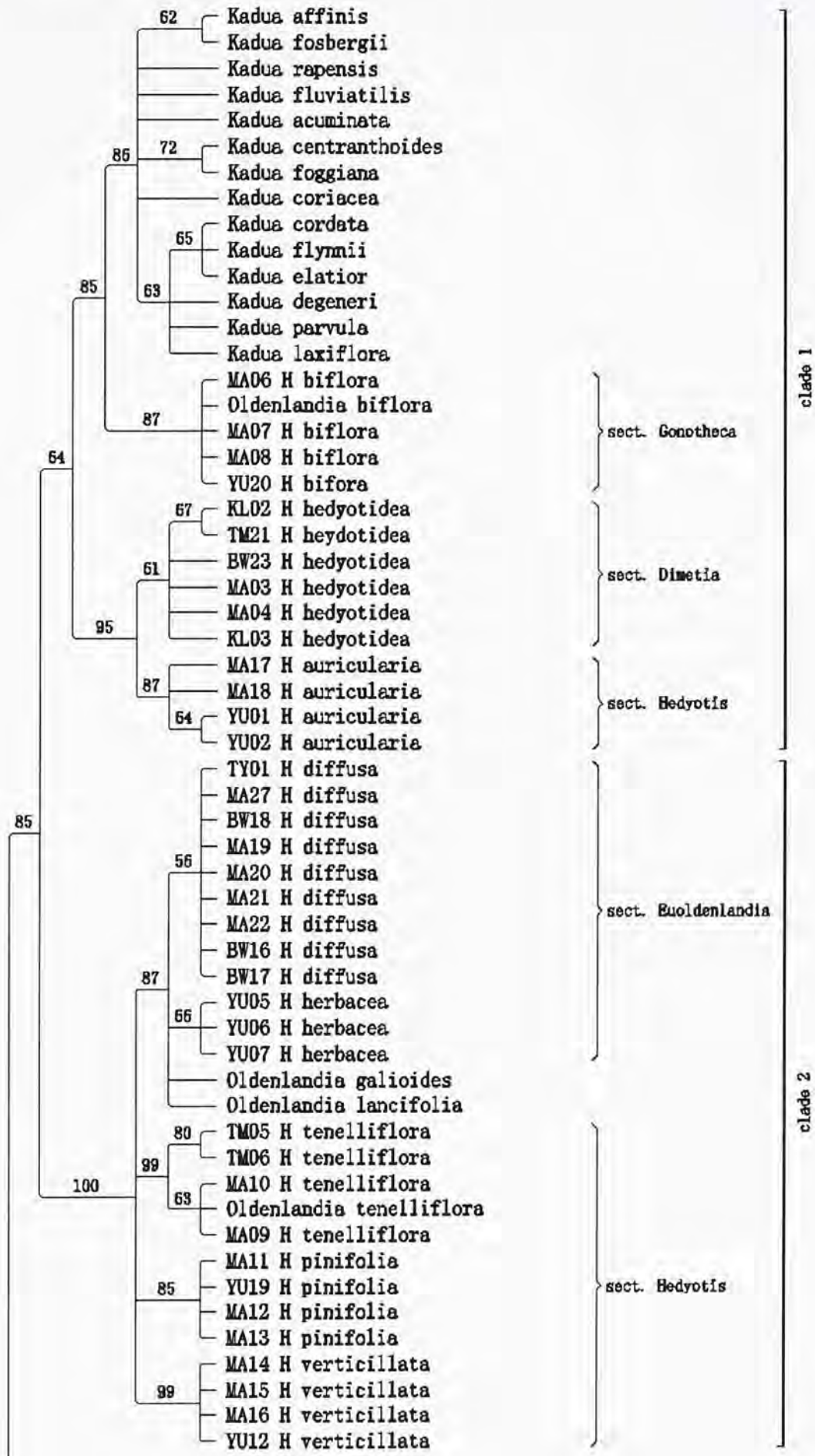




Figure 4- 17. Part two of the magnified Figure 4-15.

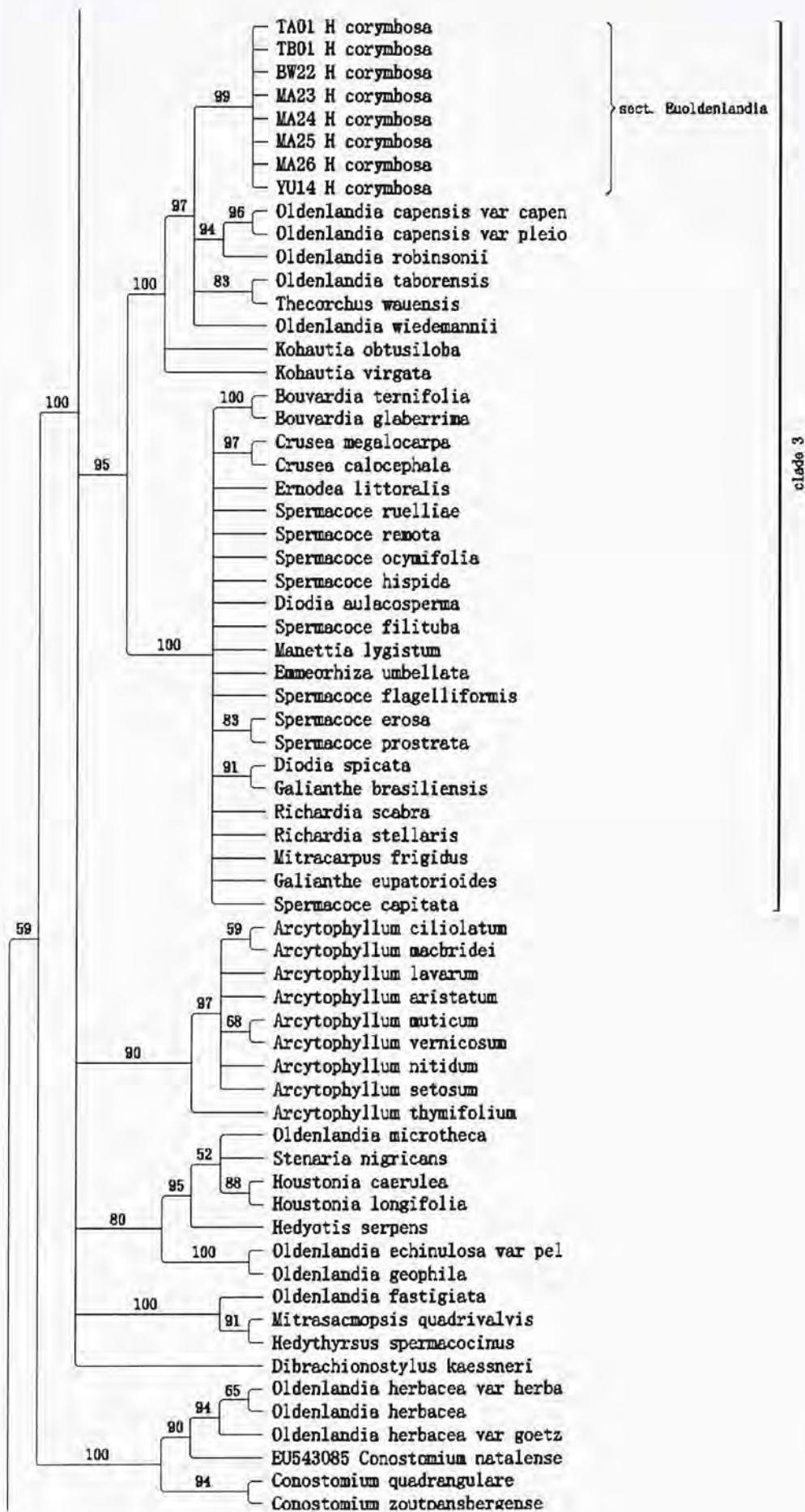


Figure 4- 18. Part three of the magnified Figure 4-15.

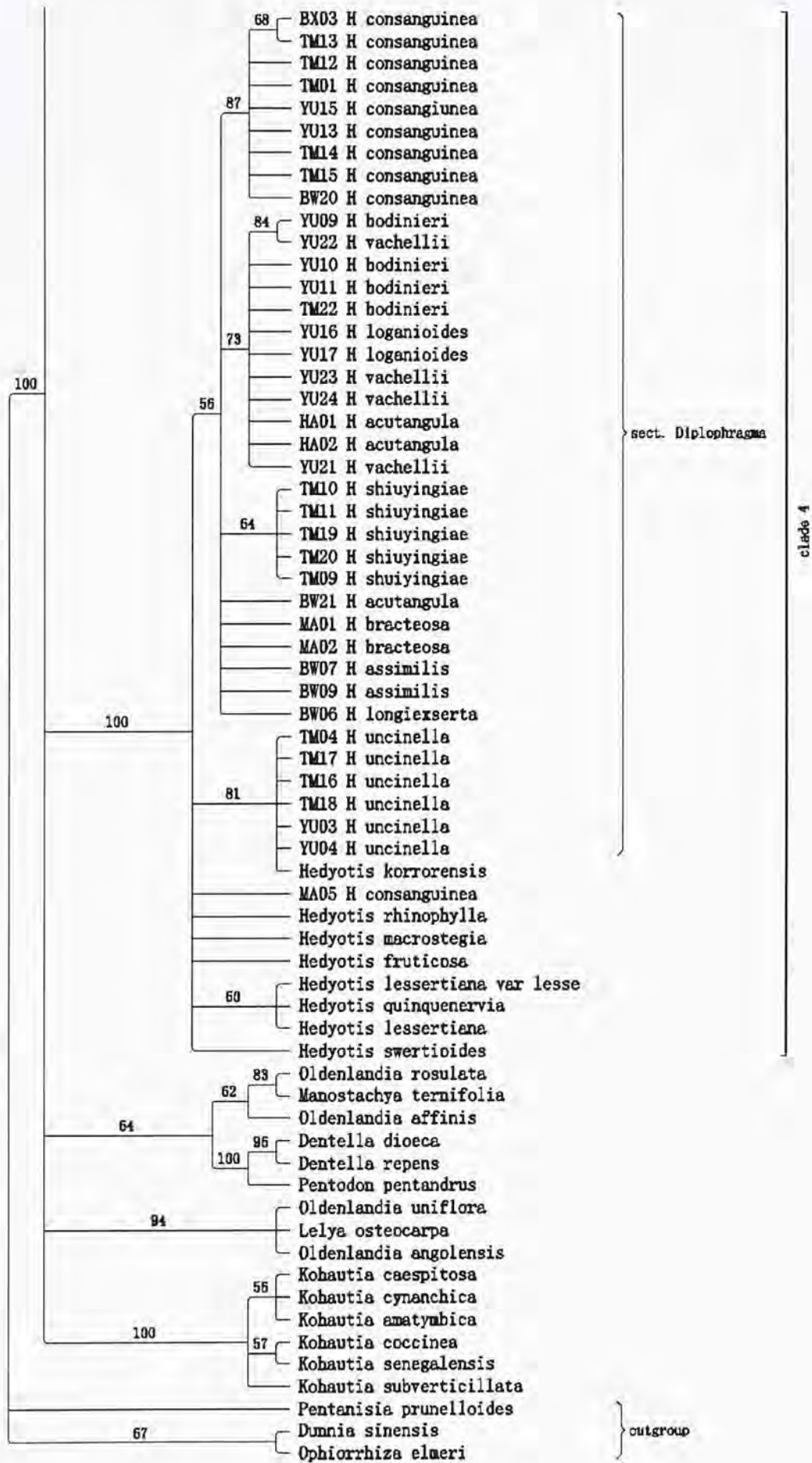




Figure 4- 19. Jackknife 50% majority-rule strict consensus tree based on the data of *trnL-F* intergenic spacer region. The jackknife value is shown above the branch. This figure was magnified to three parts shown in Figures 4- 20 – 4- 21.

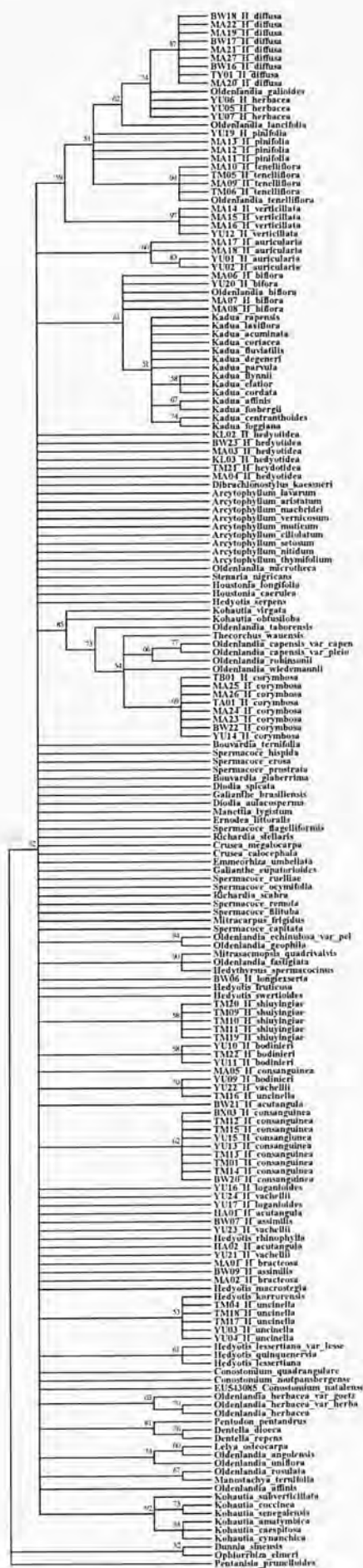


Figure 4- 20. Part three of the magnified Figure 4-19.

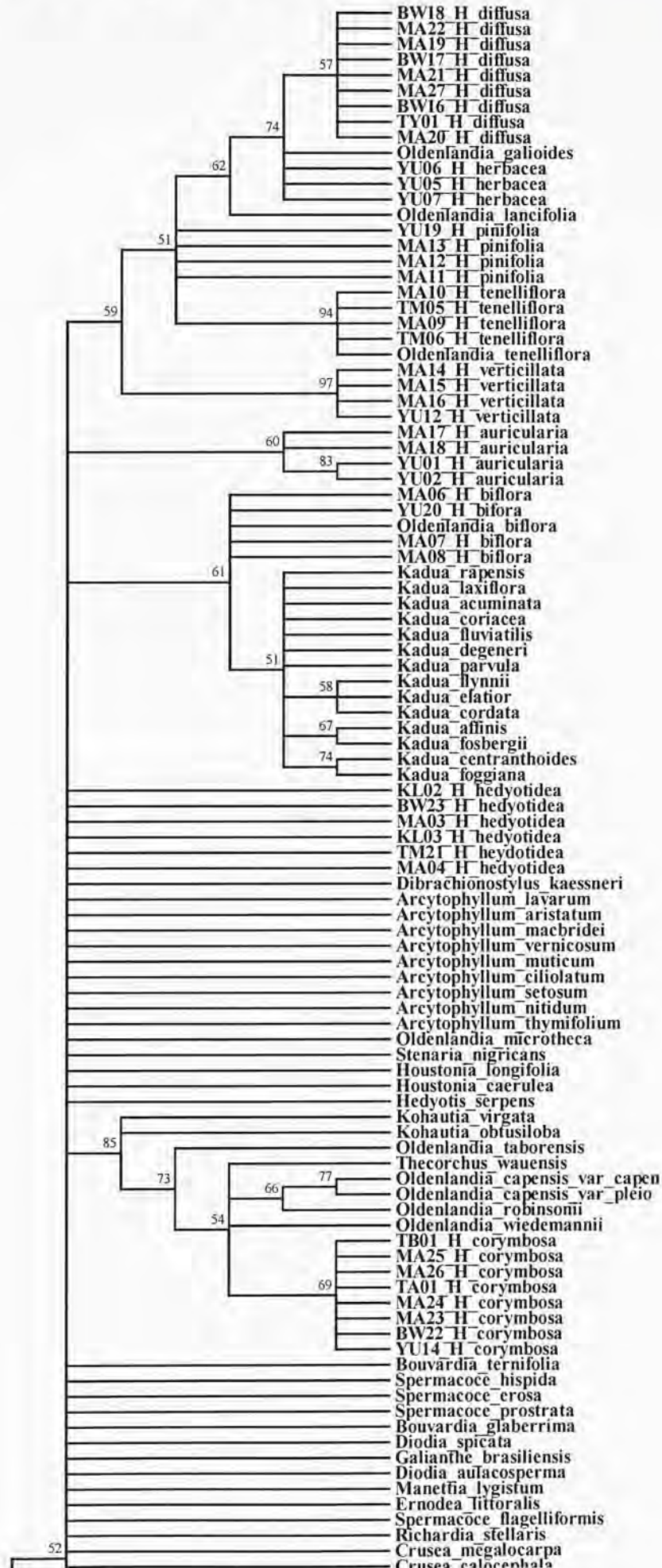
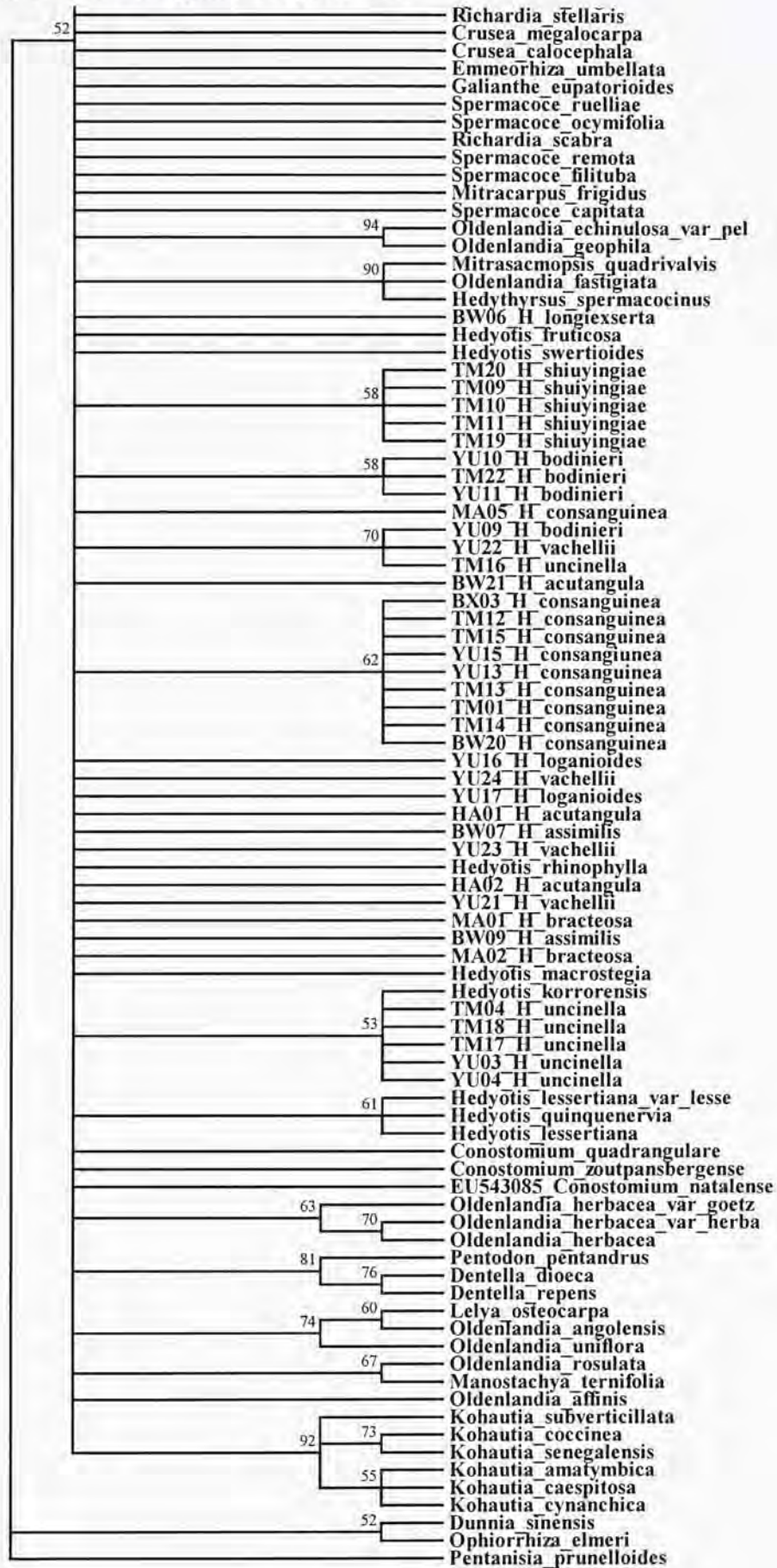




Figure 4- 21. Part three of the magnified Figure 4-15.



## 4.2.2 Plastid *rbcL* region

### 4.2.2.1 Sequence alignment

The alignment sequences matrix of *rbcL* region of 114 taxa is shown in **Appendix Table 6- 15**, sequence length ranged from 670 bp to 671 bp. Sequences obtained from Genbank of the National Center for Biotechnology Information (NCBI) are shown in **Table 4- 3**.

**Table 4- 3. List of sequences downloaded from Genbank of the National Center for Biotechnology Information (NCBI).**

Genebank accession number	Species
Z68787	<i>Agathisanthemum bojeri</i>
AJ616209	<i>Amphasma benguellense</i>
AJ288594	<i>Amphasma luzuloides</i>
AJ288595	<i>Arcytophyllum arisatum</i>
X83626	<i>Bouvardia glaberrima</i>
AJ616210	<i>Conostomium longitubum</i>
Z68792	<i>Conostomium quadrangulare</i>
AJ616211	<i>Dibrachionostylus kaessneri</i>
AJ288600	<i>Diodia sarmentosa</i>
EU145467	<i>Dunnia sinensis</i> (outgroup)
AJ288601	<i>Ernodea littoralis</i>
Z68799	<i>Hedyotis fruticosa</i>
AJ288605	<i>Hedyotis littoralis</i>
AJ288606	<i>Hedyotis nigricans</i>
AM117232	<i>Hedythyrus thamnoideus</i>
AJ288607	<i>Hemidiodia ocimifolia</i>
AJ616212	<i>Hydrophylax maritima</i>
Z68800	<i>Kohautia caespitosa</i>
AJ288609	<i>Lelya prostrata</i>
Z68803	<i>Manettia bicolor</i>
AJ616213	<i>Manostachya ternifolia</i>
AJ616214	<i>Mitrasacmopsis quadrivalvis_1</i>
AM117248	<i>Mitrasacmopsis quadrivalvis_2</i>



Z68808	<i>Oldenlandia goreensis</i>
EU145464	<i>Ophiorrhiza elmeri</i> (outgroup)
Z68813	<i>Pentanopsis fragrans</i>
AJ616208	<i>Pentanopsis gracilicaulis</i>
X83660	<i>Pentanopsis pentandrus</i>
AM117255	<i>Pentania prunelloides</i> (outgroup)
AJ288619	<i>Psyllocarpus laricoides</i>
Z68820	<i>Richardia pilosa</i>
AJ288623	<i>Spermacoce hispida</i>
Z68823	<i>Spermacoce tenuior</i>
Y18720	<i>Synaptantha tillaeacea</i>
AM117282	<i>Thecorchus wauensis</i>

#### 4.2.2.2 Phylogenetic analysis

Gaps were coded by simple indel coding method, and 675 characters (521 constant characters, 62 variables but uninformative characters, and 92 parsimony-informative characters) were got from this alignment matrix. After measurement with heuristic searches and the tree-bisection-reconnection (TBR) branch swapping algorithm, a jackknife 50% majority-rule consensus tree with 1000 jackknife replicates based on 98509 trees is shown in **Figure 4- 22**. The individual rooted tree was 334 steps long, with consistency index (CI) of 0.5569, homoplasy index (HI) of 0.4431, retention index (RI) of 0.8263, and rescaled consistent index (RC) of 0.4602. CI excluding uninformative characters was 0.4436 while HI excluding uninformative characters was 0.5564.

The sequences source of *rbcL* region of Spermacoceae was limited in GenBank compared to the *trnL-F* intergenic spacer and *rps16* region. Moreover, the *rbcL* region showed very poor resolution in inferring the phylogenetic relationship. The

position of *Diplophragma* section was placed closely with the *H. fruticosa* with 90% jackknife support. *H. corymbosa* formed a monophyletic group with low jackknife support. The *H. hedyotideia*, *H. auricularia* and *H. costata* formed a group with 86% jackknife support. *H. biflora* formed a distinct group in this cladogram with 100% jackknife support. The group of *H. diffusa*, *H. herbacea*, *H. pinifolia*, *H. tenelliflora* and *H. verticillata* showed similar situation which formed a monophyletic group with low jackknife support.

The maximum likelihood (ML) tree of *rbcL* region with 500 bootstrap replicates was conducted by Nearest-Neighbor-Interchange (NNI) heuristic method, and it was presented in **Figure 4- 26**. Gaps were treated as complete deletion here. The ML tree of *rbcL* intergenic spacer presented very poor resolution power in inferring the relationship between genera in Spermaceae.



Figure 4- 22. Jackknife 50% majority-rule strict consensus tree with 1000 replicates based on the data of *rbcL* region. This figure was magnified to three parts showed in Figures 4-23 - 4-25.

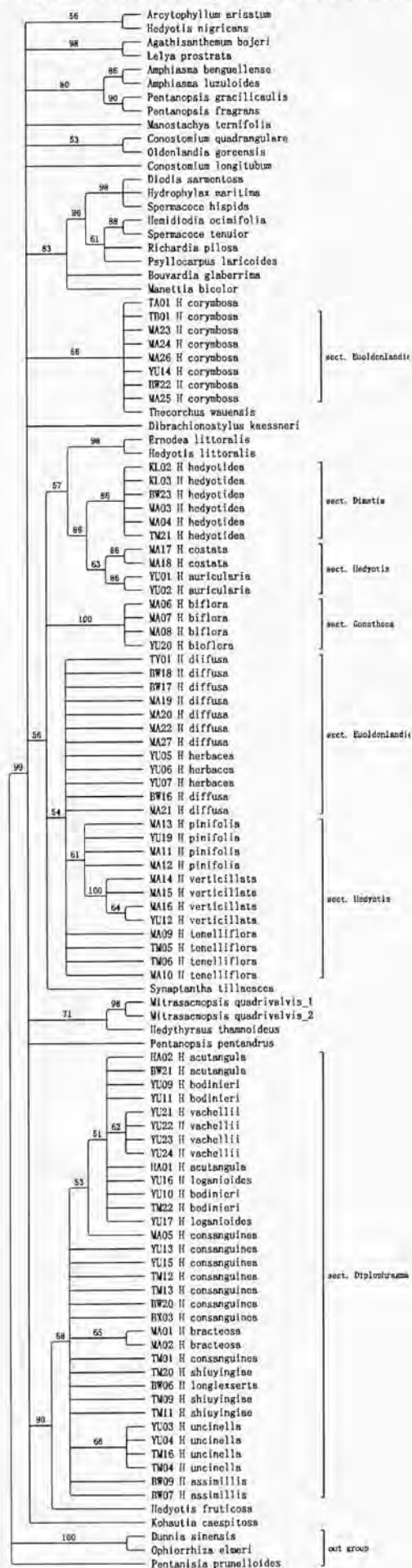


Figure 4- 23. Part one of the magnified Figure 4-22.

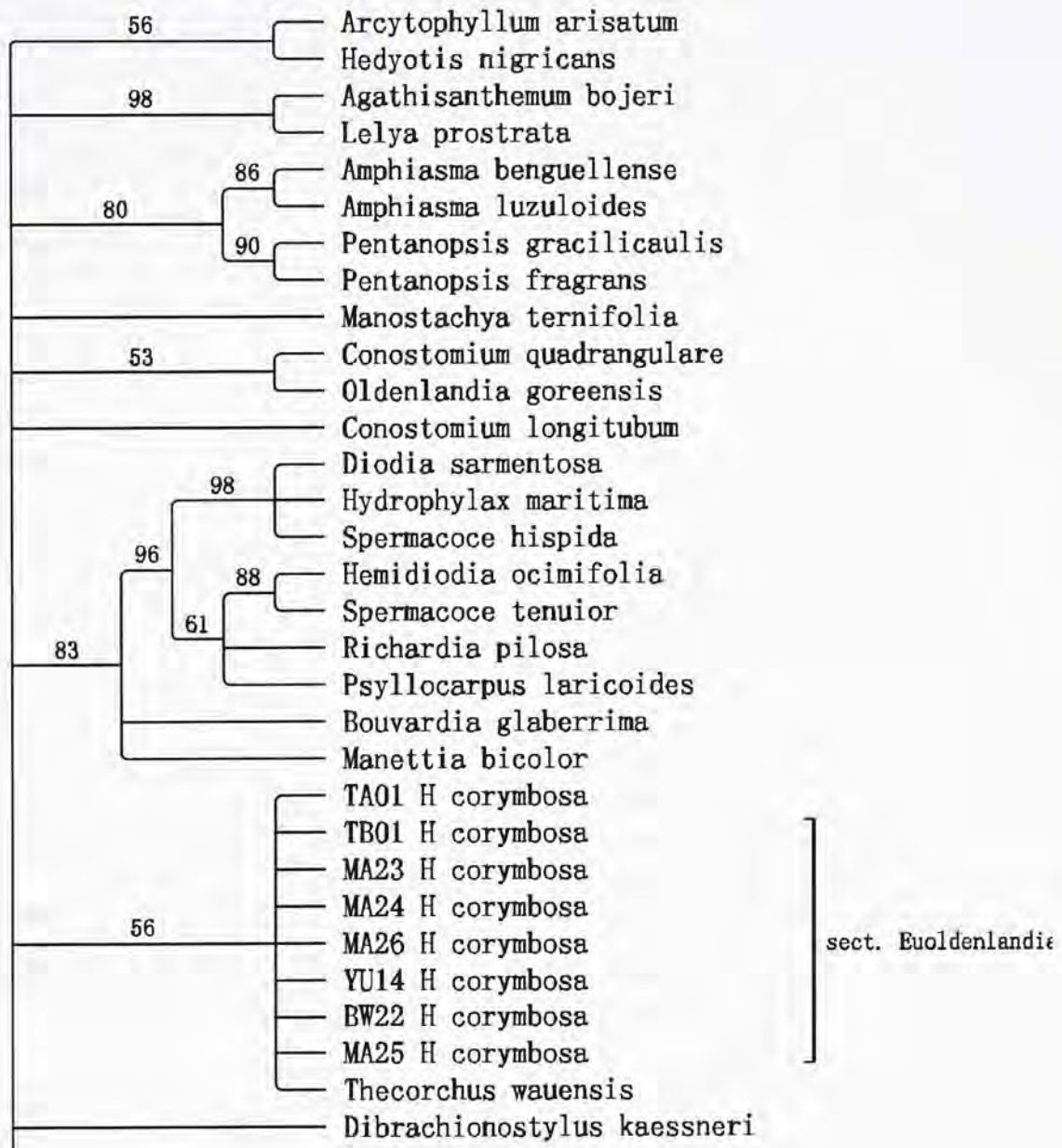




Figure 4- 24. Part two of the magnified Figure 4-22.

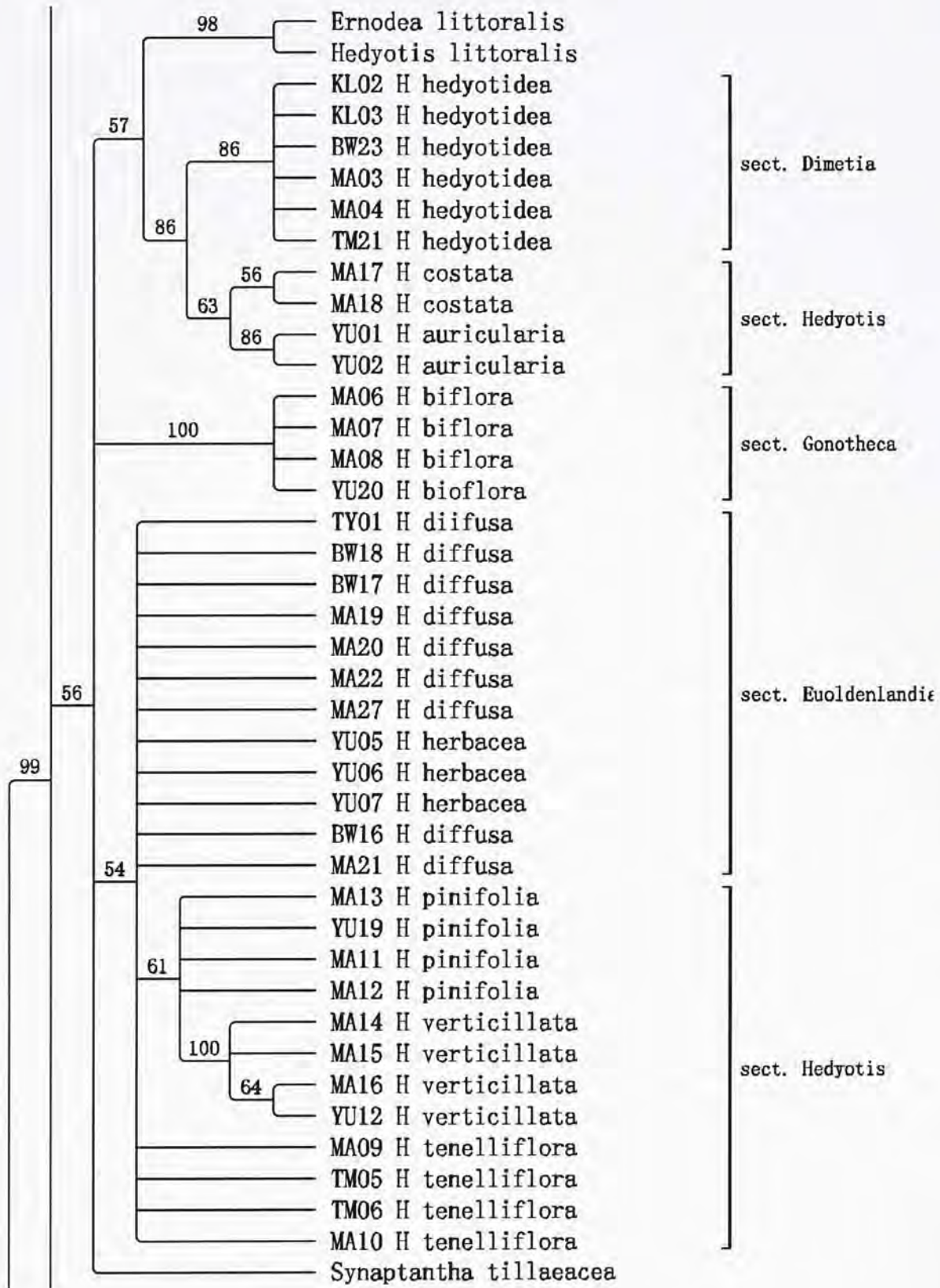
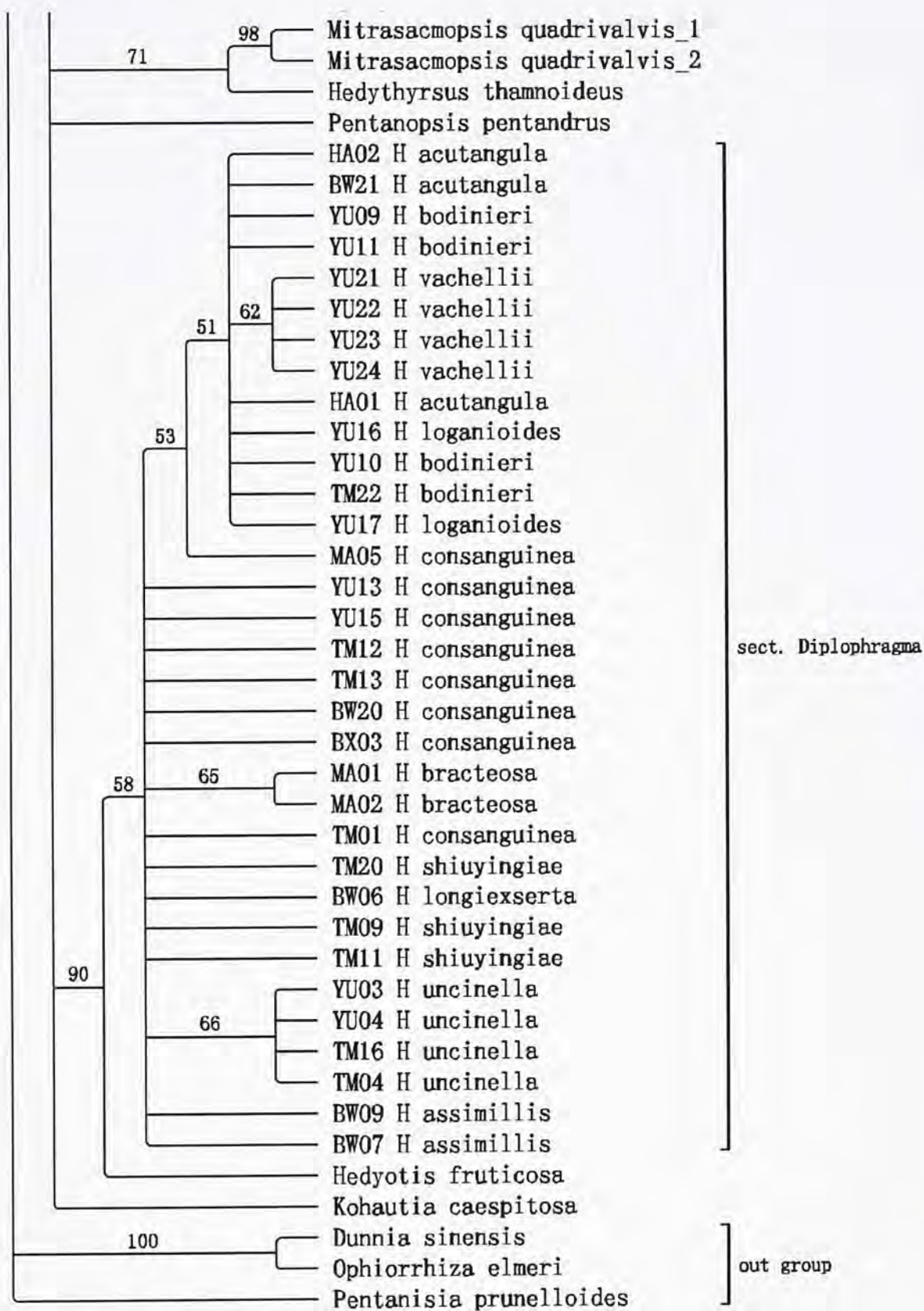


Figure 4- 25. Part three of the magnified Figure 4-22.







## 4.2.3 Plastid *rps16* region

### 4.2.3.1 Sequence alignment

The sequence length of *rps16* region was about 518 bp to 483 bp, and the alignment matrix is presented in **Appendix Table 6- 16**. Sequences selected from Genbank of the National Center for Biotechnology Information (NCBI) are listed below.

**Table 4- 4. List of sequences downloaded from Genbank of the National Center for Biotechnology Information (NCBI).**

Genebank accession number	Species
EU543018	<i>Agathisanthemum bojeri</i>
EU543019	<i>Agathisanthemum globosum</i>
AF002753	<i>Amphiasma benguellense</i>
EU543020	<i>Amphiasma luzuloides</i>
AF333348	<i>Arcytophyllum aristatum</i>
AF333350	<i>Arcytophyllum ciliolatum</i>
AF333352	<i>Arcytophyllum ericoides</i>
AF333354	<i>Arcytophyllum erilavarum</i>
AF333356	<i>Arcytophyllum macbridei</i>
AF002754	<i>Arcytophyllum muticum</i>
AF333359	<i>Arcytophyllum nitidum</i>
AF333362	<i>Arcytophyllum rivetii</i>
AF333364	<i>Arcytophyllum serpyllaceum</i>
AF002755	<i>Arcytophyllum setosum</i>
AF333367	<i>Arcytophyllum thymifolium</i>
AF333368	<i>Arcytophyllum vernicosum</i>
AF002757	<i>Borreria hispida</i>
EU543022	<i>Bouvardia glaberrima</i>
AF002758	<i>Bouvardia ternifolia</i>
AF002760	<i>Conostomium natalense</i>
AM266812	<i>Conostomium quadrangulare</i>
AF002759	<i>Crusea hispida</i>
EU543025	<i>Crusea megalocarpa</i>
AF333370	<i>Dentella repens</i>
AF002761	<i>Dibrachionostylus kaessneri</i>



---

EU543026	<i>Diodia aulacosperma</i>
AF002762	<i>Diodia sarmentosa</i>
EU543027	<i>Diodia spicata</i>
AY764288	<i>Diodia virginiana</i>
EU145519	<i>Dunnia sinensis</i> (outgroup)
AY764289	<i>Emmeorhiza umbellata</i>
AF002763	<i>Ernodea littoralis</i>
AY764290	<i>Galianthe brasiliensis</i>
EU543028	<i>Galianthe eupatorioides</i>
AF333371	<i>Hedyotis degeneri</i>
AF002765	<i>Hedyotis hillebrandii</i>
EU543029	<i>Hedyotis lessertiana</i> var <i>lessertiana</i>
EU543030	<i>Hedyotis lessertiana</i> var <i>marginata</i>
AF002767	<i>Hedyotis macrostegia</i>
AF333373	<i>Hedyotis nigricans</i>
AF333375	<i>Hedyotis parvula</i>
AF333376	<i>Hedyotis schlechtendahliana</i>
AF333377	<i>Hedyotis serpens</i>
EU543031	<i>Hedyotis swertioides</i>
EU543032	<i>Hedythyrus spermacocinus</i>
AF333379	<i>Houstonia caerulea</i>
AF002766	<i>Houstonia longifolia</i>
AM117316	<i>Hydrophylax maritima</i>
EU642523	<i>Kadua affinis</i>
EU642524	<i>Kadua axillaris</i>
EU543033	<i>Kadua centranthoides</i>
EU642525	<i>Kadua coriacea</i>
EU642526	<i>Kadua elatior</i>
EU642527	<i>Kadua fluviatilis</i>
EU642528	<i>Kadua flynnii</i>
EU642529	<i>Kadua fosbergii</i>
EU642530	<i>Kadua laxiflora</i>
EU543034	<i>Kadua littoralis</i>
EU642531	<i>Kadua rapensis</i>
EU543035	<i>Kohautia amatymbica</i>
AM117324	<i>Kohautia caespitosa</i>
EU543037	<i>Kohautia coccinea</i>
EU543038	<i>Kohautia cynanchica</i>
EU543041	<i>Kohautia subverticillata</i>
AF002769	<i>Manettia aff lygistum</i> Andersson
AF002768	<i>Manettia alba</i>
AM117328	<i>Manostachya ternifolia</i>
AF002770	<i>Mitracarpus frigidus</i>
EU543044	<i>Mitracarpus microspermus</i>

---

---

AM117329	<i>Mitrasacmopsis quadrivalvis</i>
EU543045	<i>Mitrasacmopsis quadrivalvis</i>
AF003607	<i>Nesohedyotis arborea</i>
AF003608	<i>Oldenlandia affinis</i>
EU543047	<i>Oldenlandia angolensis</i>
EU567459	<i>Oldenlandia biflora</i>
EU543048	<i>Oldenlandia capensis var capensis</i>
EU543049	<i>Oldenlandia capensis var pleiosepala</i>
EU543052	<i>Oldenlandia fastigiata</i>
EU543053	<i>Oldenlandia galioides</i>
EU543054	<i>Oldenlandia geophila</i>
EU543055	<i>Oldenlandia goreensis</i>
EU543056	<i>Oldenlandia herbacea var goetzei</i>
EU543057	<i>Oldenlandia herbacea var herbacea</i>
EU543058	<i>Oldenlandia lancifolia</i>
EU543059	<i>Oldenlandia microtheca</i>
EU543060	<i>Oldenlandia nematocaulis</i>
AF333382	<i>Oldenlandia nervosa</i>
EU543061	<i>Oldenlandia robinsonii</i>
EU543062	<i>Oldenlandia tenelliflora</i>
AY764293	<i>Oldenlandia tenuis</i>
AY764294	<i>Oldenlandia thesiifolia</i>
AY764295	<i>Oldenlandia uniflora`</i>
EU543063	<i>Oldenlandia wiedemannii</i>
EU145510	<i>Ophiorrhiza elmeri</i> (outgroup)
AM266860	<i>Pentanisia prunelloides</i> (outgroup)
EU543065	<i>Pentanopsis fragrans</i>
EU567460	<i>Pentanopsis gracilicaulis</i>
EU543066	<i>Pentodon pentandrus</i>
AF003614	<i>Richardia scabra</i>
EU543068	<i>Richardia stellaris</i>
EU543069	<i>Spermacoce capitata</i>
AF003619	<i>Spermacoce confusa</i>
EU543071	<i>Spermacoce filituba</i>
EU543072	<i>Spermacoce flagelliformis</i>
EU543073	<i>Spermacoce hispida</i>
EU543074	<i>Spermacoce ruelliae</i>
EU543075	<i>Synaptantha tillaeacea</i>
AM266901	<i>Thecorchus wauensis</i>

---



#### 4.2.3.2 Phylogenetic analysis

The sequences of *rps16* region from 186 taxa were included in this study. Gaps were coded by simple indel coding method. The aligned matrix was comprised of 744 characters, 338 of which were constant, 169 of which variables but uninformative and 237 of which were parsimony-informative. Based on 96100 trees, a jackknife 50% majority-rule consensus tree (1000 jackknife replicates) of *Hedyotis* based on *rps16* region conducted in PAUP\* 4.0b10 with heuristic searches and the tree-bisection-reconnection (TBR) branch swapping algorithm is presented in **Figure 4- 12**. The individual rooted tree was 875 steps long, with consistency index (CI) of 0.6171, homoplasy index (HI) of 0.3829, retention index (RI) of 0.8930, and rescaled consistent index (RC) of 0.5511. CI excluding uninformative characters was 0.5145 while HI excluding uninformative characters was 0.4855.

In this cladogram, three clades which included the *Hedyotis* (species in China) were selected for further analysis. The group of *H. diffusa*, *H. herbaces* in *Euoldenlandia* section, *Oldenlandia galioides* and *Oldenlandia lancifoli* was clustered with *H. pinifolia*, *H. tenelliflora* and *H. verticillata* to form a monophyletic group with 100% jackknife support. The rest members in this clade also included *H. biflora* (synonymous with *Oldenlandis biflora*), *Kadua* genus, *H. heydotis* – *H. auricularia* – *H. costata* group, *H. biflora* and two *Hedyotis* species. In the clade two, *H. corymbosa* and several *Oldenlandia* species formed a distinct group (100% jackknife support) which was placed as a sister to *Spermacoce* group. The *Diplophragma* was a

monophyletic group shown in previous analyses, and it formed a well supported group with *H. fruticosa* as the type species in *Hedyotis* s.s.

The maximum likelihood (ML) tree of *rbcL* region with 500 bootstrap replicates was conducted by Nearest-Neighbor-Interchange (NNI) heuristic method, and it was presented in **Figure 4- 31**. Gaps were treated as complete deletion here. Compared to jackknife of *rps16* region, the ML tree presented a poorer resolution power in inferring the relationship in Spermaceae.





Figure 4- 28. Part one of the magnified Figure 4-27.

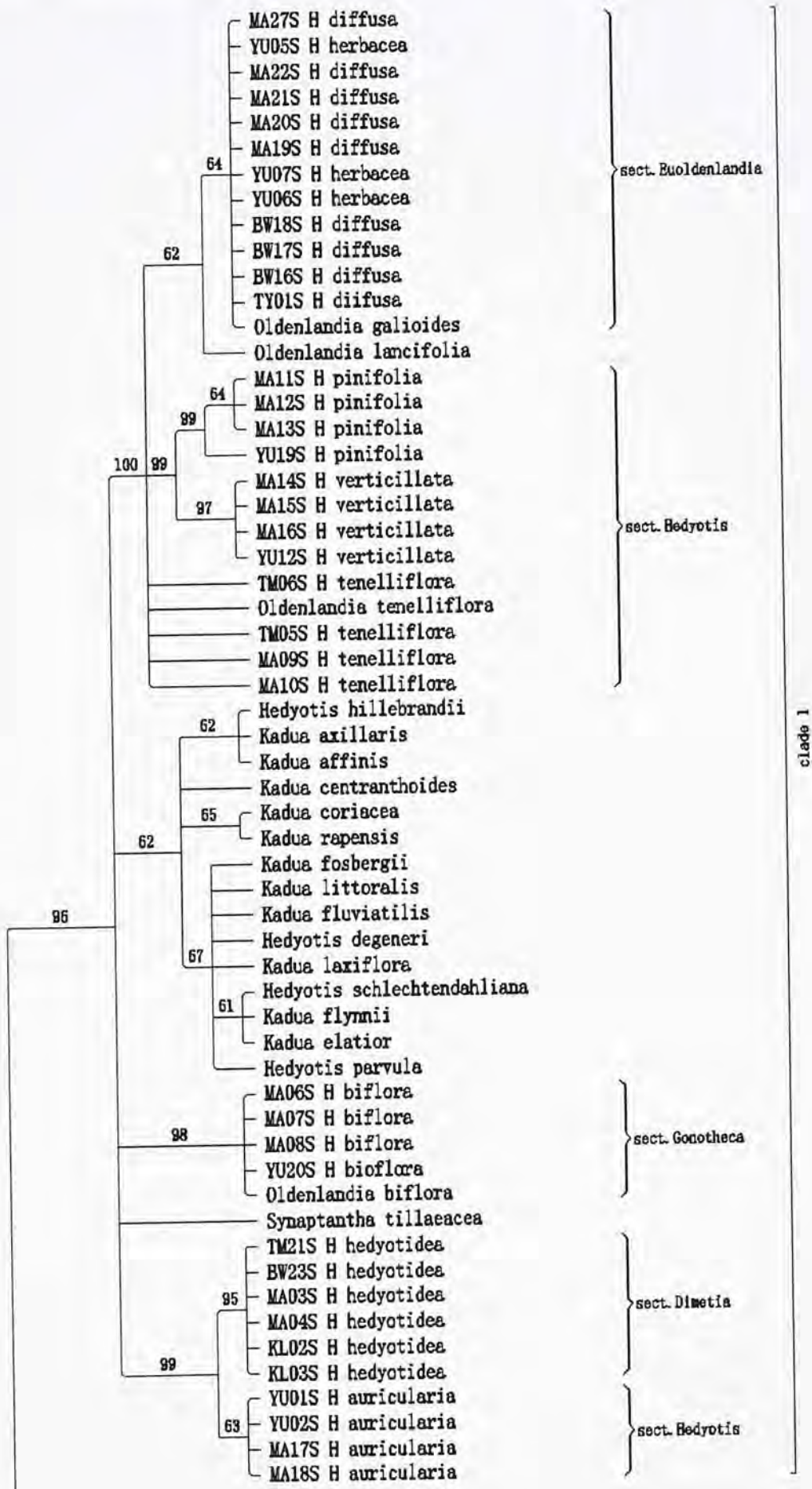




Figure 4- 29. Part two of the magnified Figure 4-27.

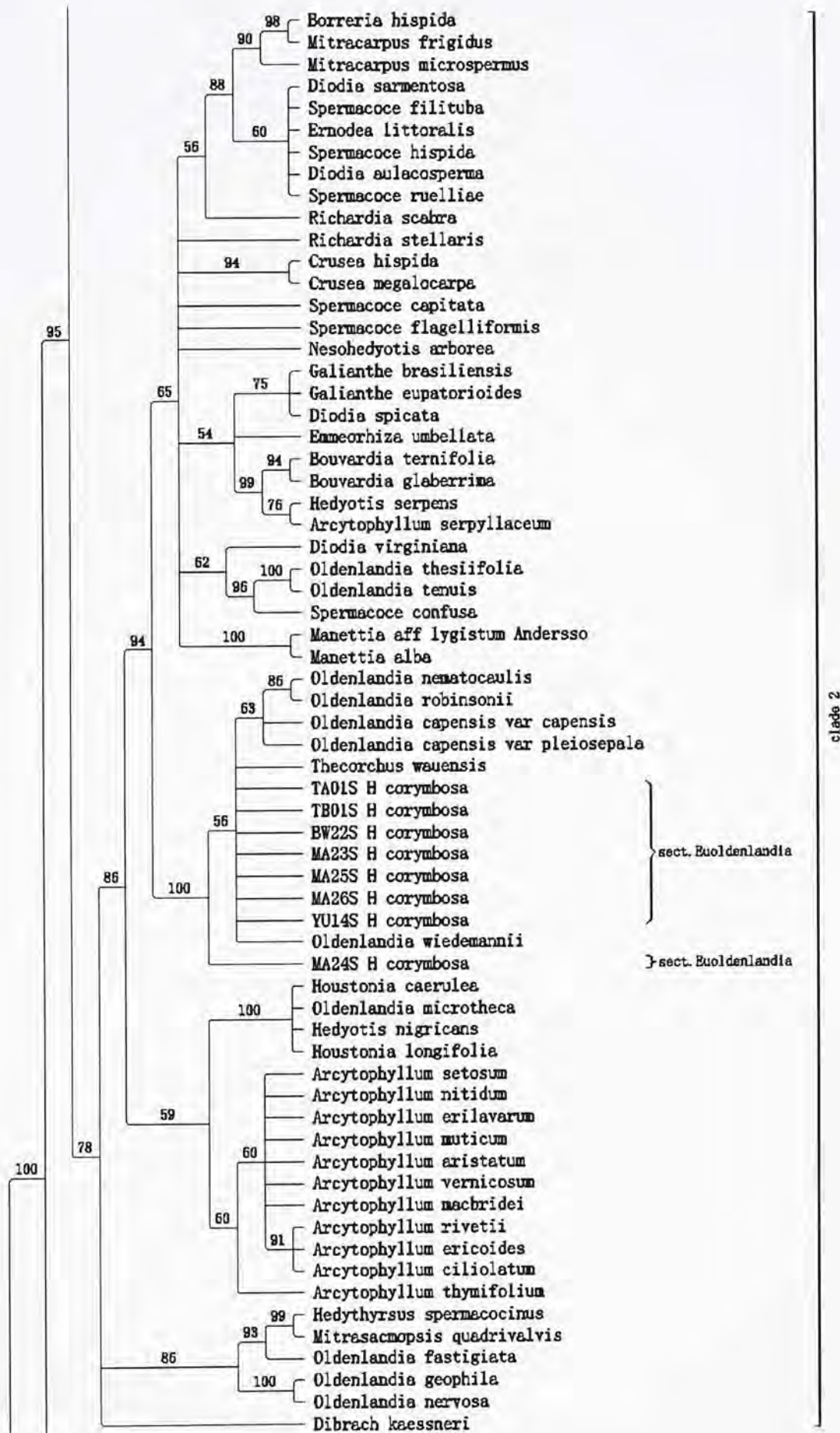


Figure 4- 30. Part three of the magnified Figure 4-27.

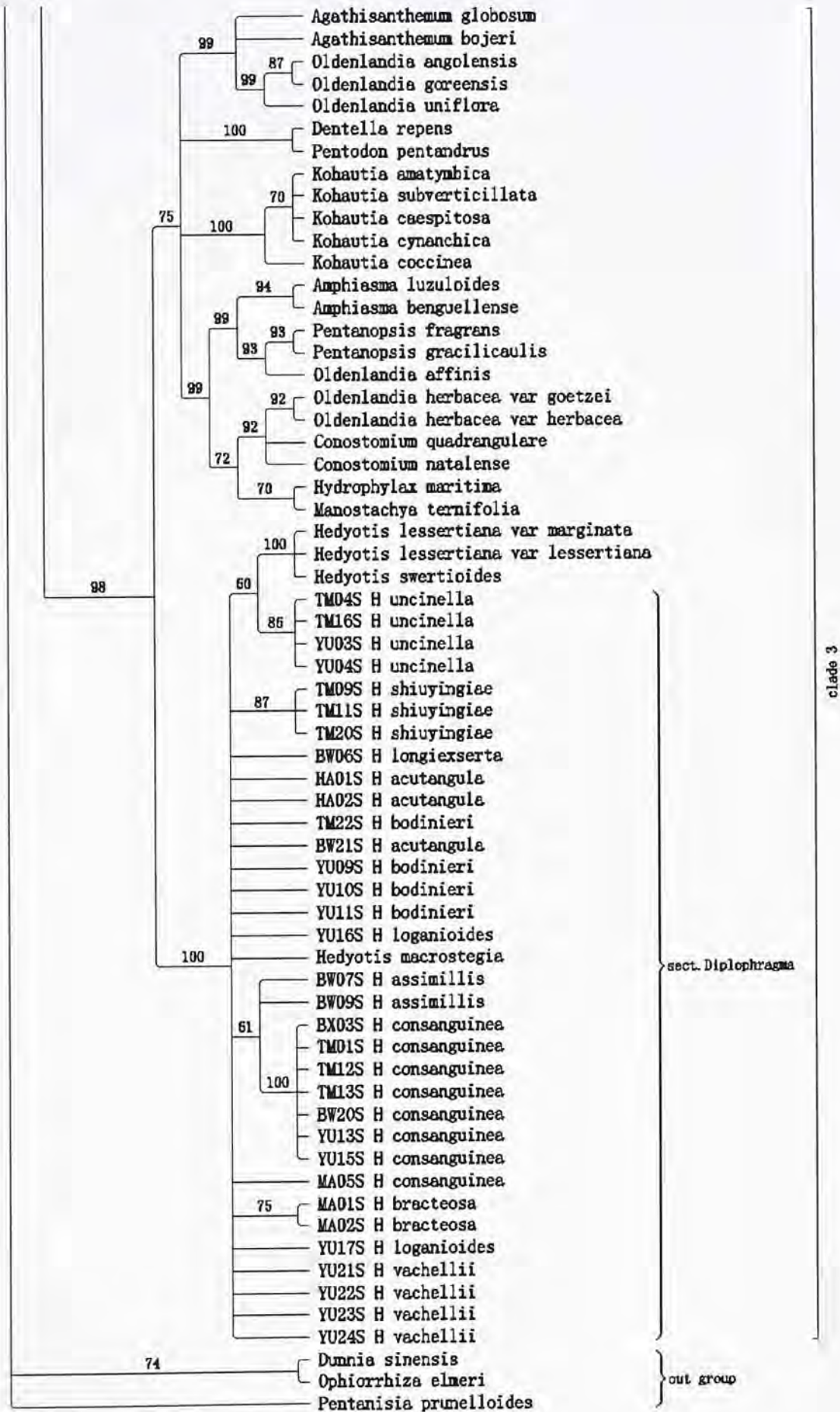




Figure 4- 31. Jackknife 50% majority-rule strict consensus tree based on the data of *rps16* region. The jackknife value is shown above the branch.

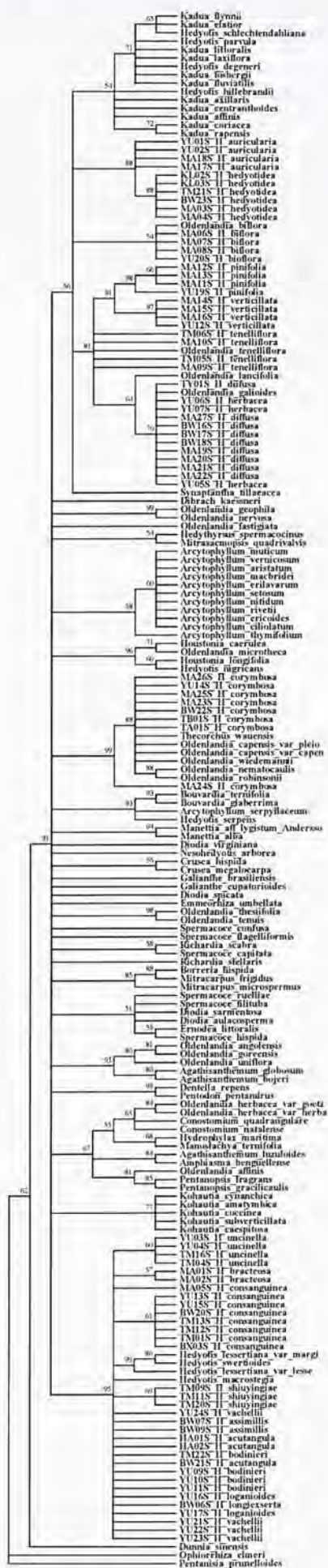
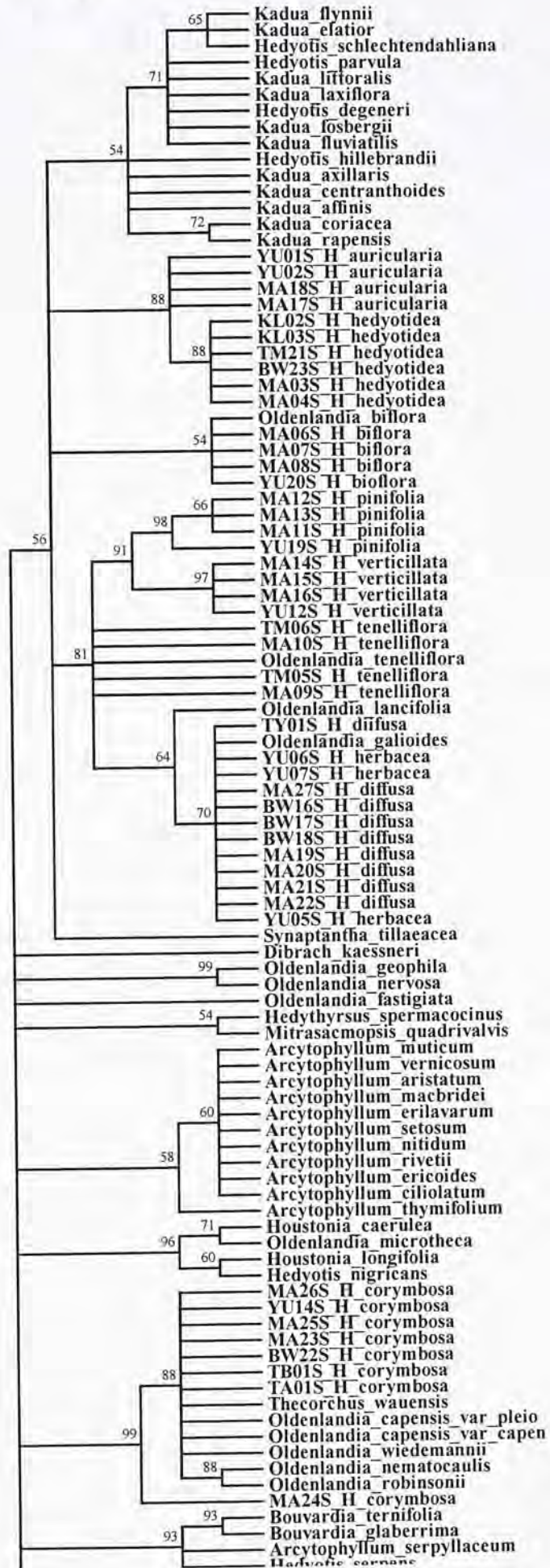


Figure 4- 32. Part three of the magnified Figure 4-31.







## 4.3 Discussion

### 4.3.1 Comparison of phylogenetic utility of the six DNA regions

In the previous molecular phylogenetic studies of Rubiaceae, 13 DNA regions including six nuclear DNA regions (ETS, ITS, NTS, *Tpi*, *pep-C* large and *pep-V* small) and 7 plastid DNA regions (*atpB-rbcL*, *ndhF*, *matK*, *rbcL*, *rps16*, *trnL-F*, *trnS-G*) (Bremer 2009) were involved, and thousands of sequences from those regions are available from GenBank on internet. Between them, *rbcL* region is very useful for family and subfamilies classification, whereas ITS, *rps16* and *trnL-F* regions are the most popular and very informative sequence markers for tribal and generic delimitation. Referring to this study, sequences of *rbcL* providing very little variability were presented very poor resolution in taxonomic relationship of *Hedyotis*. The power of ITS region was proved in resolving lower level of the hierarchy, especially for the relationship between or within genera (Stuessy 2009). In this study, sequences of ITS which produced large amount of informative characters showed good resolution in inferring the phylogenetic relationship of *Hedyotis*.

Molecular regions with a considerable amount of informative characters were chosen to compare their utilities in inferring the phylogenetic relationship of a target group. Usually, a phylogenetic study preferred a longer but less variable region to a short highly variable region (Karehed, Groeninckx et al. 2008). As with *trnH-psbA* region, it was one of the highest variable regions with fast evolution rate. However, the



limited length of this region does not produce enough characters for a fully resolved phylogeny. In this study, *trnH-psbA* region was selected to investigate the phylogenetic relationship of *Hedyotis* and compare the utility of solving phylogenetic problems with other molecular regions. The resolution quality of this region is moderate, because of its short aligned sequence length and problematic alignment of highly variable sequences.

The remaining regions, noncoding plastid gene *trnL*, *trnL-F* and *rps16* and protein coding plastid gene *matK* presented moderate resolution in inferring the phylogenetic relationship for *Hedyotis*.

### **4.3.2 Diplophragma section**

According to the *Flora Republicae Popularis Sinicae* (Ko 1999), the section of *Diplophragma* is characterized by septicidal dehiscent capsule and flat capsule apex. In the study of Li in 2010, five species in *Diplophragma* (included one new species, *H. shiuyingiae*, published in 2008) formed a monophyletic group based on ITS region (Li, Jiang et al. 2010). In this study, ten species in *Diplophragma* were included. They formed a strongly supported monophyletic clade (97%-100% jackknife support) basing on six single-gene cladograms and one combined datasets cladogram. As the poor resolution of five plastid DNA regions for this section and the conflict results between nrITS and combined datasets, the phylogenetic relationship of this section is still unclear.

However, the position of several species and confusing taxonomic issue of this section was still worth discussion here.

*H. bodinieri*, *H. shiuyingiae* and *H. vachellii* are endemic to Hong Kong. The species of *H. shiuyingiae* was published by Chen in 2008, and he included this species to *Diplophragma* section for its septicidal dehiscent capsule. Actually, it was first named as *H. taipoensis* in 1971 and then named as *H. Hongkongensis* in 1993 by Dr. S. Y. Hu, but, both names have not published (Chen 2008). Thus, it was confused with the species of *H. yangchensis* in Hong Kong Herbarium for a long time, until Chen recognized it as a new species and gave a name to it. In this study, the position of it in this section was confirmed by different molecular datasets.

Morphologically, *H. consanguinea* and *H. acutangular* were found very similar to each other even after careful examination of their leaf, stipule, flower and capsule type. The only difference is that *H. acutangular* is woody at base and branched above the woody stem. Regarding to the molecular analyses here, *H. consanguinea* was always separated far away from *H. acutangular*. In the cladograms of nrITS region and combined datasets, *H. acutangula* formed a monophyletic group with *H. vachellii*, *H. shiuyingiae*, *H. bracteosa*, *H. bodinieri* and *H. loganioides*. Interestingly, the result also showed that *H. consanguinea* collected from Macao was split from those from Hong Hong. As we only got a non-flowering *H. consanguinea* sample from Macao, further confirmation was hampered. As the type specimen of *H. consanguinea* was collected from Macao, morphological comparison for further species verification is very important. The close position of *H. bodinieri* and *H.*



*loganioides* were supported by their molecular results. They were nested with each other in the ITS region cladogram, and related closely in cladograms of cpDNA regions as well as the combined dataset cladogram. Their close relationship was also supported by their morphological similarity based on the features of leaf, flower and fruit. The only difference of them was: samples of *H. bodinieri* collected from the top of Tai Mo Shan with short stem and dense branches while the *H. loganioides* collected from Ma On Shan with sparse branches and taller stems. However, this difference could be affected by the geographic and environmental factors. Thus, a more deep morphological examination for them is in need.

A broader sampling for this section is necessary, and the further morphological examination for precise section delimitation is also required.

### **4.3.3 Hedyotis, Dimetia, Euoldendandia and Gonotheca sections**

Five species in Hedyotis section, one species in Dimetia section, three species Euoldenlandia in section and one species in section Gonotheca were collected to this study. Based on the result of different sets of molecular data, species of four sections formed a monophyletic group separating from Diplophragma section. Four cpDNA region (except the *rbcL* region) and one nuclear DNA region all presented clear phylogenetic relationship for the four sections. According to the results shown in this study, Euoldenlandia section was not a monophyletic group. Also, *H. corymbosa* was basal to this clade and *H. diffusa* and *H. herbacea* clustered with the species (*H.*

*teneflora*, *H. pinifolia* and *H. verticillata*) in Hedyotis section. The remaining species in Hedyotis section, *H. auricularia* and *H. costata* were grouped with *H. hedyotidea* from section Dimetia.

The relationship of *H. diffusa* and *H. corymbosa* is always a puzzle for many decades. From the morphological perspective, they are very similar except for the number of flowers in axillary position. However, the two species are quite different at the molecular level. As shown in the cladogram (Figs 4.1-4.7), they are separated by other species far away from each other. Digging more diagnostic characters for taxonomic studies is necessary in inferring the phylogenetic relationship of this two species.

Referring to the *Flora Republicae Popularis Sinicae* (Ko 1999), the Hedyotis section is defined by an ambiguous character state, which is indehiscent or later dehiscent capsule. For the species of *H. auricularia* and *H. costata* which has indehiscent capsule, they were grouped with *H. hedyotidea* in section Dimetia that has septicidal capsules with apical protrusion (Li, 2010). We only sampled one species of section Dimetia in which, the other species are still veiled. *H. teneflora*, *H. pinifolia* and *H. verticillata* in the section Hedyotis are characterized by loculicidal dehiscent capsules. They showed close relationship with the *H. diffusa* and *H. herbacea* in Euoldenlandia section which also has the loculicidal dehiscent capsule. However, *H. biflora* in section Gonotheca with loculicidal dehiscent capsule formed a monophyletic group, but it showed variable positions in different cladograms.



#### 4.3.4 The position of *Hedyotis* (species in China) in Spermaceae

As mentioned in the introduction, the delimitation of *Hedyotis* in the broad sense has long been problematic. In *Flora Republicae Popularis Sinicae*, genus *Hedyotis* is consisted of species of *Hedyotis* subgenus and *Oldendandia* subgenus. All the species in *Euoldenlandia* and *Gonotheca* section actually belong to the subgenus of *Oldenlandia* while some members in *Hedyotis* section also belong to it. In previous molecular studies, both *Hedyotis* s.s. and *Oldendandia* s.s. were found as polyphyletic groups (Karehed, Groeninckx et al. 2008; Groeninckx, Dessein et al. 2009). In this study, the polyphyletic position of *Hedyotis* (collected from China) was also confirmed by molecular markers of *rbcL*, *trnL-F* and *rps16*.

*H. fruticosa* is the type species for *Hedyotis* s.s. The *Diplophragma* section is always treated as an important component of *Hedyotis* s.s. In this study, the species in *Diplophragma* section were found to be clustered with *H. fruticosa* based on the cladograms of *rbcL* and *trnL-F* region in the Figs. 4.10-11. So we propose that the *Diplophragma* section in China belongs to *Hedyotis* s.s, although we do not have all the samples of this section. In order to have a better understanding of this section in China, a more thorough study with morphological support for this genus is needed, especially, comparing their seed character with *H. fruticosa*.

In this study, the result confirmed the analysis of Groeninckx in 2009 and Karehed in 2008 who have found that *Oldenlandia biflora* (synonym of *H. biflora*, as *H. biflora* belongs to *Oldenlandia*) is close to the group of *Kadua*. And Groeninckx suggested

that *O. biflora* should not be treated as a member of *Oldenlandia*, and further examination of its morphological characters is needed.

*H. auricularia* was selected by Linnaeus in 1754 as the type species of *Hedyotis* s.s. But it has indehiscent capsule which conflicts the septicidal dehiscent capsule of *Hedyotis* s.s. So Bremekamp in 1952 transferred *H. auricularia* to a new subgenus *Exallage* and selected it as the type species of this subgenus. *H. costata* and *H. lineata* which was found in China were also included in this subgenus by Bremekamp. But the genus *Exallage* was not widely accepted by other researchers. Based on the cladograms of six single-gene datasets as well as the cladogram of combined dataset, *H. hedyotidea* in section *Dimetia*, which has septicidal capsule with apical protrusion, was grouped with *H. auricularia* and *H. costata*. This matches with the interpretation made by Li (2010). These three species were also closely linked to *Kadua* and *H. biflora* group. Interestingly, the genus of *Kadua* was conventionally treated as one of the subsections in *Hedyotis*. According to the latest analysis of Terrell in 2005, the genus of *Kadua* in Hawaiian was reconstructed mainly based on seed and fruit characters. According to his work, four subgenera of *Kadua* were found. Generally, most species in *Kadua* are subshrub or small trees, but two large succulent perennial herb species are also included in *Kadua* as members of *Wiegmannia* section (Terrell, Robinson et al. 2005). *H. biflora* is an annual small herb while *H. auricularia* and *H. costata* are succulent perennial herbs. *H. hedyotidea* is a perennial climbing subshrub. Moreover, annual herbs *H. diffusa* and *H. herbacea* with the perennial species in section *Hedyotis* (*H. tenelliflora*, *H.*



*pinifolia*, and *H. verticillata*) fall in the clade of *Kadua* as well. As *H. corymbosa* is the lectotype species for genus *Oldenlandia*, and the taxonomic position of it was revealed to be clustered with the clade of Spermaceae which was obviously separated with the *H. diffusa* group. So, moving those species which were found closely related to the clade of *Kadua* out of *Oldenlandia* is proposed.

#### **4.4 Conclusions**

The section delimitation solely relied on the capsule dehiscent type is confusing and unsound. For more precise taxonomic studies, two or more diagnostic characters for section delimitation are required. The second problematic issue is that improper use of the *Hedyotis* name has caused the problem of confusing phylogenetic position of relative species in Spermaceae to some extent. As Verdcourt stated in 1976, if all the genera sunk into *Hedyotis* due to their closely relationship, it will become an awkward group containing morphologically heterogeneous species. However, before given the suggestion for a new generic circumscription, further sampling is needed. Phylogenetic study of *Hedyotis* should rely on both molecular and morphological datasets combined.

## REFERENCES

- Achille, F.; Motley, T.J.; Lowry II, P.P.; Jérémie, J. 2006. Polyphyly in Guettarda L. (Rubiaceae, Guettardeae) based on nrDNA ITS sequence data. *Annals of the Missouri Botanical Garden* 93: 103–121.
- Alvarez, I.; Wendel, J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.
- Andersson, L.; Rova, J.H.E. 1999. The *rps16* intron and the phylogeny of the Rubioideae (Rubiaceae). *Plant Systematics and Evolution* 214: 161–186.
- Andersson, L.; Rova, J.H.E.; Guarín, F.A. 2002. Relationships, circumscription, and biogeography of *Arcytophyllum* (Rubiaceae) based on evidence from cpDNA. *Brittonia* 54: 40–49.
- Andreasen, K.; Baldwin, B.G.; Bremer, B. 1999. Phylogenetic utility of the nuclear rDNA ITS region in subfamily Ixoroideae (Rubiaceae): comparisons with cpDNA *rbcL* sequence data. *Plant Systematics and Evolution* 217: 119–135.
- Andreasen, K.; Bremer, B. 2000. Combined phylogenetic analysis in the Rubiaceae–Ixoroideae: morphology, nuclear and chloroplast DNA data. *American Journal of Botany* 87: 1731–1748.
- Azuma, H.; Thien, L.B.; Kawano, S. 1999. Molecular phylogeny of *Magnolia* (Magnoliaceae) inferred from cpDNA sequences and evolutionary divergence of the floral scents. *Journal of Plant Research* 112: 291–306.
- Bartlett, S.E.; Davidson, W.S. 1992. FINS (Forensically Informative Nucleotide Sequencing): A procedure for identifying the animal origin of biological specimens. *Biotechniques* 12: 408–411.
- Baldwin B.G., Sanderson M.J., Porter J.M., Wojciechowski M.F., Campbell C.S., Donoghue M.J. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247–277.
- Beaumont, A.R. 1994. Genetics and evolution of aquatic organisms, pp. 219–249. Chapman and Hall: London.
- Bellstedt, D.U.; Linder, H.P.; Harley, E.H. 2001. Phylogenetic relationships in *Disa* based on noncoding *trnL–trnF* chloroplast sequences: evidence of numerous repeat regions. *American Journal of Botany* 88: 2088–2100.
- Blanco, M.; Prez-Martín, R.I.; Sotelo, C.G. 2008. Identification of Shark Species in Seafood Products by Forensically Informative Nucleotide Sequencing (FINS). *Journal of Agricultural and Food Chemistry* 56: 9868–9874.
- Bock, J.H.; Slightom, J.L. 1995. Fluorescence-based cycle sequencing with primers selected from a nonamer library. *Biotechniques* 19: 60–2, 64.
- Boore, J.L. 1999. Animal mitochondrial genomes. *Nucleic Acids Research* 27: 1767–1780.
- Bossier, P. 1999. Authentication of Seafood Products by DNA Patterns. *Journal of Food Science* 64: 189–193.



- Bremer, B. 1992. Phylogeny of the Rubiaceae (Chiococceae) Based on Molecular and Morphological Data – Useful Approaches for Classification and Comparative Ecology. *Annals of the Missouri Botanical Garden* 79: 380–387.
- Bremer, B. 1996a. Combined and separate analyses of morphological and molecular data in the plant family Rubiaceae. *Cladistics* 12: 21–40.
- Bremer, B. 1996b. Phylogenetic studies within Rubiaceae and relationships to other families based on molecular data. Second International Rubiaceae Conference, Proceedings 7: 33–50.
- Bremer, B. 2009. A Review of Molecular Phylogenetic Studies of Rubiaceae. *Annals of the Missouri Botanical Garden* 96:4–26.
- Bremer, B., Jansen, R.K.; Oxelman, B.; Backlund, M.; Lantz, H.; Kim, K.J. 1999. More characters or more taxa for a robust phylogeny – case study from the coffee family (Rubiaceae). *Systematic Biology* 48: 413–435.
- Bremer, B.; Andreasen, K.; Olsson, D. 1995. Subfamilial and tribal relationships in the Rubiaceae based on *rbcL* sequence data. *Annals of the Missouri Botanical Garden* 82: 383–397.
- Bremer, B.; Eriksson T. 2009. Time Tree of Rubiaceae: Phylogeny and Dating the Family, Subfamilies, and Tribes. *International Journal of Plant Sciences* 170: 766–793.
- Bremer, B.; Eriksson, O. 1992. Evolution of Fruit Characteristics and Dispersal Modes in the Tropical Family Rubiaceae. *Biological Journal of the Linnean Society* 47:79–95.
- Bremer, B.; Jansen, R.K. 1991. Comparative restriction site mapping of chloroplast DNA implies new phylogenetic–relationships within Rubiaceae. *American Journal of Botany* 78: 198–213.
- Bremer, B.; Manen, J.F. 2000. Phylogeny and classification of the subfamily Rubioideae (Rubiaceae). *Plant Systematics and Evolution* 225: 43–72.
- Bremer, B.; Struwe, L. 1992. Phylogeny of the Rubiaceae and the Loganiaceae – Congruence or conflict between morphological and molecular–data. *American Journal of Botany* 79: 1171–1184.
- Brown, W.M., George, J.M.; Wilson, A.C. 1979. Rapid evolution of mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 76: 1967–1971.
- Bult, C.J.; Sweere, J.A.; Zimmer, E.A. 1995. Cryptic sequence simplicity, nucleotide composition bias, and molecular coevolution in the large subunit of ribosomal DNA in plants: implications for phylogenetic analyses. *Annals of the Missouri Botanical Garden* 82: 235–246.
- But, P.P.H. 2000. Modernization of traditional Chinese medicine needs five finger mountain and golden head ring. In Yuan, L. (ed.) *Drug Discovery and Traditional Chinese Medicine: Science, Regulatory and Globalization*, pp. 137–144. Kluwer Academic Publishers: Netherland.
- CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences USA* 106: 12 794–12 797.
- Chan, K. 1995. Progress in traditional Chinese medicine. *Trends in Pharmacological*



- Sciences* 16: 182–7.
- Chan, W. H.; Ling, K. H.; Chiu, S. W.; Shaw, P. C.; But, P. P. H. 2011. Molecular Analyses of *Cordyceps gunnii* in China. *Journal of Food and Drug Analysis* 19: 18–25.
- Chandler, G.T.; Bayer, R.J.; Crisp, M.D. 2001. A molecular phylogeny of the endemic Australian genus *Gastrolobium* (Fabaceae: Mirbelieae) and allied genera using chloroplast and nuclear markers. *American Journal of Botany* 88: 1675–1687.
- Chapela, M.J.; Sotelo, C.G.; Calo-Mata, P.; Pérez-Martín, R.I.; Rehbein, H.; Hold, G.L.; Quinteiro, J.; Rey-Méndez, M.; Rosa, C.; Santos, A.T. 2002. Identification of Cephalopod Species (Ommastrephidae and Loliginidae) in Seafood Products by Forensically Informative Nucleotide Sequencing (FINS). *Journal of Food Science*. 67: 1672–1676.
- Chen, T. 2007. A new species in the genus *Hedyotis* (Rubiaceae) from south China. *Edinburgh Journal of Botany* 64: 331–334.
- Chen, T. 2008. Notes on the Identity of *Hedyotis Yangchunensis* (Rubiaceae) in Hong Kong. *Harvard Papers in Botany* 13: 283–288.
- Cheng, Q.T.; Zheng, B.S. 1987. *Systematic synopsis of Chinese fishes*, pp: 67–70. Science Press: Beijing.
- Chung, H.S., Jeong, H.J.; Hong, S.H.; Kim, M.S.; Kim, S.J.; Song, B.K.; Jeong, I.S.; Lee, E.J.; Ahn, J.W.; Baek, S.H.; Kim, H.M. 2002. Induction of nitric oxide synthase by *Oldenlandia diffusa* in mouse peritoneal macrophages. *Biological and Pharmaceutical Bulletin* 25: 1142–1146.
- Davis, A.P.; Govaerts, R.; Bridson, D.M.; Ruhsam, M.; Moat, J.; Brummitt, N.A. 2009. A Global Assessment of Distribution, Diversity, Endemism, and Taxonomic Effort in the Rubiaceae. *Annals of the Missouri Botanical Garden* 96: 68–78.
- Dent, D.R. 1993. The use of *Bacillus thuringiensis* as an insecticide. In Jones, D.G. (ed.) *Exploitation of Microorganisms*, pp. 19–37. Chapman and Hall: London.
- Downie, S.R.; Palmer, J.D. 1992. Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In Soltis, P.E.; Doyle, D.E.; Doyle, J.J. (eds.) *Molecular systematics of Plants*, pp. 14–35. Chapman and Hall: London.
- Doyle, J.J.; Doyle, J.L.; Palmer, J.D. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. *Systematic Botany* 20: 272–294.
- Fang, P.W. 1934. Study on the fishes referring to Salangidea of China. *Sinensia* 4: 231–268.
- Farris, J.S.; Albert, V.A.; Källersjö, M.; Lipscomb, D.; Kluge, A.G. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics*. 12: 99–124.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 39: 783–791.
- Feng, T.; Liu, S.; He, X.J. 2010. Molecular authentication of the traditional Chinese medicinal plant *Angelica sinensis* based on internal transcribed spacer of nrDNA. *Electronic Journal of Biotechnology*. 13: 1-10.



- Forrest, A.R.R.; Carnegie, P.R. 1994. Identification of gourmet meat using FINS (Forensically Informative Nucleotide Sequencing). *Biotechniques* 17: 24–26.
- Forsberg, F. R., and E. E. Terrell. 1985. A recently established exotic in west Florida and Alabama (*Hedyotis salzmannii* or *Oldenlandia salzmannii*; Rubiaceae) *Castanea*: 50: 49-51.
- Fu, C.Z.; Luo, J.; Wu, J.H.; Lopez, J.A.; Zhong, Y.; Lei, G.C.; Chen, J.K. 2005. Phylogenetic relationships of salangid fishes (Osmeridae, Salanginae) with comments on phylogenetic placement of the salangids based on mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 35: 76–84.
- Govaerts, R.; Ruhsam, M.; Andersson, L.; Robbrecht, E.; Bridson, D.M.; Davis, A.P.; Schanzer, L.; Sonke B. 2006. World Checklist of Rubiaceae. <http://www.kew.org/wesp/rubiaceae>. (accessed on 28 July 2009).
- Greuter, W.; McNeill, J.; Barrie, F.R.; Burdet, H.M.; Demoulin, V.; Filgueiras, T.S.; Nicolson, D.H.; Silva, P.C.; Skog, J.E.; Trehane, P.; Turland, N.J.; Hawksworth, D.L. 2000. International Code of Botanical Nomenclature (St Louis Code). *Regnum Vegetabile* 138: 1–474. Koeltz Scientific Books: Konigstein.
- Groeninckx, I.; Dessein, S.; Ochoterena, H.; Person, C.; Motley, T.J.; Karehed, J.; Bremer, B.; Huysmans, S.; Smets, E. 2009. Phylogeny of the Herbaceous Tribe Spermaceae (Rubiaceae) Based on Plastid DNA Data. *Annals of the Missouri Botanical Garden* 96: 109–132.
- Guha, S.; Kashyap, V. K. 2006. Molecular identification of lizard by RAPD & FINS of mitochondrial 16s rRNA Gene. *Legal Medicine*. 8:5–10.
- Halford, D.A. 1992. Review of the genus *Oldenlandia* L. (Rubiaceae) and related genera in Australia. *Austrobaileya* 3: 683–722.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 41: 95-98.
- Hamer, G. 1996. Thermophilic aerobic processes for waste sewage sludge treatment. In Moo–Young, M.; Anderson, W.A.; Chakrabarty, A.M. (eds.) *Environmental Biotechnology: Principles and Applications*, pp. 206–220. Kluwer Academic Publishers: Dordrecht.
- Hamilton, M.B. 1999. Tropical tree gene flow and seed dispersal: deforestation affects the genetic structure of the surviving forest fragments. *Nature* 401: 129–130.
- Hamilton, M.B.; Braverman, J.M.; Soria–hernanz, D.F. 2003. Patterns and relative rates of nucleotide and insertion/deletion evolution at six chloroplast intergenic regions in New World species of the Lecythidaceae. *Molecular Biology and Evolution* 20: 1710–1721.
- Hao, M.G.; Liu, Z.Q.; Wang, J.L. 2004. Application of the sequences of rDNA ITS to identify Chinese crude drug *Hedyotis diffusa*. *Journal of Anhui Normal University* 27: 188–191.
- Hesketh, T.; Zhu, W. X. 1997. Health in China – Traditional Chinese medicine: One country, two systems. *British Medical Journal* 315: 115–117.
- Hilu, K.W., Borsch, T.; Müller, K.; Soltis, D.E.; Soltis, P.S.; Savolainen, V.; Chase, M.



- W.; Powell, M.P.; Alice, L.A.; Evans, R.; Sauquet, H.; Neinhuis, C.; Slotta, T.A.B.; Rohwer, J.G.; Campbell, C.S.; Chatrou, L.W. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758–1776.
- Hooker, J.D. 1880. Rubiaceae, *Hedyotis* and *Oldenlandia*. *Flora of British India* 3: 49–70.
- Janssen, T.; Bremer, K. 2004. The age of major monocot groups inferred from 800+ *rbcl* sequences. *Botanical Journal of the Linnean Society* 146: 385–398.
- Joshi, K.; Chavan, P.; Warude, D.; Patwardhan, B. 2004. Molecular markers in herbal drug technology. *Current Science* 87: 159–165.
- Karehed, J.; Bremer B. 2007. The systematics of Knoxieae (Rubiaceae) molecular data and their taxonomic consequences. *Taxon* 56: 1051–1076.
- Karehed, J.; Groeninckx, I.; Dessein, S.; Motley, T.J.; Bremer, B. 2008. The phylogenetic utility of chloroplast and nuclear DNA markers and the phylogeny of the Rubiaceae tribe Spermacoceae. *Molecular Phylogenetics and Evolution* 49: 843–66.
- Komatsu, K.; Zhu, S.; Fushimi, H.; Qui, T.K.; Cai, S.Q.; Kadota, S. 2001. Phylogenetic analysis based on 18S rRNA gene and *matK* gene sequences of *Panax vietnamensis* and five related species. *Planta Medica*. 67: 461–465.
- Kawabata, A.; Kubota, S.; Ishito, Y.; Fujita, T.; Kitagawa, D.; Iizuka, K. 1994. Growth, maturity and standing stock size of the Ishikawa Icefish *Salangichthys ishikawae* along the Pacific Coast of Aomori Prefecture, Japan. *Bulletin of the Tohoku National Fisheries Research Institute* 56: 105–114.
- Kim, D.H.; Lee, H.J.; Oh, Y.J.; Kim, M.J.; Kim, S.H.; Jeong, T.S.; Baek, N.I. 2005. Iridoid glycosides isolated from *Oldenlandia diffusa* inhibit LDL-oxidation. *Archives of Pharmacal Research* 28: 1156–1160.
- Kim, I.S.; Park, J.Y. 2002. Freshwater fishes of Korea. Kyo-Hak Publishing: Seoul.
- Kim, Y., Park, E.J.; Kim, J.; Kim, Y.B.; Kim, S.R.; Kim Y.C. 2001. Neuroprotective constituents from *Hedyotis diffusa*. *Journal of Natural Products* 64: 75–78.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*. 16: 111–120.
- Ko, W.C. 1999. *Hedyotis*, In Luo, X.R. (ed.) *Flora Republicae Popularis Sinicae*, vol: 71, pp: 32–77. Science Press: Beijing.
- Kocher, T.D.; Thomas, W.K.; Meyer, A.; Edwards, S. V.; Paabo, S.; Villablanca, F.X.; Wilson, A.C. 1989. Dynamics of mitochondrial DNA evolution in animals: 18 amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America* 86: 6196–6200.
- Kottelat, M. 2001. Freshwater fishes of northern Vietnam. Washington, DC: Environment Social Development Unit, East Asia and Pacific Region.
- Kress, W.J.; Erickson D.L. 2007. A two-locus global DNA barcode for land plants: the coding *rbcl* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One* 2: e508.
- Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., Janzen, D.H. 2005. Use of



- DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences USA* 102: 8369-8374.
- Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wil, A.; Lopez, R.; Thompson, J.D.; Gibson, T.J.; Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23: 2947–2948.
- Lau, D.T.W.; Shaw, P.C.; Wang, J.; But, P.P.H. 2001. Authentication of medicinal *Dendrobium* species by the internal transcribed spacer of ribosomal DNA. *Planta Medica* 67: 456–460.
- Lee, J.; Hymowitz, T. 2001. A molecular phylogenetic study of the subtribe Glycininae (Leguminosae) derived from the chloroplast DNA *rps16* intron sequences." *American Journal of Botany* 88: 2064–2073.
- Lens, F.; Groeninckx, I.; Smets, E.I. Dessein, S. 2009. Woodiness within the Spermaceae–Knoxieae alliance (Rubiaceae): retention of the basal woody condition in Rubiaceae or recent innovation? *Annals of Botany* 103: 1049–64.
- Lewis, W. H. 1961. Merger of the North American *Houstonia* and *Oldenlandia* under *Hedyotis*. *Rhodora*. 63: 216-223.
- Lewis, W.H. 1962. Phylogenetic study of *Hedyotis* (Rubiaceae) in North American. *American January of Botany* 49: 855–865.
- Li, M., Jiang, R.W.; Hon, P.M.; Cheng, L.; Lie, L.L.; Zhou, J.R.; Shaw, P.C.; But, P.P.H. 2010. Authentication of the anti-tumor herb Baihuasheshecao with bioactive marker compounds and molecular sequences. *Food Chemistry* 119: 1239–1245.
- Li, M., Cao, H., But, P.P.H., Shaw, P.C. 2011. Identification of herbal medicinal materials using DNA barcodes. *Journal of Systematics and Evolution*. In press. Doi: 10.1111/j.1759-6831.2011.00132.x
- Li, S.W. 2003. Silver in water. *Agriculture in Human Province* 6:18–19.
- Li, X.; Ju, W.J.; Chi, Y.Y. 1996. Authentication of baihuasheshecao and its adulterants. *China journal of Chinese materia medica* 21: 460.
- Liang, Z.T.; Jiang, Z. H.; Leung, K.S.Y.; Zhao, Z.Z. 2006. Determination of iridoid glucosides for quality assessment of herb *oldenlandia* by high-performance liquid chromatography. *Chemical & Pharmaceutical Bulletin (Tokyo)* 54: 1131–1137.
- Liden, M.; Fukuhara, T.; Rylander, J.; Oxelman, B. 1997. Phylogeny and classification of Fumariaceae, with emphasis on *Dicentra* s.l. based on the plastid gene *rps16* intron. *Plant Systematics and Evolution* 206: 411–420.
- Lin, C.C.; Ng, L.T.; Yang, J.J.; Hsu, Y.F. 2002. Anti-inflammatory and hepatoprotective activity of peh-hue-juwa-chi-cao in male rats. *The American Journal of Chinese Medicine* 30: 225–234.
- Liston, A.; Robinson, W.A.; Oliphant, J.M. 1996. Length variation in the nuclear ribosomal DNA Internal transcribed spacer region of non-flowering seed plants. *Systematic Botany* 21: 109–120.
- Liu, Z.Q.; Ding, X.Y. Ding, B.Z. 2006. Application of species-specific PCR in the identification of *Hedyotis diffusa* Willd. *Journal of Nanjing Normal University*



- 29: 91–94.
- Liu, Z.Q.; Hao, M.G. 2005. The authentication of rDNA ITS region for Baihuasheshecao. *Shaanxi Journal of Traditional Chinese Medicine* 26: 167–169.
- Liu, Z.Q.; Hao, M.G.; Wang, J.L. 2004. Application of allele-specific primer in the identification of *Hedyotis diffusa*. *Journal of Chinese medicinal materials* 27: 484–487.
- Lu, C.M.; Yang, J.J.; Wang, P.Y.; Lin, C.C. 2000. A new acylated flavonol glycoside and antioxidant effects of *Hedyotis diffusa*. *Planta Medica* 66: 374–377.
- Manen, J.F., Natali, A.; Manen, J.F.; Natali, A.; Ehrendorfer, F. 1994. Phylogeny of Rubiaceae–Rubiaceae inferred from the sequence of a cpDNA intergene region. *Plant Systematics and Evolution* 190: 195–211.
- Manhart, J.R. 1994. Phylogenetic analysis of green plant *rbcL* sequences. *Molecular Phylogenetics and Evolution* 3: 114–27.
- McBride, L. J., Koepf, S.M.; Gibbs, R.A.; Salser, W.; Mayrand, P.E.; Hunkapiller, M.W.; Kronick, M.N. 1989. Automated DNA sequencing methods involving polymerase chain reaction. *Clinical Chemistry* 35: 2196–201.
- Mindell, D.P.; Honeycutt R.L. 1999. Ribosomal RNA in vertebrates: evolution and phylogenetic applications. *Annual Review of Ecology and Systematics* 21: 541–566
- Miya, M.; Nishida, T. 1996. Molecular phylogenetic perspective on the evolution of the deep-sea fish genus *Cyclothone* (Stomiiformes: Gonostomatidae). *Japanese Journal of Ichthyology* 43: 375–398.
- Motley, T.J.; Wurdack, K.J.; Delprete, P.G. 2005. Molecular systematics of the Catesbaeeae–Chiococceae complex (Rubiaceae): flower and fruit evolution and biogeographic implications. *American Journal of Botany* 92: 316–329.
- Mouly, A., Razafimandimbison, S.G.; Khodabandeh, A.; Bremer, B. 2009. Phylogeny and classification of the species-rich pantropical showy genus *Ixora* (Rubiaceae–Ixoreae) with indications of geographical monophyletic units and hybrids. *American Journal of Botany* 96: 686–706.
- Müller, K. 2005. SeqState – primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics*. 4: 65–69.
- Müller, K. 2006. Incorporating information from length–mutational events into phylogenetic analysis. *Molecular Phylogenetics and Evolution*. 38: 667–676.
- Mullis, K.; Faloona, F.; Scharf, S.; Saiki, R.; Horn, G.; Erlich, H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 51: 263–73.
- Mullis, K.B.; Faloona, F.A. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155: 335–50.
- Natali, A.; Manen, J.F.; Ehrendorfer, F. 1995. Phylogeny of the Rubiaceae Rubioideae, in Particular the Tribe Rubieae – Evidence from a Noncoding Chloroplast DNA–Sequence. *Annals of the Missouri Botanical Garden* 82: 428–439.
- Nelson, J. S. 1994. Fishes of the World 3<sup>rd</sup>, pp. 194–195. John Wiley & Sons Inc.:



- New York.
- Nei and Kumar 2000
- Nei, M.; Kumar, S. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press: New York.
- Neupane, S.; Dessin, S.; Motley, T.J. 2009. The *Heydotis* – *Oldenlandia* – *Kohautia* complex (Rubiaceae) in Nepal: a study of fruit, seed and pollen characters and their taxonomic significance. *Edinburgh Journal of Botany* 66: 371–390.
- Nickrent, D.L.; Soltis, D.E. 1995. A Comparison of Angiosperm Phylogenies from Nuclear 18S rDNA and rbcL Sequences. *Annals of the Missouri Botanical Garden* 82: 208–234.
- Olmstead, M.C.; Franklin, K.B. 1994. Lesions of the pedunculopontine tegmental nucleus abolish catalepsy and locomotor depression induced by morphine. *Brain Research* 662: 134–40.
- Oxelman, B.; Liden, M.; Berglund, D. 1997. Chloroplast *rps16* intron phylogeny of the tribe Sileneae (caryophyllaceae). *Plant Systematics and Evolution* 206: 393–410.
- Palmer, J.D. 1985a. Evolution of chloroplast and mitochondrial DNA in plants and algae. In MacIntyre R.J. (ed.) *Molecular evolutionary Genetics*, pp. 131–240.
- Palmer, J.D. 1985b. Comparative organization of chloroplast genomes. *Annual review of Genetics* 19: 325–354.
- Palmer, J.D. 1986. Isolation and structural analysis of chloroplast DNA. *Methods in Enzymology* 188: 167–186.
- Palmer, J.D. 1991. Plastid chromosomes: structure and evolution. In Bogorad, L.; Vasil, I.K. (eds.) *Cell culture and somatic cell genetic in plants, The molecular biology of plastid*, pp. 5–53. Academic Press: New York.
- Perler, F.; Efstratiadis, A.; Lomedico, P.; Gilbert, W.; Kolodner, R.; Dodgson, J. 1980. The evolution of genes: the chicken preproinsulin gene. *Cell* 20: 555–556.
- Robbrecht, E. 1988. Tropical woody Rubiaceae. *Opera botanica Belgica* 1: 1–271.
- Robbrecht, E. 1994a. On the delimitation of the Rubiaceae. A review. *Opera botanica Belgica* 6: 19–30.
- Robbrecht, E. 1994b. Supplement to the 1988 outline of the classification of the Rubiaceae. Index to genera. *Opera botanica Belgica* 6: 173–196.
- Robbrecht, E.; Manen, J.F. 2006. The major evolutionary lineages of the coffee family (Rubiaceae, angiosperms). Combined analysis (nDNA and cpDNA) to infer the position of *Coptosapelta* and *Luculia*, and supertree construction based on rbcL, rps16, trnL–trnF and atpB–rbcL data. A new classification in two subfamilies, Cinchonoideae and Rubioideae. *Systematics and Geography of Plants* 76: 85–145.
- Roberts, T.R. 1984. Skeletal anatomy and classification of the neotenic Asian salmoniform superfamily Salangoidea (icefishes or noodlefishes). *Proceedings of the California Academy of Science* 43: 179–220.
- Rogers, S.O.; Bendich, A.J. 1987. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Molecular Biology* 9: 509–520.
- Rova, J. H. E., Delprete, P.G.; Andersson, L.; Albert, V.A. 2002. A trnL–F cpDNA



- sequence study of the Condamineae–Rondeletiae–Sipaneeae complex with implications on the phylogeny of the Rubiaceae. *American Journal of Botany* 89: 145–159.
- Rydin, C., Razafimandimbison, S.G.; Khodabandeh, A.; Bremer, B. 2009. Evolutionary relationships in the Spermaceae alliance (Rubiaceae) using information from six molecular loci: insights into systematic affinities of *Neohymenopogon* and *Mouretia*. *Taxon* 58: 793–810.
- Rydin, C.; Kainulainen, K.; Razafimandimbison, S.G.; Smedmark, J.E.E.; Bremer, B. 2009. Deep divergences in the coffee family and the systematic position of *Acranthera*. *Plant Systematics and Evolution* 278: 101–123.
- Rydin, C.; Razafimandimbison, S.G.; Bremer, B. 2008. Rare and enigmatic genera (*Dunnia*, *Schizocolea*, *Colletocema*), sisters to species-rich clades: Phylogeny and aspects of conservation biology in the coffee family. *Molecular Biology and Evolution* 48: 74–83.
- Sahajpal, V.; Goyal, S.P. 2010. Identification of a forensic case using microscopy and forensically informative nucleotide sequencing (FINS): A case study of small Indian civet (*Viverricula indica*). *The journal of forensic science of society* 50: 94–97
- Sang, T., Crawford, D.J.; Stuessy, T.F. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120–1136.
- Santaclara, F.J.; Espiñeira, M.; Vieites, J.M. 2007. Genetic Identification of Squids (Families Ommastrephidae and Loliginidae) by PCR–RFLP and FINS Methodologies. *Journal of Agricultural and Food Chemistry* 55: 9913–9920.
- Scheid, V. 1999. The globalisation of Chinese medicine. *The Lancet*. Suppl: SIV10.: 354.
- Schumann, K. 1891. Rubiaceae. In Engler A.; Prantl K. (eds.) *Die natürlichen Pflanzenfamilien*, pp. 1–156. Engelmann: Leipzig.
- Shan, B.E.; Zhang, J.Y.; Du, X.N. 2001. Immunomodulatory activity and anti-tumor activity of *Oldenlandia diffusa* in vitro. *The medicine magazine of China*. 21: 370–4.
- Shaw, J., Lickey, E.B.; Beck, J.T.; Farmer, S.B.; Liu, W.S.; Miller, J.; Siripun, K.C.; Winder, C.T.; Schilling, E.E.; Small, R.L. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142–166.
- Shaw, P.C., Ngan, F.N.; But, P.P.H.; Wang, J. 2002. Molecular markers in Chinese medicinal materials. In: Shaw, P.C., Wang, J.; But, P.P.H. (eds.). *Authentication of Chinese Medicinal Materials by DNA Technology*, pp. 1–23. World Scientific Publishing Co. Pte. Ltd.: Singapore.
- Sokolovskaya, T.G.; Sokolovsky, A.S.; Sobolevsky, E.I. 1998. A list of fishes of the Peter the Great Bay (the Sea of Japan). *Journal of Ichthyology* 38: 5–15.
- Soltis, D.E.; Soltis, P.S. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. In: Soltis, D.E.; Soltis, P.S.; Doyle, J.J. (eds.). *Molecular Systematics of Plants II: DNA Sequencing*. pp. 1–42. Kluwer Academic



Publishers: Massachusetts.

- Stech, M.; Quandt, D.; Frey, W. 2003. Molecular circumscription of the hornworts (Anthocerotophyta) based on the chloroplast DNA *trnL-trnF* region. *Journal of Plant Research* 116: 389–398.
- Stöver, B.C.; Müller, K.F. 2010. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*. 11: 7. DOI: 10.1186/1471-2105-11-7.
- Stuessy, T.F. 2009. Plant taxonomy: the systematic evaluation of comparative data. pp: 273–294. Columbia University Press: New York.
- Swerdlow, H.; Gesteland, R. 1990. Capillary gel electrophoresis for rapid, high resolution DNA sequencing. *Nucleic Acids Res* 18: 1415–9.
- Swofford, D. L. 2001. *PAUP\*: Phylogenetic analysis using parsimony (\*and other methods)*. Sinauer Associates, Inc: Sunderland.
- Taberlet, P.; Gielly, L.; Patou, G.; Bouvet, J. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.
- Takaiwa, F.; Oono, K.; Sugiura, M. 1985. Nucleotide sequence of the 17S–25S spacer region from rice rDNA. *Plant Molecular Biology* 4: 355–364.
- Takita, T. 1995. Fishes of ariake sound. *Marine Science* 12: 105–115.
- Takita, T.; Kawaguchi, K.; Masutani, H. 1988. Distribution and morphology of the salangidish, *Neosalanx reganius*. *Japanese Journal of Ichthyology* 34: 497–503.
- Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*. 28: 2731-2739.
- Teletchea, F.; Maudeta, C.; Hännia, C. 2005. Food and forensic molecular identification: update and challenges. *Trends in Biotechnology*. 23: 359–366
- Terrell, E.E.; Robinson, H.E.; Wagner, W.L.; Lorence, D.H. 2005. Resurrection of genus *Kadua* for Hawaiian Hedyotidinae (Rubiaceae), with emphasis on seed and fruit characters and notes on South Pacific species. *Systematic Botany* 30: 818–833.
- Tracy, T.E.; Mulcahy, L.S. 1991. A simple method for direct automated sequencing of PCR fragments. *Biotechniques* 11: 68–75.
- Verdcourt, B. 1958. Remarks on the Classification of the Rubiaceae. *Bulletin du Jardin botanique de l'État a Bruxelles*. 28: 209-290
- Wang, R.J. 2007. *Hedyotis koana* R. J. Wang, a new species of Rubiaceae from China. *Acta Phytotaxonomica Sinica* 45: 696–700.
- Wang, R.J. 2008. Two new species of *Hedyotis* (Rubiaceae: Hedyotideae) from Hainan, China. *Novon: A Journal for Botanical Nomenclature* 18: 264–268.
- Wang, R.J.; Xing, F.W. 2003. Two new *Hedyotis* species in China. *Acta Phytotaxonomica Sinica*. 41: 85–88.
- Wang, Y.L., Zhang, Y.; Fang, M.; Li, Q.J.; Jiang, Q.; Ming, L. 2005. Anti-inflammatory and antibacterial effects of total flavones of *Oldenlandia diffusa* willd. *Chinese Pharmacological Bulletin* 21: 348–350.



- Wen, J.; Hua, C.Q.; Zhang, L.P.; Luo, P.; Zhao, Z.; Fan, S.G.; Su, T. 2010. The application of PCR-RFLP and FINS for species identification used in sea cucumbers (Aspidochirotida: Stichopodidae) products from the market. *Food Control* 21: 403-407.
- Westhof, E.; Michel, F. 1996. The structural domains of group I introns. In R. Green and R. Schroeder (eds.) *Ribosomal RNA and group I introns*, pp: 1-10. Chapman and Hall: New York.
- White, T.J.; Burns, T.; Lee, S.; Taylor, J. 1990. In Innis, M.; Gelfand, D.; Sninsky, J.; White, T. (eds.) *PCR Protocols: A Guide to Methods and Applications*, pp: 315-322. Academic: San Diego.
- Wight, R.; Arnott, G.A.W. 1834. *Hedyotis. Prodromus Florae Peninsulae Indiae Orientalis I*, pp. 405-418. Parbury, Allen, & Co.: London.
- Wolfe, K.H. 1991. Protein-coding genes in chloroplast DNA: compilation of nucleotide sequences, data base entries and rates of molecular evolution. In Bogorad, L.; Vasil, I.K. (eds.) *Cell culture and somatic cell genetic in plants, Vol.7, The molecular biology of plastid*, pp: 467-482. Academic Press: San Diego.
- Xie, Y.H.; Xie, H. 1997. Classification, distribution, and population ecology of Salangidae fishes. *Chinese Journal of Fisheries* 10: 11-19.
- Xie, Z.; Zhang, Y.; Lu, R. 1997. Identification of herb *Hedyotis diffusa* and its confused material herb *Hedyotis pinifolia*. *Journal of Chinese Medicinal Materials* 20: 287-290.
- Xiong, C.X. 1994. The analysis of nutrition proportion of *Neosalanx tangkahkeii* and *Neosalanx oligodonties*. *Freshwater fisheries* 24: 7-9.
- Xu, L.S.; Xu, G.J.; Wang, Z.T. 1997. *Species systematization and quality evaluation of commonly used Chinese traditional drugs. South-China Edition*, pp: 658. Science and Technology Press: Fujian.
- Yin, H.; Hai, S. 1989. The morphological and histological authentication of baihuasheshecao and its adulterants. *China journals of Chinese materia medica* 14: 9.
- Yoshida, Y.; Wang, M.Q.; Liu, J.N.; Shan, B.E.; Yamashita, U. 1997. Immunomodulating activity of Chinese medicinal herbs and *Oldenlandia diffusa* in particular. *International Journal of Immunopharmacology* 19: 359-370.
- Zhang, J., Li, M.; Xu, M.Q.; Takita, T.; Wei, F.W. 2007. Molecular phylogeny of icefish Salangidae based on complete mtDNA cytochrome b sequences, with comments on estuarine fish evolution. *Biological Journal of the Linnean Society* 91: 325-340.
- Zhang, Y.B.; Shaw, P.C.; Sze, C.W.; Wang, Z.T.; Tong, Y. 2007. Molecular Authentication of Chinese Herbal Materials. *Journal of Food and Drug Analysis* 15:1-9.
- Zhang, Y.L. 1993. The phylogenetic biogeography of Salangidae. In G.X. Zhang (ed.) *Scientific Treatise on Systematic and Evolutionary zoology* 2<sup>ed</sup>, pp: 65-77. Science and Technology Press of China: Beijing.
- Zhang, Y.L.; Qiao, X.G. 1994. Study on phylogeny and zoogeography of fishes of the



- family Salangidae. *Acta Zool Taiwanica* 5: 95–115.
- Zheng, C.C. 1990. The authentication study of baihuasheshecao and its adulterants in Hainan Province. *Hainan Medical Journal* 1: 57.
- Zhu YJ, Chen SL, Yao H, Tan R, Song JY, Luo K, Lu J. 2010. DNA barcoding the medicinal plants of the genus Paris. *Acta Pharmaceutica Sinica* 45: 376–382
- Zurawski, G.; Perrot, B.; Bottomley, W.; Whitfeld, P.R. 1981. The structure of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from spinach chloroplast DNA. *Nucleic Acids Research* 9: 3251–3270.
- Zurawski, G.; Whitfeld, P.R.; Bottomley, W. 1986. Sequence of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from pea chloroplasts. *Nucleic Acids Research* 14: 3975.

## APPENDIX





















































































































Table 6-9 Sequences alignment matrix of Hedyotis based on *trnH-psbA* intergenic spacer region.

	5	15	25	35	45	55	65	75	85	95	105	115
B907P_H_as	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
B908P_H_as	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
HA01P_H_ac	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
HA02P_H_ac	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
Y103P_H_un	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
Y104P_H_un	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM04P_H_un	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM17P_H_un	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM18P_H_un	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM19P_H_un	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
Y109P_H_bo	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
Y110P_H_bo	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM22P_H_bo	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM11P_H_sh	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM20P_H_sh	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM10P_H_sh	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM19P_H_sh	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
B906P_H_lo	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
MA01P_H_br	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
MA02P_H_br	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM01P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
Y113P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM12P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
B921P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
BX03P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM14P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM13P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM15P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
Y115P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
MA05P_H_co	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
MA09P_H_te	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
MA10P_H_te	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC
TM05P_H_te	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC	CGAATTC





















































































































































































































YU14\_H\_cor  
 M424\_H\_cor  
 AF072038\_P

YU1	YU2	YU3	YU4	YU5	YU6	YU7	YU8	YU9	YU10	YU11	YU12	YU13	YU14	YU15	YU16	YU17	YU18	YU19	YU20	YU21	YU22	YU23	YU24	YU25	YU26	YU27	YU28	YU29	YU30	YU31	YU32	YU33	YU34	YU35	YU36	YU37	YU38	YU39	YU40	YU41	YU42	YU43	YU44	YU45	YU46	YU47	YU48	YU49	YU50	YU51	YU52	YU53	YU54	YU55	YU56	YU57	YU58	YU59	YU60	YU61	YU62	YU63	YU64	YU65	YU66	YU67	YU68	YU69	YU70	YU71	YU72	YU73	YU74	YU75	YU76	YU77	YU78	YU79	YU80	YU81	YU82	YU83	YU84	YU85	YU86	YU87	YU88	YU89	YU90	YU91	YU92	YU93	YU94	YU95	YU96	YU97	YU98	YU99	YU100
YU10_H_bod	YU11_H_bod	YU12_H_bod	YU16_H_log	YU17_H_log	YU09_H_bod	HA01_H_bac	HA02_H_bac	BP21_H_bac	YU21_H_vac	YU22_H_vac	YU23_H_vac	YU24_H_vac	TM09_H_shi	TM19_H_shi	TM1_H_shi	TM20_H_shi	TM10_H_shi	MA01_H_bac	MA02_H_bac	SA03_H_cor	YU13_H_cor	TM14_H_cor	TM15_H_cor	TM01_H_cor	TM12_H_cor	TM13_H_cor	YU15_H_cor	BP20_H_cor	BP06_H_cor	BP07_H_bac	BP08_H_bac	MA05_H_cor	YU03_H_unc	YU04_H_unc	TM04_H_unc	TM17_H_unc	TM18_H_unc	TM16_H_unc	BP17_H_bif																																																												

TM18_H_unc	TM16_H_unc	BP17_H_bif	MA17_H_bif	BP16_H_bif	MA19_H_bif	MA20_H_bif	MA22_H_bif	MA21_H_bif	BP18_H_bif	TY01_H_bif	YU05_H_bac	YU06_H_bac	YU07_H_bac	MA09_H_bac	MA10_H_bac	TM05_H_bac	TM06_H_bac	YU12_H_bac	MA14_H_bac	MA15_H_bac	MA11_H_bac	MA12_H_bac	MA13_H_bac	YU19_H_bac	SL02_H_bac	SL03_H_bac	MA03_H_bac	TM21_H_bac	MA04_H_bac	BP23_H_bac	MA17_H_bac	MA18_H_bac	YU01_H_bac	YU02_H_bac	MA07_H_bif	MA08_H_bif	MA06_H_bif	YU20_H_bif	BP22_H_cor	MA23_H_cor	MA25_H_cor	TA01_H_cor	MA26_H_cor	TS01_H_cor
------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------	------------

















































































































































BW16S_H_d1	-ITMUA	EMTMUR	TMERTUR	MTORU-
M407S_H_d1	-ITMUA	EMTMUR	TMERTUR	MTORU-
YU07S_H_be	-ITMUA	EMTMUR	TMERTUR	MTORU-
YU06S_H_be	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
YU01S_H_d1	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
M411S_H_p1	-ITMUA	EMTMUR	TMERTUR	MTORU-
M412S_H_p1	-ITMUA	EMTMUR	TMERTUR	MTORU-
M413S_H_p1	-ITMUA	EMTMUR	TMERTUR	MTORU-
YU19S_H_p1	-ITMUA	EMTMUR	TMERTUR	MTORU-
M414S_H_ve	-ITMUA	EMTMUR	TMERTUR	MTORU-
M415S_H_ve	-ITMUA	EMTMUR	TMERTUR	MTORU-
M416S_H_ve	-ITMUA	EMTMUR	TMERTUR	MTORU-
YU12S_H_ve	-ITMUA	EMTMUR	TMERTUR	MTORU-
M410S_H_te	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
TM05S_H_te	-ITMUA	EMTMUR	TMERTUR	MTORU-
M409S_H_te	-ITMUA	EMTMUR	TMERTUR	MTORU-
TM06S_H_te	-ITMUA	EMTMUR	TMERTUR	MTORU-
Hedyotis_h	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_axil	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_affi	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_cori	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_repe	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_cent	-ITMUA	EMTMUR	TMERTUR	MTORU-
Hedyotis_d	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_litt	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_fosb	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_flyw	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_laai	-ITMUA	EMTMUR	TMERTUR	MTORU-
Hedyotis_s	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_flya	-ITMUA	EMTMUR	TMERTUR	MTORU-
Kadua_ejat	-ITMUA	EMTMUR	TMERTUR	MTORU-
Hedyotis_p	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
M406S_H_b1	-ITMUA	EMTMUR	TMERTUR	MTORU-
M407S_H_b1	-ITMUA	EMTMUR	TMERTUR	MTORU-
M408S_H_b1	-ITMUA	EMTMUR	TMERTUR	MTORU-
YU20S_H_b1	-ITMUA	EMTMUR	TMERTUR	MTORU-
Symplocaria	-ITMUA	EMTMUR	TMERTUR	MTORU-
KL00S_H_be	-ITMUA	EMTMUR	TMERTUR	MTORU-
KL00S_H_be	-ITMUA	EMTMUR	TMERTUR	MTORU-
TM21S_H_be	-ITMUA	EMTMUR	TMERTUR	MTORU-
BK23S_H_be	-ITMUA	EMTMUR	TMERTUR	MTORU-
M403S_H_be	-ITMUA	EMTMUR	TMERTUR	MTORU-
M404S_H_be	-ITMUA	EMTMUR	TMERTUR	MTORU-
M417S_H_a1	-ITMUA	EMTMUR	TMERTUR	MTORU-
M418S_H_a1	-ITMUA	EMTMUR	TMERTUR	MTORU-
YU01S_H_a1	-ITMUA	EMTMUR	TMERTUR	MTORU-
YU02S_H_a1	-ITMUA	EMTMUR	TMERTUR	MTORU-
Bocconia_h	-ITMUA	EMTMUR	TMERTUR	MTORU-
Mitracarpu	-ITMUA	EMTMUR	TMERTUR	MTORU-
Mitracarpu	-ITMUA	EMTMUR	TMERTUR	MTORU-
Diodia_sar	-ITMUA	EMTMUR	TMERTUR	MTORU-
Spermacoce	-ITMUA	EMTMUR	TMERTUR	MTORU-
Eriosema_l1	-ITMUA	EMTMUR	TMERTUR	MTORU-
Spermacoce	-ITMUA	EMTMUR	TMERTUR	MTORU-
Diodia_pul	-ITMUA	EMTMUR	TMERTUR	MTORU-
Spermacoce	-ITMUA	EMTMUR	TMERTUR	MTORU-
Richardia_	-ITMUA	EMTMUR	TMERTUR	MTORU-
Richardia_	-ITMUA	EMTMUR	TMERTUR	MTORU-
Spermacoce	-ITMUA	EMTMUR	TMERTUR	MTORU-
Spermacoce	-ITMUA	EMTMUR	TMERTUR	MTORU-
Crosea_his	-ITMUA	EMTMUR	TMERTUR	MTORU-
Crosea_neg	-ITMUA	EMTMUR	TMERTUR	MTORU-
Gallanth_	-ITMUA	EMTMUR	TMERTUR	MTORU-
Gallanth_	-ITMUA	EMTMUR	TMERTUR	MTORU-
Diodia_spi	-ITMUA	EMTMUR	TMERTUR	MTORU-
Eriosema	-ITMUA	EMTMUR	TMERTUR	MTORU-
Nesohedyot	-ITMUA	EMTMUR	TMERTUR	MTORU-
Diodia_vir	-ITMUA	EMTMUR	TMERTUR	MTORU-
Bouvardia_	-ITMUA	EMTMUR	TMERTUR	MTORU-
Bouvardia_	-ITMUA	EMTMUR	TMERTUR	MTORU-
Hedyotis_s	-ITMUA	EMTMUR	TMERTUR	MTORU-
Acrotyphi	-ITMUA	EMTMUR	TMERTUR	MTORU-
Manettia_a	-ITMUA	EMTMUR	TMERTUR	MTORU-
Manettia_a	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
Spermacoce	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
01denlandi	-ITMUA	EMTMUR	TMERTUR	MTORU-
Thecoccubus	-ITMUA	EMTMUR	TMERTUR	MTORU-
TM01S_H_co	-ITMUA	EMTMUR	TMERTUR	MTORU-
TM01S_H_co	-ITMUA	EMTMUR	TMERTUR	MTORU-





Pentania	-itruva	erutuuf	muutuhi	arokoo-
Dania_sin	-itruva	erutuuf	muutuhi	arokoo-
Ophiorhiz	-itruva	erutuuf	muutuhi	arokoo-





CUHK Libraries



004866467